Annu. Rev. Psychol. 2004. 55:207–34
doi: 10.1146/annurev.psych.55.090902.141415

First published online as a Review in Advance on September 29, 2003

CUMULATIVE PROGRESS IN FORMAL THEORIES OF ATTENTION

Gordon D. Logan
Department of Psychology, Vanderbilt University, Nashville, Tennessee 37203; email: gordon.logan@vanderbilt.edu

Key Words signal detection, similarity, choice, attention, mathematical psychology

■ **Abstract** Formal theories of attention based on similarity-choice theory and signal-detection theory are reviewed to document cumulative progress in theoretical understanding of attention from the 1950s to the present. Theories based on these models have been developed to account for a wide variety of attentional phenomena, including attention to dimensions, attention to objects, and executive control. The review describes the classical similarity-choice and signal-detection theories and relates them to current theories of categorization, Garner tasks, visual search, cuing procedures, task switching, and strategy choice.

CONTENTS

0066-4308/04/0204-0207$14.00

INTRODUCTION

Since the beginning of the cognitive revolution in the 1950s, attention has been a central topic in experimental psychology. In recent years, research on attention has been extended to neuroscience, in studies of humans and monkeys, and to clinical science, in studies of psychopathology. A cynic might argue that the history of research on attention has been a series of unrelated fashions and fads, focusing on different experimental paradigms. In the 1950s and 1960s, the focus was on selective listening. In the 1970s, it was automaticity and dual-task performance. In the 1980s, it was visual search, negative priming, and cuing. In the 1990s, it was the psychological refractory period and the attentional blink. Since the turn of the century, the focus has been on task switching. A cynic might argue that this constant shifting from topic to topic has led to little cumulative progress in our theoretical understanding of attentional phenomena. An optimist might argue that there has been substantial cumulative progress from the 1950s to the present at a deeper level of theory that integrates and explains the relations between the various empirical phenomena. I am an optimist and my purpose in writing this chapter is to document that cumulative progress by reviewing recent developments in formal theories of attention (for earlier reviews of formal theories of attention, see Bundesen 1996 and Swets 1984; for an earlier argument for cumulative progress in studies of attention, see Posner 1982).

The review will be organized around two "families" of theory that derive from seminal work in the 1950s. One class of theory adapts concepts from Shepard's work on similarity scaling and Luce's work on choice to problems of attention. The other class adapts concepts from Green's, Tanner's, and Swets's work on signal-detection theory to problems of attention. The two classes of theory are families in two senses. First, they represent successive generations of theory, with each new theory building on an ancestral theory by elaborating its assumptions or adding new assumptions and by extending its domain of applicability to new problems not addressed by the ancestor. Second, theories within each family share a common formal structure—similarity-choice theory or signal-detection theory—that is analogous to the genetic endowment shared by members of a family. These familial features—the successive elaboration of a powerful mathematical structure across several generations of theory—provide the basis for cumulative progress in understanding attention.

My goal in writing this review is to document cumulative progress and to show relations between formal theories of attention. In meeting this goal, I will ignore much of the recent progress in empirical studies of attention and much of the informal theorizing that accompanies it, hoping that other reviews, past and future, will cover these. My focus on similarity-choice theory and signal-detection theory necessarily excludes other work on formal models of attention, particularly connectionist models. The theories I do cover represent ideas that have been at the core of theoretical psychology since the beginning of the cognitive revolution, and I think it is worthwhile emphasizing their longstanding contributions to the field.

The review begins with a brief overview of similarity-choice theory and signal-detection theory that points out similarities and differences between them and describes conditions under which they are mathematically equivalent. The remainder of the review is organized around three main topics: (*a*) attention to dimensions, which includes subsections on categorization and the tasks introduced by Garner to study dimensional attention; (*b*) attention to objects, which includes subsections on visual search and cuing tasks; and (*c*) executive control of attention, which includes strategies and task switching.

SIMILARITY-CHOICE THEORY

Shepard-Luce Choice Rule

The initial work on similarity-choice theory was done by Shepard (1957) and Luce (1959, 1963). The main goal of similarity-choice theory was to predict choice probabilities from estimates of similarity and bias. According to the theory, the probability of choosing response i given stimulus x is given by the Shepard-Luce choice rule:

$$P(i|x) = \frac{\eta(x,i)\beta_i}{\sum_{j\in R}\eta(x,j)\beta_j}, \tag{1}$$

where $\eta(x, i)$ is the similarity between object x and a representation of response category i and β_i is the subject's bias for giving response i in the choice situation. The probability of choosing response i for object x increases with the similarity between x and i and with the bias for i. The probability of choosing i given x decreases with the similarity between x and the other responses j in the response set R.

Classical similarity-choice theory assumes that objects and categories can be represented as points in multidimensional space and similarity is an exponential function of distance in that space (Shepard 1987). Thus,

$$\eta(x,i) = \exp[-s \cdot d_{xi}],$$

where s is a sensitivity parameter reflecting the steepness of the generalization gradient and d_{xi} is the distance between x and i in the multidimensional space:

$$d_{xi} = \left(\sum_{h=1}^{H}|u_{xh} - u_{ih}|^r\right)^{\frac{1}{r}}. \tag{2}$$

Distance between points is computed by calculating the distance between them along each of the H dimensions in the multidimensional space, raising the dimensional distances to the rth power, summing them over dimensions, and taking the rth root of the sum. The parameter r determines the distance metric. If $r = 1$, the distance metric is city-block; if $r = 2$, the distance metric is Euclidean.

Reaction Time and Response Selection

Similarity-choice theory predicts choice probabilities but not reaction times. Marley & Colonius (1992) and Bundesen (1993) showed that, under very general conditions, independent race models predict the same choice probabilities as the similarity-choice model. That is, for a given similarity-choice model, it is possible to construct an independent race model that gives exactly the same choice probabilities. Independent race models use time to choose among competing alternatives: The first alternative to finish is chosen. The equivalence of similarity-choice models and race models adds a temporal dimension to similarity-choice models and allows them to predict reaction time as well as response probability.

Bundesen (1990) interpreted his theory of visual attention as a race model, in which each categorization of each object raced against the other. He interpreted the elements of the choice equation [e.g., $\eta(x, i)\beta_i$] as rate parameters for exponential distributions. The finishing time for the winner of a race between exponential distributions is itself exponentially distributed with a rate parameter that is the sum of the rate parameters for the individual exponential distributions in the race. The mean finishing time for an exponential distribution is simply the reciprocal of the rate parameter. The mean finishing time for a race involving R categorizations of a single object is

$$FT = \frac{1}{\sum_{j \in R} \eta(x, j)\beta_j}, \tag{3}$$

which is 1 over the denominator of Equation 1. Equation 1 provides the choice probabilities and Equation 3 provides the time to make the choice.

The simple race model expressed in Equations 1 and 3 cannot deal with conflict situations, such as the Stroop (1935) task or the Eriksen & Eriksen (1974) flanker task, in which prepotent responses compete with the required response (Palmeri 1997). The race model predicts that conflict will create difficulty, but it places the effect of difficulty in the wrong dependent variable. The prepotent response likely will be chosen first, so conflict will appear in error rate. The finishing time of the first response determines reaction time, so conflict either will produce no cost in reaction time or it will speed up reaction time. The data show a different pattern: Conflict slows reaction time but has little effect on error rate.

To account for conflict situations, Logan (1996) and Nosofsky & Palmeri (1997a) added response selection processes to the basic similarity-choice model. The response selection processes accumulate evidence provided by the similarity-choice model. The race is run repeatedly and each categorization that comes out of the race is added to an accumulator for one response or another. In Logan's (1996) counter model, a response is selected when one of the accumulators reaches an absolute threshold (i.e., when it accumulates K categorizations). In Nosofsky & Palmeri's (1997a) exemplar-based random walk model, a response is selected when one of the accumulators reaches a relative threshold (i.e., when it accumulates K more categorizations than any other accumulator). The exemplar-based random walk model is particularly powerful. It is related formally to Ratcliff's (1978; Ratcliff et al. 1999) diffusion model, which is the most powerful model of reaction time available today. The diffusion model is a generalization of the random walk model, in which time and evidence are both continuous variables. It accounts for response probabilities and distributions of reaction times for correct and incorrect responses in terms of a small number of parameters. It provides estimates of the rate at which information accumulates (drift rate) but it does not explain why the rates take on different values in different conditions. An important contribution of the exemplar-based random walk model is to provide a theoretical account of variation in drift rate between conditions.

Cumulative Developments

Elements of similarity-choice theory have been used pervasively throughout cognitive psychology. The Shepard-Luce choice rule is used to predict choice probabilities in a variety of theories. Two major families of theory represent cumulative development, one addressing categorization and one addressing attention. The categorization family began with Medin & Schaffer's (1978) context model of classification, which used the Shepard-Luce choice rule to predict classification probabilities. Nosofsky's (1984, 1986, 1988) generalized context model extended Medin & Schaffer's model to include the similarity assumptions inherent in the choice rule. Kruschke (1992) added learning assumptions to Nosofsky's model, and Lamberts (2000) extended the assumptions about similarity to allow dynamic changes in similarity within the course of a single experimental trial. Nosofsky & Palmeri (1997a) combined Nosofsky's model with Logan's (1988) instance theory of automaticity to create the exemplar-based random walk model. The attention family began with Bundesen's (1987) fixed capacity independent race model, which Bundesen (1990) extended to the theory of visual attention. Logan (1996) added a perceptual front end to Bundesen's theory and Logan & Gordon (2001) extended that theory to dual-task performance and executive control. Logan (2002) combined the two families in a single instance theory of attention and memory that includes each of the ancestral theories as a special case.

SIGNAL-DETECTION THEORY

Sensitivity, Bias, and Similarity

Signal-detection theory was developed in the 1950s and 1960s by Tanner & Swets (1954) and Green & Swets (1966), among others. It had antecedents in Thurstone's (1927) work on comparative judgment. Whereas Thurstone was interested in developing psychophysical scales from comparative judgments, the main goal of signal-detection theory was to separate sensitivity from bias in a variety of perceptual discriminations. A key concept in signal-detection theory is the idea of internal noise: Variance in perceptual representations will cause similar stimuli to have overlapping representations. The classic signal-detection situation is a yes-no discrimination task, in which subjects must determine whether a stimulus—a signal—has been presented. The subject gets a single sample from the perceptual system and compares it to a decision criterion, deciding "yes" if the sample exceeds the criterion and "no" if it does not. The overlap of the distributions for signal (technically, signal plus noise) and no signal (noise) determines the subjects' sensitivity. The greater the overlap, the lower the sensitivity. The position of the decision criterion on the decision axis reflects the subject's response bias. A low criterion reflects a bias for saying "yes" and a high criterion reflects a bias for saying "no." The probability of a correct "yes" response—a hit—is proportional to the area of the signal-plus-noise distribution that exceeds the criterion, and the probability of an erroneous "yes" response—a false alarm—is proportional to the area of the noise distribution that exceeds the criterion.

Over the years, signal-detection theory has been used in many ways in many different applications. Often, it is used to generate dependent measures that separate sensitivity and bias (d' and β, respectively) without much theoretical commitment to the underlying processes. However, it has also become the core of several theories of memory, attention, and categorization, providing a language in which to articulate these more complex processes. Classical signal detection, applied to yes-no discrimination tasks, assumes univariate normal distributions with equal variance for signal-plus-noise and noise distributions, though other distributions and other assumptions about variance have been investigated thoroughly (e.g., Green & Swets 1966, Wickens 2002). Many of the applications to more complex processes assume normal distributions and several assume multivariate normal distributions. The general recognition theory of Ashby and his colleagues is a notable example (e.g., Ashby & Lee 1991, Ashby & Maddox 1993, Ashby & Perrin 1988, Ashby & Townsend 1986).

Signal-detection theory assumes that objects are represented as distributions in multidimensional space. As with similarity-choice theory, similarity can be interpreted as distance between objects in the multidimensional space. Distance is stochastic, however, because objects are represented as distributions, not points. Ashby & Perrin (1988) interpreted similarity in terms of overlap between distributions. Objects that are more similar are represented by distributions that overlap more; objects that are less similar are represented by distributions that overlap less. In the special case of multivariate normal distributions with equal variances and covariances, similarities calculated from overlap of distributions are equivalent to similarities calculated from distances between points (i.e., the means of the distributions; Ashby & Maddox 1993, Nosofsky 1992).

Reaction Time and Response Selection

Classical signal-detection approaches deal primarily with situations in which stimuli are weak or confusable, so accuracy is the primary dependent measure. In order to apply signal-detection theory to phenomena of attention, researchers often transform situations that are usually studied with reaction time methods into situations that can be studied with accuracy, by limiting exposure duration or masking the stimuli (e.g., Palmer 1994). Another approach is to assume that reaction time is proportional to the distance from the percept to the criterion. For example, Maddox & Ashby (1996) tested two assumptions about the relation between reaction time and distance, an exponential function

$$RT = \alpha \exp^{-\beta D} + c, \tag{4}$$

following Murdock (1985), and a power function

$$RT = \alpha D^{-\beta} + c, \tag{5}$$

where D is the distance between the percept and the decision boundary and α and β are constants and c is an intercept parameter. In their theory, response selection depends on the region the percept falls in (i.e., where it falls with respect to the decision boundaries) and reaction time depends on the distance from the percept to the decision boundary. This approach seems more descriptive than explanatory because no process interpretation is given for the relation between reaction time and distance. Recently, Ashby (2000) developed a stochastic version of general recognition theory that drives a multivariate diffusion process. This model is related to Ratcliff's (1978; Ratcliff et al. 1999) diffusion model and it provides a process interpretation of the relation between reaction time and distance from the percept to the boundary.

Cumulative Developments

Signal-detection theory may be even more pervasive than similarity-choice theory. It appears in many theories of perception, attention, memory, and categorization. There have been several different threads of cumulative development, but they have not yet been woven together in a single fabric, as Logan (2002) did with similarity-choice theory approaches to attention and categorization. The impressive cumulative developments from signal-detection theory include (*a*) the work of Sperling and his colleagues on focusing and switching attention (e.g., Reeves & Sperling 1986, Shih & Sperling 2002, Sperling & Reeves 1980, Sperling & Weichselgartner 1995); (*b*) the general recognition theory developed by Ashby and colleagues to account for identification and categorization (e.g., Ashby & Lee 1991, Ashby & Maddox 1993, Ashby & Perrin 1988, Ashby & Townsend 1986); (*c*) the work of Dosher & Lu on cuing and focusing attention (Dosher & Lu 2000a,b, Lu & Dosher 1998); and (*d*) the work of Palmer and colleagues on visual search (e.g., Eckstein et al. 2000; Palmer 1994, 1998; Palmer et al. 1993). It would be very interesting to see a formal integration of these different theories, but that must await future research.

ATTENTION TO DIMENSIONS: CATEGORIZATION

Attention researchers do not discuss categorization much, but the concept of attention to dimensions plays an important role in categorization research. Early rule-based theories of categorization assumed that subjects attended to dimensions relevant to categorization and ignored other dimensions (e.g., Bruner et al. 1956). In many experiments, the primary task was to discover which dimensions were relevant (Trabasso & Bower 1968). More recent similarity-based theories of categorization also consider attention to dimensions to be an important process in categorization, including similarity-choice theories and signal-detection theories.

Similarity-Choice Theory

Luce (1963) and others applied the choice rule to a variety of choice tasks that usually involved choosing a single response to a single stimulus (for a review, see Luce 1977). Nosofsky (1984, 1986, 1988) applied the choice rule to categorization, on the assumption that categories were represented as collections of instances. In Nosofsky's generalized context model, the probability of choosing category i for object x was directly related to the sum of the similarities between object x and the various instances of category i the subject had encountered, and inversely related to the sum of the similarities between object x and the various instances of the categories in the response set R. That is,

$$P(i|x) = \frac{\sum_{m=1}^{N_i} \eta(x, i_m)\beta_i}{\sum_{j \in R} \sum_{m=1}^{N_j} \eta(x, j_m)\beta_j}, \tag{6}$$

where N_i is the number of instances in category i.

An important contribution of the generalized context model was to relate choice probabilities in identification tasks to choice probabilities in categorization tasks. Shepard et al. (1961) applied Equations 1 and 2 to identification probabilities, extracted estimates of the similarities, and used them in Equations 1 and 6 to predict categorization probabilities. This attempt was a famous failure, which led Shepard et al. to conclude that the principles that governed identification were separate from the principles that governed categorization. Nosofsky (1984, 1986, 1988) noted that subjects might attend to the dimensions of the stimuli differently in identification and categorization. He rewrote Equation 2 to include a parameter, w_h, that represents the attention given to dimension h:

$$d_{xi} = \left(\sum_{h=1}^{H} w_h \left| u_{hx} - u_{hi} \right|^r \right)^{\frac{1}{r}}. \quad (7)$$

Nosofsky allowed the attention weight, w_h, to take on different values in identification and categorization and succeeded where Shepard et al. had failed: He was able to predict categorization probabilities from identification probabilities and vice versa. Categorization and identification could be explained by the same principles—those underlying similarity-choice theory.

Nosofsky (1984, 1986, 1988) assumed that subjects chose attention weights that optimized performance. Kruschke (1992) extended the generalized context model to include a connectionist model that learned attention weights. Consequently, Kruschke's model provides a better account of category learning than the generalized context model (Nosofsky et al. 1994, Nosofsky & Palmeri 1996).

The idea of dimensional attention weights is very powerful. Nosofsky et al. (1994) extended it to account for rule-based categorization. In their view, a rule amounts to exclusive attention to a single dimension. Johansen & Palmeri (2002) modeled the transition from rule-based categorization to instance-based categorization in terms of a shift from attending exclusively to one dimension to distributing attention across dimensions.

Nosofsky & Palmeri's (1997a) exemplar-based random walk model extends the generalized context model to account for reaction time phenomena as well as for choice probabilities. The random walk allows the exemplar-based random walk model to respond more deterministically than the generalized context model and provides a process interpretation for extensions to the generalized context model that allows it to respond deterministically. Moreover, the random walk process coupled with the idea that learning involves accumulating instances allows the exemplar-based random walk model to account for the effects of frequency of presentation on categorization performance. Nosofsky & Palmeri (1997b) showed that subjects responded faster to the more frequently presented of two stimuli that were equally distant from the decision bound (also see Verguts et al. 2003).

Signal-Detection Theory

General recognition theory accounts for categorization in terms of decision bounds that divide multidimensional space into regions corresponding to each category. A stimulus is classified according to the region in which it falls (e.g., Ashby & Lee 1991, Ashby & Maddox 1993). Ashby & Lee (1991) applied general recognition theory to similarity ratings and identification probabilities and then to the identification and categorization data of Nosofsky (1986). They found that general recognition theory fit Nosofsky's data better than the generalized context model. Moreover, general recognition theory fit the identification and categorization data without assuming a different distribution of attention weights in the two tasks. However, it did require different decision bounds and different assumptions about perceptual independence and perceptual separability. Maddox et al. (1998)

extended general recognition theory to account for reaction time distributions in perceptual categorization tasks. Recently, Maddox et al. (2002) extended general recognition theory to include a perceptual attention component that affects the variance of the multivariate distributions as well as a decisional attention component represented by decision bounds (cf. Bundesen 1990).

In the domain of categorization, general recognition theory competes fiercely with the generalized context model and the exemplar-based random walk model. Each theory provides an impressively exact account of an impressive amount of data. Indeed, the predictions of the two theories are often very similar to each other, despite fundamental differences in the underlying assumptions about the representations and processes used to perform the tasks. Ashby & Maddox (1993) and Nosofsky (1992) have shown conditions under which they are formally equivalent and make exactly the same predictions. It would be interesting to see the same effort extended to compare signal-detection and similarity-choice theories of other aspects of attention.

ATTENTION TO DIMENSIONS: GARNER TASKS

Attention to dimensions was investigated most thoroughly by Garner and his colleagues (see, e.g., Garner 1974). Prominent among their experiments is a filtering task that examines subjects' ability to disregard variation in irrelevant dimensions. For example, subjects may judge the height of rectangles while attempting to ignore their width. Two major phenomena must be explained: Garner interference and redundancy gains. Garner interference is measured by comparing a baseline task with a filtering task. In the baseline task, the relevant dimension varies but the irrelevant dimension is held constant (e.g., judging whether wide rectangles are tall or short). In the filtering task, the two dimensions vary independently (e.g., tall rectangles are wide and narrow; short rectangles are wide and narrow). If subjects are able to filter out variation in the irrelevant dimension, there should be no difference between baseline and filtering conditions—there will be no Garner interference. If subjects are not able to filter out variation in the irrelevant dimension, reaction time will be longer in the filtering task than in the baseline task—there will be Garner interference. Many studies have found that the presence or absence of Garner interference depends on the relation between the dimensions. If the dimensions are separable, like hue and height or brightness and width, then there is no Garner interference. If the dimensions are integral, like hue and brightness or height and width, then there is Garner interference.

Redundancy gain is observed by comparing correlated and orthogonal filtering tasks (Garner 1974). In a correlated task, both dimensions vary but their values are correlated (e.g., tall rectangles are wide; short rectangles are narrow), whereas in an orthogonal task, the two dimensions vary independently (e.g., tall rectangles are wide and narrow; so are short rectangles). A redundancy gain is observed if subjects are faster with the correlated task than with the orthogonal task. Again, whether redundancy gains are observed depends on the relation between the dimensions.

Separable dimensions usually show no redundancy gain; integral dimensions show strong redundancy gains.

Garner's research poses three key questions for formal theories to answer: What causes Garner interference? What causes redundancy gain? And what does it mean for dimensions to be separable or integral? Models based on similarity-choice theory and models based on signal-detection theory have answered these questions.

Similarity-Choice Theory

Nosofsky & Palmeri's (1997a) exemplar-based random walk model accounts for the difference between separable and integral dimensions in terms of the distance metric (i.e., the parameter r in Equations 2 and 7). If $r = 1$, the distance metric is city block and the dimensions are separable. Changing the distance on one dimension has no effect on the distance on the other dimension, so the dimensions can be processed independently without one intruding on the other. If $r = 2$, the distance metric is Euclidean and the dimensions are integral. Changing the distance on one dimension also changes the distance between the objects, so variation in the irrelevant dimension intrudes on judgments of the relevant dimension.

The exemplar-based random walk model accounts for Garner's results with separable dimensions in terms of attention weights. With separable dimensions, all of the attention weight can be given to the relevant dimension, so the irrelevant dimension has no influence. This distribution of attention weight stretches the relevant dimension and collapses the irrelevant dimension. Consequently, there is no difference between baseline and filtering tasks and no difference between orthogonal and correlated filtering tasks.

The exemplar-based random walk model accounts for Garner's results with integral dimensions in terms of differential repetition effects and differential confusions between stimuli. The baseline condition involves fewer stimuli than the filtering condition. In typical experiments, the baseline condition involves only two stimuli whereas the filtering condition involves four. Thus, stimuli and responses are more likely to repeat in the baseline condition, and repetition of stimuli and responses reduces reaction time (Nosofsky & Palmeri 1997a).

The different numbers of stimuli in the baseline and filtering tasks also creates differential confusions between stimuli. Consider a baseline condition in which subjects judge the height of narrow rectangles and a filtering condition in which they judge the height of wide and narrow rectangles. Imagine the subject is presented with a tall narrow rectangle. In the baseline condition, there is only one stimulus to be confused with the target—the short narrow rectangle. In the filtering condition, however, there are two stimuli to be confused with the target—the short narrow rectangle and the short wide rectangle. Consequently, the probability of choosing the target will be lower in the filtering condition and this will slow the random walk. The lower the choice probability, the slower the rate at which target categorizations accumulate. Moreover, the lower the choice probability, the more incorrect categorizations will accumulate in the other counter. The exemplar-based random walk model chooses a response when one counter has K more categorizations than any other, so the greater the number of categorizations in the incorrect counter, the greater the number of correct categorizations that must be accumulated. Nosofsky & Palmeri (1997a, 1997b) predict the data quantitatively.

The exemplar-based random walk model accounts for the difference between correlated and orthogonal filtering tasks in terms of differential repetition and differential distance between the alternatives. As with the baseline-filtering contrast, the contrast between correlated and orthogonal filtering tasks involves different numbers of stimuli. In a typical design, the correlated condition involves two stimuli whereas the orthogonal condition involves four. As with the baseline-filtering contrast, there are more opportunities for stimulus-response repetitions in the correlated condition than in the orthogonal condition, and this differential facilitation from repetition accounts for some of the difference in reaction time (Nosofsky & Palmeri 1997a).

The difference in distance between alternatives can be seen by imagining a set of four stimuli that vary on two dimensions arrayed as a square in multidimensional space. The orthogonal task requires subjects to discriminate the left two stimuli from the right two stimuli. The correlated task requires subjects to discriminate the bottom left stimulus from the top right stimulus. Imagine the bottom left stimulus has been presented. In the orthogonal task, the nearest (most confusable) alternative is the bottom right stimulus. In the correlated task, the nearest alternative is the top right stimulus. The bottom right stimulus is closer to the target than the top right stimulus, so choice probability will be lower, and reaction time will be longer. Nosofsky & Palmeri (1997a) predicted this difference quantitatively. Nosofsky & Palmeri (1997b) extended these predictions to reaction time distributions.

Signal-Detection Theory

In general recognition theory, the concepts of integrality and separability are closely tied to the concepts of perceptual and decisional independence (Ashby & Maddox 1994, Ashby & Townsend 1986). The theory distinguishes between perceptual separability and decisional separability. Consider a set of four stimuli produced by factorially combining two levels of two components, A and B. Perceptual separability holds if the perceptual effects of one stimulus component are independent of the perceptual effects of another. This can be assessed by comparing the marginal distributions of perceptual effects, $g_{AiB1}(x)$ and $g_{AiB2}(x)$, for level i of component A at levels 1 and 2 of component B. Perceptual separability holds if $g_{AiB1}(x) = g_{AiB2}(x)$ for $i = 1$ and 2. Perceptual separability is violated if $g_{AiB1}(x) \neq g_{AiB2}(x)$ for $i = 1$ and 2. Ashby & Maddox (1994) consider two cases in which perceptual separability can be violated: mean shift integrality and variance shift integrality. In mean shift integrality, the mean of the distribution of perceptual effects for one level of one component is different from the mean of the distribution of perceptual effects for the other level (i.e., the distributions have the same shape but the means are shifted relative to each other). In variance shift integrality, the means of the distributions are the same but the variances are different. Both mean-shift integrality and variance-shift integrality will produce violations of perceptual separability.

Decisional separability occurs if the decision about one component is unaffected by the perceptual effect of the other component. This occurs when the decision boundaries are parallel to the coordinate axes of the multidimensional space. Perceptual independence is assessed from the covariance between the distributions of perceptual effects in multidimensional space. Perceptual independence holds if the covariances are zero (for a complete discussion of independence, see Ashby & Townsend 1986).

The results in Garner's tasks are predicted from these assumptions about separability and integrality and from the idea that reaction time is proportional to the distance between the percept and the decision bound (Equations 4 and 5). If perceptual and decisional separability hold, then the percepts will be the same distance from the decision bound in the baseline and filtering tasks—there will be no Garner interference. If perceptual separability fails and decisional separability holds, then some percepts will be closer to the decision bound in the filtering condition than in the baseline condition, resulting in slower reaction times. In this way, general recognition theory predicts Garner interference with (perceptually) integral dimensions.

In orthogonal and correlated filtering tasks, the percepts will be the same distance from the decision bound if perceptual and decisional separability hold—there will be no redundancy gain. Maddox & Ashby (1996) noted that the optimal decision bounds were not orthogonal to the coordinate axes in the correlated filtering task (e.g., when the mean of one distribution is in the top left and the mean of the other is in the bottom right). With this configuration of distributions, the optimal decision bound is a diagonal line (going from the top right to the bottom left). This decision bound is optimal in the sense that it maximizes the distance between the percept and the decision bound, and that will speed reaction time. Consequently, Maddox & Ashby (1996) predicted redundancy gains with perceptually separable stimuli in the correlated filtering task, and they found them. In terms of their theory, they predicted that the correlated task would violate decisional separability.

ATTENTION TO OBJECTS: VISUAL SEARCH

Identification and categorization tasks represent a universe in which there is only one object. The real world and many attention tasks represent a universe in which there are several objects, and choosing an object to respond to is a significant problem. For the past 20 years, much research has been done on visual search tasks, in which subjects are faced with a display of many objects and must decide whether the display contains a target object. The main independent variable is display size—the number of objects in the display—and the main dependent variable is a measure of search efficiency—the extent to which reaction time or accuracy or both are affected by variation in the number of objects in the display. If search is efficient, reaction time and accuracy are largely unaffected by the number of objects in the display. If search is inefficient, reaction time and accuracy are strongly affected by the number of objects in the display. Much of the research has focused on factors that determine search efficiency and a number of alternative hypotheses have been proposed, some of which are informal and some of which are formal. Similarity-choice theory and signal-detection theory have been applied to these problems, suggesting their own hypotheses about the determinants of search efficiency.

Similarity-Choice Theory

The main contribution of Bundesen's (1990) theory of visual attention was an application of similarity-choice theory to the problem of object selection. His initial work focused primarily on partial report tasks (Sperling 1960), in which subjects are presented with several stimuli and are required to identify some of them. In partial report tasks, the objects to be identified typically share a property or set of properties that are independent of the properties relevant to identification. For example, subjects might be shown a display of red and black letters and asked to report the red ones. This kind of object selection was called stimulus set (Broadbent 1971), input selection (Treisman 1969), and filtering (Kahneman & Treisman 1984) in classical theories of attention.

In the theory of visual attention, objects are assigned attentional weights, and the probability of selecting an object increases with the attentional weight on the object. For a display of objects that are homogeneous (in the sense of Bundesen 1990, p. 524) the probability that object x is the first object selected is given:

$$P_\pi(x) = \frac{\sum_{k \in S} \eta(x, k)\pi_k}{\sum_{z \in D} \sum_{k \in S} \eta(z, k)\pi_k}, \tag{8}$$

where $\eta(x, k)$ is the extent to which object x has the property k in the stimulus set S (the set of target-defining properties), and D is the set of objects that are displayed on a given trial. The parameter π_k reflects the priority given to objects with property k. It has the same effect as the bias parameter in the Shepard-Luce choice rule but it is represented by a different symbol because it has a different function: π reflects the priority given to objects in the stimulus set, whereas β reflects the bias given to different categories in the response set. In the language of classical theories of attention, π reflects stimulus set or input selection, whereas β reflects response set (Broadbent 1971) or analyzer selection (Treisman 1969).

Changing priority changes the distribution of attention across the objects in the display. Objects whose properties are in the stimulus set S get more attention than objects whose properties are not in the stimulus set. By Equation 8, the $P_\pi(z)$ values sum to 1.0 across the display, so increasing the priority of one object necessarily decreases the priority of the others. This constraint limits the theory of visual attention's processing capacity (Bundesen 1990, Logan 2002).

Bundesen (1990) combined stimulus set and response set multiplicatively, in order to account for stimulus selection and response selection (for an explanation of the multiplicative combination, see Logan 2002). In the theory of visual attention, the probability of choosing object x and identifying it as a member of category i is given by:

$$P(x \cap i) = \frac{\eta(x, i)\beta_i P_\pi(x)}{\sum_{z \in D} \sum_{j \in R} \eta(z, j)\beta_j P_\pi(z)}. \tag{9}$$

Bundesen (1990) applied this equation to partial and whole report tasks and extended it to a variety of attention tasks, including efficient and inefficient visual search.

EFFICIENT AND INEFFICIENT SEARCH The results from similarity-choice theories are generally consistent with more qualitative models (Treisman & Gelade 1980) or simulation models (Cave & Wolfe 1990, Wolfe 1994) that assume that efficient search is done in parallel and inefficient search is done in series. Bundesen (1990) modeled Treisman & Gelade's (1980) feature search task, in which a target differs from a set of homogeneous distractors in terms of a single feature (e.g., a red target among green distractors). If the target is not similar to the distractors, this task yields very efficient search. Bundesen modeled target-present responses as a race between alternatives. The distractors are not very similar to the target, so they intrude little on the race and reaction time is largely independent of the number of distractors. He modeled target-absent responses in terms of a deadline. The model responded "absent" if a target was not found by the time the deadline expired. However, Chun & Wolfe (1996) suggest that temporal deadlines may not be the best way to terminate search on target-absent trials. Logan (2002) extended the theory of visual attention to include the similarity assumptions of the generalized context model and the exemplar-based random walk model. He noted that subjects can vary attention weights (see Equation 7) in feature search tasks to increase the distance between the targets and the distractors in multidimensional similarity space, which increases choice probability and speeds reaction time.

Bricolo et al. (2002) provided serial and parallel models of inefficient search tasks, inspired by the theory of visual attention, that address changes in reaction time distributions as a function of the number of objects in the display (display size) and serial position effects on mean reaction time. The results of the display size experiment showed that the minimum of the reaction time distribution increased with display size for both target-present and target-absent responses but the increase was much larger for target-absent responses. This result was predicted by a serial self-terminating model and by a parallel self-terminating model with fixed capacity that is reallocated after each object is finished processing. The results of the serial position experiment showed an increase in reaction time with serial position, which could be predicted by a serial self-terminating model and by a parallel self-terminating model with fixed capacity that is not reallocated after each object is finished processing. The same serial model could account for both sets of results, but different parallel models were required for the two experiments, so Bricolo et al. (2002) favored the serial model.

Logan (2002) noted that subjects could not vary attention weights to optimize performance in many tasks that yield inefficient search. In conjunction search tasks, in which the target shares one feature with half of the distractors and another feature with the other half of the distractors, varying attention weight to move one of the distractors further away from the target moves the other distractor closer to the target, resulting in no net gain in discriminability. Logan (1996) provided decision rules for conjunction search tasks and noted that they predicted faster decisions for targets that differed from distractors on two features (triple conjunctions) than for targets that differed from distractors on one feature (double conjunctions; see Wolfe et al. 1989).

DISTANCE AND GROUPING BY PROXIMITY A common finding in a variety of attention tasks is that performance is affected by the proximity of the objects. Nearby distractors impair target performance more than distant distractors (e.g., Eriksen & Eriksen 1974). Grouping by proximity is powerful and can have strong effects on performance. Targets that are grouped together with distractors are harder to find. Targets that are isolated from distractors are easier to find (e.g., Banks & Prinzmetal 1976). Bundesen's (1990) theory of visual attention cannot account for these effects. It assumes all objects in the display are processed in parallel and their processing rates are the same regardless of their spatial arrangement. Logan (1996) extended the theory of visual attention to account for the effects of proximity and grouping by proximity, adding Van Oeffelen & Vos's (1982, 1983) COntour DEtector (CODE) theory of perceptual grouping by proximity as a "front end" to produce the CODE theory of visual attention.

The CODE theory assumes that the representation of object location is not a point, but rather, is distributed over space (also see Ashby et al. 1996). The distributions representing nearby objects overlap substantially; distributions representing distant objects overlap very little. According to CODE, the distributions add together to produce a CODE surface that looks something like a mountain range. Grouping by proximity is determined by imposing a threshold on the CODE surface (i.e., drawing a plane of a certain height across the CODE surface) to create one or more above-threshold regions. The CODE theory claims that objects that fall within the same above-threshold region are part of the same perceptual group. CODE predicts subjects' grouping by proximity very well in textbook demonstrations (Van Oeffelen & Vos 1982, 1983) and reasonably well in random dot patterns (Compton & Logan 1993, 1999).

The CODE theory of visual attention assumes that the theory of visual attention samples information from the perceptual groups defined by above-threshold regions on the CODE surface. The probability that the theory of visual attention will sample a given object is equal to the proportion of the area or volume of the distribution of that object that falls within the sampled above-threshold region. Objects whose centers fall within the region are likely to be sampled because the central parts of their distributions fall within the above-threshold region. Objects whose centers fall outside the region (i.e., objects in different perceptual groups) will also be sampled with a probability equal to the proportion of their distribution that falls within the sampled above-threshold region, but that probability is lower because only the tails of the object intrude in the sampled above-threshold region. Moreover, the more distant the object, the smaller the proportion of its distribution will fall within the sampled above-threshold region, so the smaller the impact it will have on performance. Logan (1996) showed that these ideas accounted for distance and grouping by proximity effects in a variety of attention tasks, and Logan & Bundesen (1996) showed they accounted for distance and grouping effects in partial report tasks.

Signal-Detection Theory

Signal-detection approaches to search generally assume there are no perceptual capacity limitations on search and interpret search efficiency in terms of the impact of discriminability on a parallel decision process, inconsistent with feature integration theory (Treisman & Gelade 1980) and guided search (Cave & Wolfe 1990, Wolfe 1994, Wolfe et al. 1989), and consistent with parallel models (Duncan & Humphreys 1989, Heinke & Humphreys 2003, Humphreys & Müller 1993). Palmer and colleagues applied signal-detection theory to a variety of visual search tasks, including those that produce efficient and inefficient search (Eckstein et al. 2000; Palmer 1994, 1998; Palmer et al. 1993). In their models, a decision process takes a sample from each object in the display and decides "target present" if the largest sample exceeds a criterion and "target absent" if it does not exceed the criterion—that is, by applying a "max rule." Display size effects occur because of the max rule. On target-absent trials, the noise distribution is the distribution of the maximum of the values sampled from N distractors, and the mean of this distribution increases with N. On target-present trials, the signal-plus-noise distribution is the distribution of the maximum of $N - 1$ values sampled from the distractors and one value sampled from the target. The mean of this distribution also increases with N. The discriminability of these distributions of maxima depends on the discriminability of a single target from a single distractor—that is, on d'. The smaller the d' for a single discrimination, the greater the overlap between the noise and signal-plus-noise distributions of maxima, and the overlap increases as N increases. Thus, search tasks with hard discriminations produce large display size effects (inefficient search) and search tasks with easy discriminations produce small display size effects (efficient search). The same parallel decision process is used for easy and hard search. The difference in search efficiency is produced entirely by noise in the decision process.

Eckstein et al. (2000) extended the model to conjunction search tasks, in which targets and distractors vary on several dimensions. In their model, information from different dimensions is collapsed onto a single decision variable, which produces distributions of maxima for target-present and target-absent trials. If there are f relevant feature dimensions and the target differs from the distractors along r of them, then sensitivity for the conjunction discrimination, $d'_{r,f}$, is

$$d'_{r,f} = \frac{r d'_0}{\sqrt{f}},$$

where d'_0 is the sensitivity along each of the r dimensions, which is assumed to be equal for each dimension. This model produces a good quantitative description of the differences between feature, conjunction, and triple-conjunction conditions.

Geisler & Chou (1995) presented a signal-detection model of search performance in which sensitivity varied with eccentricity. They presented displays of single objects at known locations and measured subjects' ability to discriminate the target from the distractor. The known locations varied in eccentricity and discrimination performance decreased as eccentricity increased. Then they presented displays in which a target was or was not superimposed on a uniform texture made from the distractor pattern and in which target location was unknown. Performance on this task was completely predictable from the decline in discriminability with eccentricity, which suggests there were no capacity limits. This conclusion may be limited by presenting the distractors as a uniform field. The equivalence of performance when location is known and unknown may be another example of the general finding that search for a single target is not facilitated by knowledge of target location (Shiu & Pashler 1994).

These signal-detection approaches to visual search are impressive, but the situations they model are different in important ways from the situations in which visual search is usually studied. In all cases, the displays are presented briefly so that accuracy is the main dependent measure. The brief displays prevent sequential sampling from the display, either with covert attention or with overt eye movements. By contrast, typical search tasks present the display until the subject responds, allowing plenty of time for sequential sampling. Eye movements and covert shifts of attention may be important phenomena in visual search but these signal-detection approaches ignore them. Perhaps it should not be surprising that displays too brief to allow sequential sampling can be modeled with parallel processes. More importantly, it is not clear how the models applied to brief displays can be extended to account for reaction times with response-terminated displays. That is an important direction for future research.

ATTENTION TO OBJECTS: CUING PROCEDURES

Attention to objects is often studied by presenting subjects with displays of multiple objects and giving them cues that indicate the target's location or some other salient property. In some procedures, each object in the display is a potential target and the cue indicates which object to judge or report (e.g., Eriksen & Hoffman 1972, Sperling 1960). Consequently, subjects cannot respond to the target without first responding to the cue. In other procedures, the target differs from the distractors in some way and the cue merely indicates its position (e.g., Posner et al. 1980).

In these cases, subjects can respond to the target without first responding to the cue; nevertheless, the cue influences performance. Valid cues that indicate the correct location of the target facilitate performance, speeding reaction time and increasing accuracy. Invalid cues that indicate a location that does not contain the target impair performance, slowing reaction time and decreasing accuracy.

Similarity-Choice Theory

Bundesen (1987, 1990) applied his theory of visual attention to a variety of partial report tasks, including classical data from Sperling (1960). In his own experiments, he varied exposure duration and the number of targets and distractors. In all cases, the theory fit the data very well (also see Bundesen 2002, Logan & Bundesen 1996, Shibuya & Bundesen 1988). Recently, Duncan et al. (1999) used the theory of visual attention to analyze partial and whole report performance in neglect patients. In partial report, patients showed a bias against contralesional targets but a preserved bilateral ability to prioritize targets. In whole report, they showed a bilateral reduction in processing capacity, which suggests a bilateral component to neglect.

Bundesen (1990, 1998) applied the theory of visual attention to the cuing effects reported by Posner et al. (1980). The cue altered the priority (π in Equation 8) given to the cued location but had no effect on the bias (β in Equation 9) given to the target. Increasing priority increases the processing rate for the target on valid cue trials, speeding reaction time. Because the $P_{\pi}(z)$ values are constrained to sum to 1.0 across the display, increasing priority in an invalid location (on invalid cue trials) necessarily decreases processing rate for the target and slows reaction time.

Logan (2002) extended Bundesen's analysis, contrasting situations in which the cue was necessary to specify the target (e.g., Eriksen & Hoffman 1972, Sperling 1960) and situations in which the target could be specified independent of the cue (e.g., Posner et al. 1980). When the cue specifies the target, cuing can be accounted for entirely in terms of priority (π in Equation 8). However, when the target is specified independent of the cue and the cue merely indicates its location, both priority and bias (β in Equation 9) contribute to the cuing effect. Thus, from the perspective of the theory of visual attention, researchers interested in separating priority from bias should study situations in which cues specify the target and researchers interested in the interaction between priority and bias should study situations in which the cue merely indicates the target's location.

Signal-Detection Theory

Over the past 20 years, Sperling and colleagues have developed an episodic attention theory that accounts for performance in a variety of attention tasks (Reeves & Sperling 1986, Shih & Sperling 2002, Sperling & Reeves 1980, Sperling & Weichselgartner 1995). The most detailed applications of the theory have been to situations that involve rapid serial visual presentation (RSVP) to measure the reaction time of attention shifts, the discrete versus continuous nature of attention shifts, and the trajectory of attention through time and space. Sperling & Reeves (1980) measured reaction time by presenting two RSVP streams. When subjects detected a target in one of the streams, they were to shift attention to the other stream and report the first item they could. The lag between the target in one stream and the item reported in the other reflects the attention reaction time. Reeves & Sperling (1986) elaborated the procedure and the theory, requiring subjects to report as many items as they could from the second stream after detecting the target in the first. They found that subjects had good information about item identity (up to the limit of short-term memory) but poor information about the temporal order in which the items appeared. They modeled performance in terms of an attention gate that opened at the second stream some time after the target was detected in the first stream. The attention gate took the form of a gamma function (the convolution of two exponentials), and the strength of the tendency to report an item depended on the area it subtended under the gamma function. The values along the abscissa were determined by the presentation rate of the items, and the values along the ordinate were determined by the height of the gamma function. This attention gating model provided an excellent account of the data.

Sperling & Weichselgartner (1995) extended the attention gating theory to create the episodic theory of attention. In their theory, visual attention consists of a series of discrete attention episodes. Each episode is characterized by a three-dimensional distribution of attention that extends in time and space. A central contribution of this theory was to model attention shifts as discrete shifts from one spatial distribution to another, contrary to previous attempts to model attention shifts as continuous movements of a "spotlight" across space. Sperling & Weichselgartner (1995) fit their model to the data of several experiments that claimed to find evidence of spatially continuous shifts of attention. They found that their model accounted for the data very well without assuming a continuous shift.

Recently, Shih & Sperling (2002) extended the RSVP procedure and the model further to measure the trajectory of attention through time and space. Their RSVP procedure involves presenting successive displays of three rows of three items and cuing subjects to report items from one of the rows with a tone that varied in pitch (cf. Sperling 1960). They measured the trajectory of attention through time by noting which displays the reported letters came from and they measured the trajectory of attention through space by noting which rows the reported letters came from. They also tested their subjects on partial report tasks in which the delay of the cue and a poststimulus mask were varied and on whole report tasks in which mask delay was varied, using the same model to account for performance on these tasks and the RSVP task. They relate their theory to other procedures intended to measure shifts of spatial attention and to other models of visual spatial attention, such as Bundesen's (1990) theory of visual attention.

On the one hand, the RSVP procedure used in these experiments is quite different from most procedures used to measure cuing attention, so it is not immediately clear how the results from this procedure relate to results from the other procedures. Moreover, the model deals with accuracy of responding and it is not clear how it would deal with reaction time. On the other hand, Sperling and colleagues have taken great pains to apply their model to other procedures, formally in some cases and informally in others. The relations between the procedures can be seen through the model. Indeed, the main purpose of theories and models is to show the relations between disparate procedures.

Dosher & Lu (2000a,b; Lu & Dosher 1998) used a noisy perceptual template model to investigate the costs and benefits of valid and invalid cues, using external noise to distinguish among attention mechanisms. They added increasing amounts of external noise (white Gaussian random noise) to a visual stimulus and observed the effects on contrast thresholds. Typically, adding external noise has no effect on contrast threshold as long as external noise is smaller than the internal noise in the system. When external noise exceeds internal noise, contrast thresholds increase approximately linearly with the amount of noise, when plotted in log-log coordinates. Dosher & Lu distinguished three different attention mechanisms that affected these log-log plots in different ways: "Signal enhancement" shows an increased threshold for invalidly cued trials in the flat part of the log-log plot where external noise is smaller than internal noise. "External noise exclusion" shows an increased threshold for invalidly cued trials in the linear increasing part of the log-log plot where external noise is larger than internal noise. "Internal noise reduction" shows an increased threshold for invalidly cued trials throughout the log-log plot (i.e., in both the flat and the linearly increasing parts of the function). When there were only two locations, the pattern of performance was consistent with signal enhancement (Lu & Dosher 1998, Lu et al. 2000). When there were four or more locations, the pattern of performance was consistent with external noise exclusion (Dosher & Lu 2000a,b).

Again, the procedure required to assess the effects of adding external noise differs substantially from the usual procedures used to measure attention. It is not clear how models of near-threshold performance would extend to reaction times to stimuli that are exposed until the subject responds. These are important questions for future research.

EXECUTIVE CONTROL OF ATTENTION

Executive control is the process by which the mind programs itself. It is involved in understanding instructions, choosing among strategies, preparing and adopting a task set, monitoring performance, and disengaging task sets. It is an important and popular topic in cognitive science, neuroscience, clinical science, developmental science, and the study of individual differences. Optimization of performance is an important feature of executive control. Most theories assume optimality in one way or another but leave the task of optimization to an omnipotent homunculus that is outside the theory. A key idea in many studies of executive control is that an executive process programs subordinate processes. As influential as that idea is, it is empty without some specification of the subordinate process that says how it can be programmed. Formal theories of attention provide this specification and flesh out the idea.

Similarity-Choice Theory

Nosofsky (1984, 1986, 1988) assumed that attention weights were distributed across dimensions to optimize performance. Indeed, the best-fitting attention weights are often the ones that optimize performance. Kruschke (1992) replaced Nosofsky's homunculus with a connectionist module that learns to distribute attention weights across dimensions based on feedback during classification learning. It accounts for asymptotic categorization performance as well as the generalized context model and it provides a better account of classification learning (see, e.g., Nosofsky et al. 1994, Nosofsky & Palmeri 1996).

Logan & Gordon (2001) addressed the idea that an executive process programs a subordinate by proposing a theory of executive control in which Bundesen's (1990) theory of visual attention and Nosofsky & Palmeri's (1997a) exemplar-based random walk model were the subordinates. In the theory of visual attention and the exemplar-based random walk model, priority and bias parameters and the threshold for the random walk are determined by the homunculus, whereas the similarity parameters are determined by the quality of the stimulus information and the subject's experience with the members of the relevant categories. The theory of visual attention and the exemplar-based random walk model can be programmed—set to perform different tasks—by manipulating priority, bias, and threshold parameters. In Logan & Gordon's (2001) theory, a task set is a set of parameters that are sufficient to cause the theory of visual attention and the exemplar-based random walk model to perform particular tasks. Task sets are constructed by deriving parameters from propositional representations of the instructions in working memory. Task sets are enabled by transmitting the parameters from working memory to the place where the models reside in the processing system. Deriving the parameters in working memory is not sufficient to enable or change a task set. The task set must also be instantiated in the theory of visual attention and the exemplar-based random walk model. By analogy, having a program on disk is not sufficient to make it run. It must be loaded into core memory before it can be executed. Transmission of parameters takes time, and that time accounts for task-switching costs.

Logan & Gordon (2001) applied their theory to a dual-task situation called the psychological refractory period procedure, in which subjects must make separate responses to two stimuli that appear close together in time. In principle, the theory of visual attention and the exemplar-based random walk model could perform these two tasks in parallel or in series, but Logan and Gordon showed through simulations that performance was faster and more accurate if the models performed the tasks in series. They were able to account for task-switching costs and backward crosstalk from the second stimulus to the first with the model.

Logan & Bundesen (2003) extended the model to task-switching situations in which a cue indicating which task to perform is presented before each target stimulus. They developed formal models that assumed that the time-course functions reflect the cumulative distributions of finishing times for processes that encoded the cue and switched task sets (also see Sperling & Weichselgartner 1995). Reaction time is slow when cue encoding and set switching have not finished and fast when they have finished. Increasing the interval between the cue and the target increases the probability that cue encoding and set switching are finished. Consequently, reaction time decreases as cue-to-target interval increases. The mean cue-encoding time and set-switching time can be estimated by fitting these models to the time-course functions (also see Logan & Bundesen 1996).

Signal-Detection Theory

The idea of optimality has been a central tenet of signal-detection theory since the 1950s. It remains a central tenet in the general recognition theory (Ashby & Lee 1991, Ashby & Maddox 1993). Signal-detection theory was intended to describe normative decision making under uncertainty by an ideal observer (Geisler 1989, Green & Swets 1966). The β parameter in signal-detection theory, which reflects the placement of the criterion on the decision axis, incorporates the values of correct responses and errors as well as their probabilities. An ideal observer chooses a value of β that maximizes gains and minimizes losses (Green & Swets 1966).

Perhaps the most thorough signal-detection analysis of optimality in attention tasks was done by Sperling & Dosher (1986). They addressed single- and dual-task performance with accuracy and reaction time as dependent measures, examining tradeoffs between concurrent tasks and tradeoffs between speed and accuracy in single tasks. A key concept in their analysis is the idea of a performance-operating characteristic, which plots one measure of performance against another. In classical signal-detection theory, a receiver-operating characteristic plots the probability of hits against the probability of false alarms (Green & Swets 1966). In dual-task studies, an attention-operating characteristic plots performance on one task against performance on another (Sperling & Melchner 1978). Sperling & Dosher (1986) defined a strategy as a choice of a point on a performance-operating characteristic and they argued that choice of a strategy was determined by the expected utility that could be gained at that point.

CONCLUSIONS

In the spirit of competitive hypothesis testing that is so pervasive in psychology, it seems natural to ask, Which theory wins? That is a hard question to answer. It has been clear from the beginning that similarity-choice theory and signal-detection theory make very similar predictions (e.g., Broadbent 1971, Luce 1963). Recent analyses have shown that under some assumptions, the theories mimic each other exactly (Ashby & Maddox 1993, Nosofsky 1992). Moreover, the theories have been applied to so many different domains that it is hard to evaluate the outcome of a single battle in the context of such a large-scale war. Perhaps the theories should be evaluated in terms of other criteria, such as the explanations they provide for attentional phenomena and the assumptions they require to do so.

My personal impression is that similarity-choice theories provide better explanations. They provide processing interpretations of the phenomena that give insight into the underlying computations. Bundesen (1990) describes object selection in terms of competition instantiated as a race between perceptual objects. Nosofsky (1984, 1986, 1988) describes classification in terms of a competition instantiated as a race between memory traces. By contrast, signal-detection approaches seem more descriptive than explanatory. I find it hard to imagine the processes underlying a decision that an object falls within a region of similarity space (but see Ashby 2000). Perhaps my impression is due to differential familiarity—I have spent many more hours thinking about how to interpret attentional phenomena in terms of similarity-choice theory than in terms of signal-detection theory.

Similarity-choice theories and signal-detection theories differ fundamentally in their assumptions about noise. Similarity-choice theories assume no noise in perceptual representations; objects are represented as points in similarity space. Signal-detection theories assume noisy representations; objects are represented as distributions in similarity space. On the one hand, it seems reasonable to think of representations of category exemplars as relatively noise free. Subjects often examine exemplars at their leisure and have plenty of time to encode the nuances of the stimuli. It seems less reasonable to think that the representations are so noisy that their distributions overlap enough to cause frequent confusions. On the other hand, it seems reasonable to think of representations of brief stimuli as noisy. Indeed, it is hard to imagine how adding external noise to a weak stimulus would produce the bilinearity that Dosher & Lu (2000a,b) observe in log-log plots if the internal representation were not noisy. Similarity-choice theories may have a hard time dealing with results from that procedure. Perhaps this issue can be resolved by quantifying noise in neural representations of objects.

Similarity-choice theories and signal-detection theories both assume that objects can be represented as points or distributions in multidimensional similarity space. This representation is limited. It is not clear how structured objects can be represented as points or distributions. Moreover, the idea that similarity is proportional to distance in multidimensional space is problematic. Tversky spent much of his career challenging the metric assumptions of multidimensional similarity models (e.g., Tversky 1977, Tversky & Gati 1982, Tversky & Hutchinson 1986). Indeed, the similarity between structured objects may not be captured very well by a multidimensional similarity space. Medin et al. (1993) showed that relational measures of similarity may be more appropriate for such stimuli.

Ultimately, the most important question in evaluating similarity-choice and signal-detection theories may be this: What are the alternatives? No other theory of attention has a legacy as rich and as powerful as either of these theories. Their mathematical structure allows strong inferences and precise predictions. They make sense of diverse phenomena, bringing order to the fads and fashions that dominate empirical research on attention. Theories that adopt more complex assumptions about object representations are much more specialized, and consequently, lack the generality of similarity-choice and signal-detection approaches (e.g., Heinke & Humphreys 2003, Humphreys & Müller 1993). The theories reviewed here set a high standard for clarity, consistency, and longevity. They have allowed cumulative progress in our understanding of attention over the last 50 years and they promise to increase our cumulative understanding for many years to come. I am optimistic.

ACKNOWLEDGMENTS

This research was supported by grants BCS 0133202 and 0218507 from the National Science Foundation. I would like to thank Claus Bundesen and Tom Palmeri for valuable comments on the manuscript.

The *Annual Review of Psychology* is online at http://psych.annualreviews.org

LITERATURE CITED

Ashby FG. 2000. A stochastic version of general recognition theory. *J. Math. Psych.* 44:310–29

Ashby FG, Lee WW. 1991. Predicting similarity and categorization from identification. *J. Exp. Psychol.: Gen.* 120:150–72

Ashby FG, Maddox WT. 1993. Relations between prototype, exemplar, and decision bound models of categorization. *J. Math. Psych.* 37:372–400

Ashby FG, Maddox WT. 1994. A response time theory of separability and integrality in speeded classification. *J. Math. Psych.* 38:423–66

Ashby FG, Perrin NA. 1988. Toward a unified theory of similarity and recognition. *Psychol. Rev.* 95:124–50

Ashby FG, Prinzmetal W, Ivry R, Maddox WT. 1996. A formal theory of feature binding in object perception. *Psychol. Rev.* 103:165–92

Ashby FG, Townsend JT. 1986. Varieties of perceptual independence. *Psychol. Rev.* 93:154–79

Banks WP, Prinzmetal W. 1976. Configurational effects in visual information processing. *Percept. Psychophys.* 19:361–67

Bricolo E, Gianesini T, Fanini A, Bundesen C, Chelazzi L. 2002. Serial attention mechanisms in visual search: a direct behavioral demonstration. *J. Cogn. Neurosci.* 14:980–93

Broadbent DE. 1971. *Decision and Stress.* London: Academic

Bruner JS, Goodnow JJ, Austin GA. 1956. *A Study of Thinking*. New York: Wiley

Bundesen C. 1987. Visual attention: race models for selection from multielement displays. *Psychol. Res.* 49:113–21

Bundesen C. 1990. A theory of visual attention. *Psychol. Rev.* 97:523–47

Bundesen C. 1993. The relationship between independent race models and Luce's choice axiom. *J. Math. Psych.* 37:446–71

Bundesen C. 1996. Formal models of visual attention: a tutorial review. In *Converging Operations in the Study of Visual Selective Attention*, ed. AF Kramer, MGH Coles, GD Logan, pp. 1–44. Washington, DC: Amer. Psychol. Assoc.

Bundesen C. 1998. Visual selective attention: outlines of a choice model, a race model, and a computational theory. *Vis. Cogn.* 5:287–309

Bundesen C. 2002. A general theory of visual attention. In *Psychology at the Turn of the Millennium: Vol. 1. Cognitive, Biological,* *and Health Perspectives*, ed. L Bäckman, C von Hofsten, pp. 179–200. Hove, UK: Psychology Press

Cave KR, Wolfe JM. 1990. Modeling the role of parallel processing in visual search. *Cogn. Psychol.* 22:225–71

Chun MM, Wolfe JM. 1996. Just say no: How are visual searches terminated when there is no target present? *Cogn. Psychol.* 30:39–78

Compton BJ, Logan GD. 1993. Evaluating a computational model of perceptual grouping by proximity. *Percept. Psychophys.* 53:403–21

Compton BJ, Logan GD. 1999. Judgments of perceptual groups: reliability and sensitivity to stimulus transformation. *Percept. Psychophys.* 61:1320–35

Dosher BA, Lu Z-L. 2000a. Mechanisms of perceptual attention in precuing of location. *Vis. Res.* 40:1269–92

Dosher BA, Lu Z-L. 2000b. Noise exclusion in spatial attention. *Psychol. Sci.* 11:139–46

Duncan J, Bundesen C, Olson A, Humphreys G, Chavda S, Shibuya H. 1999. Systematic analysis of deficits in visual attention. *J. Exp. Psychol.: Gen.* 128:450–78

Duncan J, Humphreys GW. 1989. Visual search and stimulus similarity. *Psychol. Rev.* 96:433–58

Eckstein MP, Thomas JP, Palmer J, Shimozaki SS. 2000. A signal detection model predicts the effects of set size on visual search accuracy for feature, conjunction, triple conjunction, and disjunction displays. *Percept. Psychophys.* 62:425–51

Eriksen BA, Eriksen CW. 1974. Effects of noise letters upon the identification of a target letter in a nonsearch task. *Percept. Psychophys.* 16:143–49

Eriksen CW, Hoffman JE. 1972. Temporal and spatial characteristics of selective encoding from visual displays. *Percept. Psychophys.* 12:201–4

Garner WR. 1974. *The Processing of Information and Structure.* New York: Wiley

Geisler WS. 1989. Sequential ideal-observer analysis of visual discriminations. *Psychol. Rev.* 96:267–314

Geisler WS, Chou KL. 1995. Separation of low level and high level factors in complex tasks: visual search. *Psychol. Rev.* 102:356–78

Green DM, Swets JA. 1966. *Signal Detection Theory and Psychophysics.* New York: Krieger

Heinke D, Humphreys GW. 2003. Attention, spatial representation, and visual neglect: simulating emergent attention and spatial memory in the selective attention for identification model (SAIM). *Psychol. Rev.* 110:29–87

Humphreys GW, Müller HJ. 1993. SEarch via Recursive Rejection (SERR): a connectionist model of visual search. *Cogn. Psychol.* 25:43–110

Johansen MK, Palmeri TJ. 2002. Are there representational shifts during category learning? *Cogn. Psychol.* 45:482–553

Kahneman D, Treisman A. 1984. Changing views of attention and automaticity. In *Varieties of Attention*, ed. R Parasuraman, DR Davies, pp. 29–61. New York: Academic

Kruschke JK. 1992. ALCOVE: an exemplar-based connectionist model of category learning. *Psychol. Rev.* 99:22–44

Lamberts K. 2000. Information accumulation theory of categorization response times. *Psychol. Rev.* 107:227–60

Logan GD. 1988. Toward an instance theory of automatization. *Psychol. Rev.* 95:492–527

Logan GD. 1996. The CODE theory of visual attention: an integration of space-based and object-based attention. *Psychol. Rev.* 103:603–49

Logan GD. 2002. An instance theory of attention and memory. *Psychol. Rev.* 109:376–400

Logan GD, Bundesen C. 1996. Spatial effects in the partial report paradigm: a challenge for theories of visual-spatial attention. In *The Psychology of Learning and Motivation*, Vol. 35, ed. DL Medin, pp. 243–82. San Diego, CA: Academic

Logan GD, Bundesen C. 2003. Clever homunculus: Is there an endogenous act of control in the explicit task-cuing procedure? *J. Exp. Psychol.: Hum. Percept. Perform.* 29:575–99

Logan GD, Gordon RD. 2001. Executive control of visual attention in dual task situations. *Psychol. Rev.* 108:393–434
Lu Z-L, Dosher BA. 1998. External noise distinguishes attention mechanisms. *Vis. Res.* 38:1183–98
Lu Z-L, Lui CQ, Dosher BA. 2000. Attention mechanisms for multi-location first- and second-order motion perception. *Vis. Res.* 40:173–86
Luce RD. 1959. *Individual Choice Behavior.* New York: Wiley
Luce RD. 1963. Detection and recognition. In *Handbook of Mathematical Psychology*, ed. RD Luce, RR Bush, E Galanter, pp. 103–89. New York: Wiley
Luce RD. 1977. The choice axiom after twenty years. *J. Math. Psychol.* 15:215–33
Maddox WT, Ashby FG. 1996. Perceptual separability, decisional separability, and the identification-speeded classification relationship. *J. Exp. Psychol.: Hum. Percept. Perform.* 22:795–817
Maddox WT, Ashby FG, Gottlob LR. 1998. Response time distributions in multidimensional perceptual categorization. *Percept. Psychophys.* 60:620–37
Maddox WT, Ashby FG, Waldron EM. 2002. Multiple attention systems in perceptual categorization. *Mem. Cogn.* 30:325–39
Marley AAJ, Colonius H. 1992. The "horse race" random utility model for choice probabilities and reaction times, and its competing risks interpretation. *J. Math. Psychol.* 36:1–20
Medin DL, Goldstone RL, Gentner D. 1993. Respects for similarity. *Psychol. Rev.* 100: 254–78
Medin DL, Schaffer MM. 1978. Context theory of classification. *Psychol. Rev.* 85:207–38
Murdock BB. 1985. An analysis of the strength-latency relationship. *Mem. Cogn.* 13:511–21
Nosofsky RM. 1984. Choice, similarity, and the context theory of classification. *J. Exp. Psychol.: Learn. Mem. Cogn.* 10:104–14
Nosofsky RM. 1986. Attention, similarity, and the identification-categorization relationship. *J. Exp. Psychol.: Gen.* 115:39–57
Nosofsky RM. 1988. Exemplar-based accounts of relations between classification, recognition, and typicality. *J. Exp. Psychol.: Learn. Mem. Cogn.* 14:700–8
Nosofsky RM. 1992. Similarity scaling and cognitive process models. *Annu. Rev. Psychol.* 43:25–53
Nosofsky RM, Gluck MA, Palmeri TJ, McKinley SC, Glauthier P. 1994. Comparing models of rule-based classification learning: a replication and extension of Shepard, Hovland, and Jenkins (1961). *Mem. Cogn.* 22:352–69
Nosofsky RM, Palmeri TJ. 1996. Learning to classify integral-dimension stimuli. *Psychon. Bull. Rev.* 3:222–26
Nosofsky RM, Palmeri TJ. 1997a. An exemplar-based random walk model of speeded classification. *Psychol. Rev.* 104:266–300
Nosofsky RM, Palmeri TJ. 1997b. Comparing exemplar-retrieval and decision-bound models of speeded perceptual classification. *Percept. Psychophys.* 59:1027–48
Nosofsky RM, Palmeri TJ, McKinley SC. 1994. Rule-plus-exception model of classification learning. *Psychol. Rev.* 101:53–79
Palmer J. 1994. Set-size effects in visual search: the effect of attention is independent of the stimulus for simple tasks. *Vis. Res.* 34:1703–21
Palmer J. 1998. Attentional effects in visual search: relating search accuracy and search time. In *Visual Attention*, ed. RD Wright, pp. 348–88. New York: Oxford Univ. Press
Palmer J, Ames CT, Lindsey DT. 1993. Measuring the effect of attention on simple visual search. *J. Exp. Psychol.: Hum. Percept. Perform.* 19:108–30
Palmeri TJ. 1997. Exemplar similarity and the development of automaticity. *J. Exp. Psychol.: Learn. Mem. Cogn.* 23:324–54
Posner MI. 1982. Cumulative development of attentional theory. *Am. Psychol.* 37:168–79
Posner MI, Snyder CRR, Davidson BJ. 1980. Attention and the detection of signals. *J. Exp. Psychol.: Gen.* 109:160–74
Ratcliff R. 1978. A theory of memory retrieval. *Psychol. Rev.* 85:59–108
Ratcliff R, van Zandt T, McKoon G. 1999.

Connectionist and diffusion models of reaction time. *Psychol. Rev.* 106:261–300

Reeves A, Sperling G. 1986. Attention gating in short-term visual memory. *Psychol. Rev.* 93:180–206

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

Shepard RN. 1987. Toward a universal law of generalization for psychological space. *Science* 237:1317–23

Shepard RN, Hovland CI, Jenkins HM. 1961. Learning and memorization of classifications. *Psychol. Monogr.* 75:1–42

Shibuya H, Bundesen C. 1988. Visual selection from multielement displays: measuring and modeling the effects of exposure duration. *J. Exp. Psychol.: Hum. Percept. Perform.* 14:591–600

Shih S-I, Sperling G. 2002. Measuring and modeling the trajectory of visual spatial attention. *Psychol. Rev.* 109:260–305

Shiu L, Pashler H. 1994. Negligible effect of spatial precuing on identification of single digits. *J. Exp. Psychol.: Hum. Percept. Perform.* 20:1037–54

Sperling G. 1960. The information available in brief visual presentations. *Psychol. Monogr.* 74:(11, Whole No. 498)

Sperling G, Dosher BA. 1986. Strategy and optimization in human information processing. In *Handbook of Perception and Performance*, Vol. 1, ed. K Boff, L Kaufman, J Thomas, pp. 2-1– 2-65. New York: Wiley

Sperling G, Melchner MJ. 1978. The attention operating characteristic: some examples from visual search. *Science* 202:315–18

Sperling G, Reeves A. 1980. Measuring the reaction time of an unobservable response: a shift of visual attention. In *Attention and Performance VIII*, ed. RS Nickerson, pp. 347–60. New York: Academic

Sperling G, Weichselgartner E. 1995. Episodic theory of the dynamics of spatial attention. *Psychol. Rev.* 102:503–32

Stroop JR. 1935. Studies of interference in serial verbal reactions. *J. Exp. Psychol.* 18:643–62

Swets JA. 1984. Mathematical models of attention. In *Varieties of Attention*, ed. R Parasuraman, DR Davies, pp. 183–242. San Diego, CA: Academic

Tanner WP, Swets JA. 1954. A decision-making theory of visual detection. *Psychol. Rev.* 61:401–9

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

Trabasso TR, Bower GH. 1968. *Attention and Learning: Theory and Research.* New York: Wiley

Treisman A. 1969. Strategies and models of selective attention. *Psychol. Rev.* 76:282–99

Treisman A, Gelade G. 1980. A feature-integration theory of attention. *Cogn. Psychol.* 12:97–136

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 JW. 1986. Nearest neighbor analysis of psychological spaces. *Psychol. Rev.* 93:3–22

Van Oeffelen MP, Vos PG. 1982. Configurational effects on the enumeration of dots: counting by groups. *Mem. Cogn.* 10:396–404

Van Oeffelen MP, Vos PG. 1983. An algorithm for pattern description on the level of relative proximity. *Pattern Recogn.* 16:341–48

Verguts T, Storms G, Tuerlinckx F. 2003. Decision-bound theory and the influence of familiarity. *Psychon. Bull. Rev.* 10:141–48

Wickens TD. 2002. *Elementary Signal Detection Theory.* New York: Oxford Univ. Press

Wolfe JM. 1994. Guided search 2.0: a revised model of visual search. *Psychonom. Bul. Rev.* 1:202–38

Wolfe JM, Cave KR, Franzel SL. 1989. Guided search: an alternative to the feature integration model for visual search. *J. Exp. Psychol.: Hum. Percept. Perform.* 15:419–33