An instance theory of attention and memory — **Source link** ⬈

Gordon D. Logan

Institutions: Vanderbilt University

Topics: Categorization

Related papers:

- Toward an instance theory of automatization

- A theory of visual attention.

- An Exemplar-Based Random Walk Model of Speeded Classification

- Attention, similarity, and the identification-categorization relationship.

- Neural Mechanisms of Selective Visual Attention

An Instance Theory of Attention and Memory

Gordon D. Logan
Vanderbilt University

An instance theory of attention and memory (ITAM) is presented that integrates formal theories of attention and memory phenomena by exploiting commonalities in their formal structure. The core idea in each theory is that performance depends on a choice process that can be modeled as a race between competing alternatives. Attention and categorization are viewed as different perspectives on the same race. Attention selects objects by categorizing them; objects are categorized by attending to them. ITAM incorporates each of its ancestors as a special case, so it inherits their successes.

Imagine yourself on your way home from work. You walk into the parking lot and look for your car. It takes you a second, perhaps. Now imagine your colleagues analyzing the simple act of cognition underlying that look. A student of attention would be interested in how your gaze went to the cars rather than other structural features. A student of categorization would be interested in how you knew those were cars in the parking lot. And a student of memory would be interested in how you did (or did not) pick your own car out of the group. These differences in perspective reflect the divide-and-conquer approach to understanding cognition that has been prevalent among researchers for the last 20 or 30 years. Different researchers focus on the details of different parts of cognition, hoping that their work will interface with the rest. Attention may begin with perception and end when a target is found. Categorization may begin when a target is found and end with a concept. Memory may begin with a concept and end with a recollection or a feeling of familiarity.

This article takes a different approach, trying to integrate theories developed under the divide-and-conquer strategy. It views the simple act of cognition as a single phenomenon and interprets attention, categorization, and memory as different perspectives on the same simple act. To attend is to categorize; to categorize is to remember; to remember is to attend. This idea is expressed in a formal theory that combines existing theories of attention, categorization, and memory that have similar formal structures. I show how the theory derives from its ancestors and I show how the ancestors can be viewed as special cases of the theory.

The Family Tree

The divide-and-conquer strategy separates empirical phenomena as well as theories. Empirical studies of attention focus on the problem of selecting one of several objects—finding your car among the other cars in the parking lot. Empirical studies of categorization and memory focus on the problem of classifying a single object—knowing that cars are cars and your car is yours. The empirical phenomena are different and seem to require different explanations. Theories of attention focus on similarities and differences between objects, whereas theories of categorization and memory focus on features or dimensions of single objects. The new theory provides a conceptual bridge that connects the empirical phenomena and explicates relations between them. It provides a unified account of three different kinds of attention that are required to explain the phenomena: attention to objects, attention to categories, and attention to dimensions.

The new theory is the child of two families of theory, an attention family and a memory family, depicted in Figure 1. In view of this heritage, I call the new theory the *instance theory of attention and memory* (ITAM). Its ancestors on each branch are members of a family in the sense that they share formal mathematical and computational structures, like members of a family share genes. The ancestors are also like a family in that they represent several generations of cumulative theoretical developments. The two families have common roots that go back almost 50 years to ideas about similarity and choice by Shepard (1957) and Luce (1959). Each theory in the family tree was created from its immediate ancestor by generalizing existing assumptions or adding new assumptions to increase the theory's power. The new theory culminates these developments by combining theories of attention with theories of memory. It integrates the ancestors by expressing them in a common mathematical structure that includes the ancestors as special cases. The theories that can be viewed as special cases of the new theory are connected by bidirectional arrows in Figure 1. Because of this common formal structure, the new theory inherits the successes of its special cases, which are numerous and impressive.

The attention branch begins with the *fixed-capacity independent race* model (FIRM; Bundesen, 1987; Bundesen, Pedersen, & Larsen, 1984; Bundesen, Shibuya, & Larsen, 1985; Shibuya & Bundesen, 1988), which accounts for performance on partial and whole report tasks. Bundesen (1990, 1998a, 1998b; Bundesen & Harms, 1999) generalized FIRM to create the *theory of visual attention* (TVA), which accounts for performance on a broad range of attention tasks, including feature search, cuing, single-item identification, and priority learning. I combined TVA with Van Oeffelen and Vos's (1982, 1983) COntour DEtector (CODE) theory of perceptual grouping by proximity to create the *CODE*

Attention Branch *Memory Branch*

FIRM

TVA

CTVA

ECTVA

Context Model

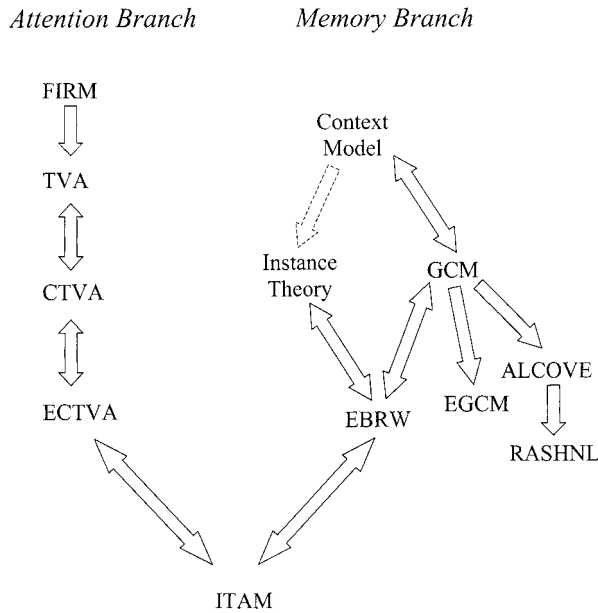Instance Theory

GCM

ALCOVE

EGCM

EBRW

RASHNL

ITAM

Figure 1. A family tree expressing the relations between ancestors on the attention branch and the memory branch. Theories connected by bidirectional arrows can be made equivalent to each other, through ITAM. See the text for a description of the theories. FIRM = fixed-capacity independent race model; TVA = theory of visual attention; CTVA = COntour DEtector (CODE) theory of visual attention; ECTVA = executive control of TVA; GCM = generalized context model; EBRW = exemplar-based random walk; EGCM = extended generalized context model; ALCOVE = attention learning COVEring theory; RASHNL = rapid attention SHifts 'N' Learning theory; ITAM = instance theory of attention and memory.

theory of visual attention (CTVA; Logan, 1996; Logan & Bundesen, 1996), which accounts for a variety of distance and grouping effects in flanker tasks and feature- and conjunction-search tasks. Recently, R. D. Gordon and I proposed a theory of executive control of TVA to account for crosstalk and set-switching effects in dual-task situations. We called this theory *executive control of TVA* (ECTVA; Logan & Gordon, 2001). TVA, CTVA, and ECTVA are special cases of ITAM, so ITAM inherits these predictions. ITAM can be configured in a form that is exactly like each of these ancestors and so generates the same predictions.

The attention branch provides ITAM with a formal theory of attention to objects and attention to categories, but it does not explain attention to dimensions. Its formal arguments rely heavily on the idea of similarity between targets and distractors, but it does not provide a theory of similarity that explains how similarity is represented or computed. The attention branch also assumes that categories are learned and past experiences are remembered, but it has no explanation of learning and memory. The memory branch solves these problems.

The memory branch begins with Medin and Schaffer's (1978) *context model of classification*, which formalized the *instance* or *exemplar* approach to classification and accounted for several effects previously attributed to prototypes. Nosofsky (1984, 1986, 1988) extended the context model's assumptions about similarity to create the generalized context model (GCM), which accounts for an impressive array of classification data, including the relation

between classification and identification and the relation between classification and episodic recognition. Nosofsky and Palmeri (1997; also see Palmeri, 1997) combined GCM with my *instance theory of automaticity* (Logan, 1988, 1990, 1992) to create the *exemplar-based random-walk* model (EBRW), which accounts for classification speed as well as accuracy and a variety of learning effects, including automatization. GCM, instance theory, and EBRW are all special cases of the new theory, so it inherits their successes as well.

The memory branch also contains Lamberts's (2000) *extended generalized context model* (EGCM), Kruschke's (1992) *attention learning covering theory* (ALCOVE), and Kruschke and Johansen's (1999) *rapid attention shifts 'n' learning* theory (RASHNL), which change the representational assumptions of GCM and so are not related as directly to the new theory. They represent promising directions for extension and development of the new theory, which I leave for future research.

The memory branch provides ITAM with a formal theory of attention to categories and attention to dimensions, but it does not explain attention to objects. It provides a formal theory of similarity and mechanisms that determine the effects of similarity on classification performance. It also provides a formal theory of learning (i.e., instance theory and EBRW) and interfaces readily with theories of other kinds of learning (e.g., ALCOVE and RASHNL). The theories in the memory branch complement the theories in the attention branch. Together, they provide ITAM with the necessary tools to account for empirical phenomena in both literatures.

The goal of this article is to present the new theory, explain how it works, and explain how it relates to its ancestors. In a nutshell, ITAM assumes that attention and categorization are both choice processes and that both are instantiated as races between competing alternatives. Attention involves choice between competing objects in the display, whereas categorization involves choice between competing classifications of display objects. ITAM assumes that the races underlying these choices run simultaneously and in fact are one and the same. An object is selected and a classification of that object is selected in the same act of cognition. The choice processes are driven by similarities between display objects and memory representations of the alternative categories. Categories are represented as collections of instances, and learning occurs through the accumulation of instances over practice. The output of object selection and category selection is input to a random-walk response selection process. In the remainder of this article, I crack the nutshell and unpack its contents, explaining how the parts of the theory fit together and how the theory works as a whole. I proceed by considering fundamental similarities and incidental differences between the theories, building the theory one issue at a time and ending with the theory as a whole.

Attention and Categorization as Choice

The most fundamental relationship among all the theories in the family tree is that each one is a choice process. Memory theories look inward, choosing among the categorizations available in memory. Attention theories look outward, choosing among perceptual objects available in the environment. Both choices are massively parallel. Memory theories assume that a probe accesses all traces in memory simultaneously, and attention theories assume

that each object in the display is processed at the same time. In the new theory, these two choices are one and the same. Objects are selected and categorized in a single step.

The importance of choice in attention and memory goes beyond the formal models in Figure 1. Attention has been treated as a choice process from the beginning of experimental psychology (e.g., James, 1890). To attend is to choose between alternative stimuli or alternative courses of action. Categorization also involves choice. The same object can be categorized in many ways (Brown, 1958). Categorizing an object involves choice among alternative categories. Moreover, attention and categorization are studied experimentally in procedures that require choice: Is a target present or absent? Is this stimulus a member of Category A or Category B? Understanding subjects' behavior in those experiments means understanding their choices and the processes that underlie them.

Each theory in the family tree assumes that the choice of a categorization for object x is a function of the strength of the evidence that x is a member of one of the categories j in the response set **R**. The probability of choosing category i for object x increases with evidence that x is i and decreases with evidence that x is a member of some other category in the response set. This core idea is expressed in each theory in a choice rule based on an equation developed by Shepard (1957) and Luce (1959, 1963), the Shepard-Luce choice rule:

$$P(\text{``}x \text{ is } i\text{''}) = \frac{v(x, i)}{\sum_{j \in R} v(x, j)}. \tag{1}$$

The probability of choosing category i for object x is given by the ratio of the strength of evidence that x is i to the sum of the strengths of evidence that x belongs to each category in the response set.

Choice Models as Race Models

The theories differ in their interpretation of the mechanism of choice in a way that turns out to be superficial. Some theories, such as GCM, interpret the choice process as an instance of the Shepard-Luce choice rule. The v parameter is broken down into a product of an *evidence parameter* (η for evidence) and a *bias parameter* (β for bias) according to the following equation:

$$v(x, i) = \eta(x, i)\beta_i. \tag{2}$$

The η parameter represents the quality of the sensory evidence that object x is a member of category i, or alternatively, the similarity between object x and members of category i. The stronger the evidence or the greater the similarity, the larger the η. The η value is determined by the environment and the person's history with members of category i. Its interpretation will be discussed in depth later in the article (see *Similarity and Choice*; *Instance Representation*; *Learning*). The β parameter represents the importance of choosing objects in category i. It is determined by the person's homunculus (Bundesen, 1990; Logan & Gordon, 2001) and represents the bias for members of category i. Its interpretation will be discussed in more detail shortly.

Other theories, such as TVA, interpret the choice process in terms of the outcome of a race between independent processes. In these theories, the v values in Equation 1 are interpreted as processing rates (v for velocity), so the categorization with the fastest processing rate is most likely to win the race. If the v parameters are interpreted as processing rates for exponential distributions of categorization times, then Equation 1 describes the probability that "x is i" wins the race exactly. When there is only one stimulus in the display, TVA represents the v parameters just as GCM does, as products of evidence parameters (ηs) and bias parameters (βs; i.e., Equation 2). The race model has the same structure as the choice model.

Marley and Colonius (1992) and Bundesen (1993) showed that a large class of independent race models is equivalent to the class of Luce choice models in that a race model can be constructed that mimics exactly the choice probabilities of a given Luce model. The equivalence of independent race models and Luce choice models is essential to the theoretical integration in ITAM. It means that each theory in Figure 1 can be construed both as a race model and as a choice model. That adds considerably to the predictive power of the models. Models based on the Shepard-Luce choice rule can predict reaction time (RT) as well as choice probability. If the v values in Equation 1 are interpreted as rate parameters for exponential distributions of finishing times, then the winner of the race will also be distributed exponentially with a rate parameter that is the sum of the rate parameters of all the runners in the race (i.e., the denominator of Equation 1). The mean finishing time (FT) is 1.0 over the denominator of Equation 1 (Bundesen, 1990; Logan, 1996; Townsend & Ashby, 1983):

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

Attention to Categories

The multiplication of η and β in Equation 2 represents ITAM's mechanism for attention to categories. ITAM assumes that η is set obligatorily in a brief encoding stage that precedes TVA (see Bundesen & Harms, 1999; Logan & Gordon, 2001), but β is set by the person's homunculus (i.e., ECTVA; see Logan & Gordon, 2001). The ηs can be thought of as strengths of resonance between structures in the environment and structures in memory, and the βs can be thought of as gain controls. Increasing β turns up the gain on desired categorizations, amplifying their resonance with the environment (increasing the v values) and increasing the probability they will be selected. Decreasing β turns down the gain on undesired categorizations, attenuating their resonance and decreasing the probability they will be selected. The ηs may be high for many categorizations of many display objects, but only those objects with high η and high β have a chance of being selected. Thus, ITAM assumes that selection of a classification or a category is a voluntary process that is mediated by attention.

ITAM's mechanism for attention to categories is inherited from both branches of the family tree and represents a fundamental compatibility between the families. The mechanism has deeper roots in the classical attention literature. The choice of categories to distinguish for a given input is an act of attention that was called *response set* (Broadbent, 1971) and *analyzer selection* (Treisman, 1969) in classical analyses of attention. In ITAM, the choice among categories is a choice among β values.

Selecting Objects; Selecting Categorizations of Objects

The choice machinery in Equations 1–3 works in an ideal universe in which there is just one object. The real world typically presents us with several objects at once, and that creates new computational problems. We are often required to act on one of the objects—you choose your own car in the parking lot—and that creates the problem of object selection: How do you choose the object you categorize? Traditionally, researchers have adopted the divide-and-conquer approach to this issue, assuming that object selection and category selection were done in separate steps. Attention researchers studied object selection and largely ignored the issues surrounding category selection. Categorization researchers studied category selection and ignored issues of object selection.

ITAM takes a different perspective. It assumes that object selection and category selection are the same thing. An object is selected by categorizing it, and a category is selected by finding an object that instantiates it (Bundesen, 1990; Logan, 1996; Logan & Gordon, 2001; also see Desimone & Duncan, 1995; Duncan, 1996). ITAM assumes that object selection occurs in parallel over the set of objects z in the display set $\mathbf{D}$, just as category selection occurs in parallel over all categories j in the response set $\mathbf{R}$. Consequently, there is a processing rate $v(z, j)$ for every categorization of every display object. As in Equation 1, choice probabilities depend on the ratio of the processing rate for categorizing a particular object to the sum of the processing rates for all categorizations of all objects in the display. Thus, the probability of choosing object x and classifying it as a member of category i depends on the processing rate for that categorization—$v(x, i)$:

$$P(x \cap i) = \frac{v(x, i)}{\sum_{z \in D} \sum_{j \in R} v(z, j)}. \quad (4)$$

The mean processing time for this choice of object and category is the reciprocal of the sum of the processing rates for all categorizations of all display objects:

$$FT = \frac{1}{\sum_{z \in D} \sum_{j \in R} v(z, j)}. \quad (5)$$

Equations 4 and 5 represent the probability and finishing time for a single choice. Performance in many tasks, such as visual search and partial report, requires several choices like these before a response can be chosen (see *Response Selection*), so Equations 4 and 5 do not always predict performance directly. They are useful because they provide insight into the advantages of ITAM's assumption that objects and categories are selected simultaneously.

Early, Late, and Simultaneous Selection

ITAM's assumption of simultaneous selection opposes long-standing tradition in the attention literature. Since the 1950s, attention researchers have treated object selection and category selection as separate, sequential stages. They have always agreed there are two stages and that one follows the other, but they have never agreed on the order of the stages or the boundary between the stages, a controversy known as the *locus of selection*. Advo-

cates of *early selection* argue that objects are selected before they are categorized (e.g., Broadbent, 1958, 1971; Treisman & Gelade, 1980). Advocates of *late selection* argue that objects are categorized before they are selected (e.g., Deutsch & Deutsch, 1963; Norman, 1968). The controversy remains unresolved, with evidence supporting both positions (Kahneman & Treisman, 1984). ITAM provides a new perspective on the issue, suggesting a third alternative—simultaneous selection—that looks like early selection in some respects and late selection in other respects (Logan, 1996). The current evidence may be more consistent with simultaneous selection than with early or late selection (also see Logan, 1995a; Logan & Zbrodoff, 1999; Van der Heijden, 1992).

ITAM provides some insight into the controversy over the locus of selection because it can implement a version of early selection and a version of late selection that can be compared with each other and with ITAM's simultaneous selection. I show that simultaneous selection produces the same accuracy as these versions of early and late selection, but it produces faster processing times.

Early selection. According to early selection theories, objects are first selected and then categorized. The choice machinery in Equations 1–4 can implement each of these steps. The probability of choosing object x from a set of objects in the display $\mathbf{D}$ is the sum of the probabilities that x will be classified in each of the relevant categories. The processing rates for object x are summed over all categories j in the response set $\mathbf{R}$, and that sum is divided by the sum of processing rates over all categories j in the response set $\mathbf{R}$ over all objects z in the display $\mathbf{D}$. That is,

$$P(x) = \frac{\sum_{j \in R} v(x, j)}{\sum_{z \in D} \sum_{j \in R} v(z, j)}. \quad (6)$$

Once object x is chosen, the category it belongs to may be chosen using what is essentially Equation 1:

$$P(i \mid x) = \frac{v(x, i)}{\sum_{j \in R} v(x, j)}. \quad (7)$$

Assuming that i is the correct categorization and x is the correct object, the probability of responding correctly in both steps is given by the product of Equations 6 and 7:

$P(x \cap i) = P(x)P(i \mid x)$

$$= \frac{\sum_{j \in R} v(x, j)}{\sum_{z \in D} \sum_{j \in R} v(z, j)} \frac{v(x, i)}{\sum_{j \in R} v(x, j)} = \frac{v(x, i)}{\sum_{z \in D} \sum_{j \in R} v(z, j)}. \quad (8)$$

Note that the far right side of Equation 8 is the same as the right-hand side of Equation 4. Early selection (Equation 8) and simultaneous selection (Equation 4) predict the same choice probabilities. They mimic each other's accuracy predictions. Early selection provides no advantage over simultaneous selection in terms of accuracy.

Processing-time predictions show a disadvantage for early selection. Early selection involves two stages: object selection and category selection. If the stages are sequential and independent, then processing time is the sum of the stages' durations; that is,

$$FT_{\text{early}} = \frac{1}{\sum_{z \in D} \sum_{j \in R} v(z, j)} + \frac{1}{\sum_{j \in R} v(x, j)}. \qquad (9)$$

The duration of the first stage in early selection—object selection—is the same as the duration of the only stage—object and category selection—in simultaneous selection (compare the first term on the right-hand side of Equation 9 with Equation 5), so early selection is necessarily slower than simultaneous selection. Thus, ITAM's assumption of simultaneous selection is faster but no less accurate than early selection. ITAM can configure itself to perform early selection or simultaneous selection, but it would prefer simultaneous selection because it is faster.

Late selection. According to late selection theories, objects are first categorized, and then one of the categorized objects is selected. This can be implemented in ITAM by calculating the probability of categorizing objects in category i and then calculating the probability that object x will be selected from those in category i. The probability of choosing a member of category i from all objects z in the display $\mathbf{D}$ is

$$P(i) = \frac{\sum_{z \in D} v(z, i)}{\sum_{z \in D} \sum_{j \in R} v(z, j)}. \qquad (10)$$

The probability that object x is selected from among the objects in category i is

$$P(x \mid i) = \frac{v(x, i)}{\sum_{z \in D} v(z, i)}. \qquad (11)$$

The probability that i and x are both chosen is the product of Equations 10 and 11:

$$P(x \cap i) = P(i)P(x \mid i)$$

$$= \frac{\sum_{z \in D} v(z, i)}{\sum_{z \in D} \sum_{j \in R} v(z, j)} \frac{v(x, i)}{\sum_{z \in D} v(z, i)} = \frac{v(x, i)}{\sum_{z \in D} \sum_{j \in R} v(z, j)}. \qquad (12)$$

Note that the far right-hand side of Equation 12 is identical to the far right-hand side of Equation 8 and identical to the right-hand side of Equation 4: Early, late, and simultaneous selection produce the same choice probabilities.

The processing time for late selection is the sum of the durations of the two sequential independent steps, category selection and object selection:

$$FT_{\text{late}} = \frac{1}{\sum_{z \in D} \sum_{j \in R} v(z, j)} + \frac{1}{\sum_{z \in D} v(z, i)}. \qquad (13)$$

The first step in late selection takes as long as the only step in simultaneous selection (compare Equations 4 and 13), so late selection necessarily takes longer than simultaneous selection. Simultaneous selection produces the same accuracy in less time and so is preferable to late selection. A beast endowed with ITAM's choice machinery could choose strategically between early, late, and simultaneous selection, but the advantage would usually go to simultaneous selection because it is faster and no less accurate than the alternatives.

Simultaneous selection of a display object and a categorization of the object illustrates one sense in which attention and categorization are the same thing. The act that results in apprehension of a display object is the same act that results in a categorization of the display object. We apprehend by categorizing. Classical early- and late-selection views assumed that categorization and object selection were separate steps, and that encouraged the divide-and-conquer strategy of studying them separately. ITAM's idea that object selection and categorization are one and the same and occur in one step discourages studying them separately and strongly recommends understanding them together in a single integrated theory.

Object Selection and Stimulus Set

Imagine yourself in a parking lot looking for a car you just rented. The clerk gave you a key with the license number on it and told you the car was red. You have to find the car with the matching license number, but it may be helpful to know the car is red. You can restrict your attention to red cars, examining the license plate only if the car is red. If there is only one red car in the lot, this strategy will lead you directly to it. More likely there will be a few red cars among several non-red ones. Restricting your attention to the red cars will reduce the number of license plates you need to examine.

Your experience in the parking lot has been studied extensively by attention researchers for decades. You chose your car on the basis of one property (color) and responded to it on the basis of another (license number). Many attention tasks similarly require people to select objects on the basis of one property and to select responses on the basis of another. Subjects might be shown a display of letters and asked to report the identity of the red one (e.g., C. W. Eriksen & Hoffman, 1972), or they might be shown a row of three letters and asked to report the middle one (e.g., B. A. Eriksen & Eriksen, 1974). Subjects might be shown a display of items arranged in rows and asked to report the identities of the items in the row cued by a tone or a bar marker (e.g., Sperling, 1960). Kahneman and Treisman (1984) called this broad class of tasks *filtering tasks*. The property used to select objects serves as a filter that screens out distractor objects that do not share the property, reducing the number of objects that require attention and allowing attention to focus on the target object. Examples of filtering tasks are illustrated in Figure 2.

In classical analyses of attention, the process used to select the object is called *stimulus set* (Broadbent, 1971) or *input selection* (Treisman, 1969) and the process used to select the categorization of the object that is to be reported is called *response set* (Broadbent, 1971) or *analyzer selection* (Treisman, 1969). In classical theories, stimulus set and response set are separate, sequential stages. The object is first selected by one process and then categorized by another process. The processes may change roles in different contexts. You may find a red object and classify it as X or O in one context and find an X and classify it as red or green in another context. The processes play only one role in a single context—stimulus set or response set—and they play their roles sequentially—first stimulus set, then response set (Treisman,

Identify the white letter

P $\mathbb{R}$ B X D M

Identify the middle letter

S H S

Identify the letters in the bottom row

X T K S
D L C R
B N P C

Figure 2. Examples of filtering tasks.

1969). First you find the red car, then you compare license numbers.

ITAM also distinguishes between stimulus set and response set. It agrees with classical theories in assuming that stimulus set selects objects and response set selects categories, but it disagrees with the classical theories in two respects. First, it assumes that response set participates in object selection as well as category selection. In other words, it assumes that two processes participate in object selection—stimulus set and response set—whereas classical theories assumed that one process was involved. This assumption makes ITAM similar to guided search theory (K. R. Cave & Wolfe, 1990; Wolfe, 1994; Wolfe, Cave, & Franzel, 1989), which assumes that target categories have a top-down influence on object selection. I show that ITAM's combination of stimulus set and response set produces better object selection than response set alone. Second, ITAM assumes that stimulus set and response set occur simultaneously, not sequentially. The object is selected and categorized in a single step. The arguments developed earlier about the advantages of simultaneous selection over early or late selection (Equations 4–13) also apply to simultaneous versus sequential stimulus set and response set. ITAM can be configured to select objects and categories either serially or simultaneously, and the simultaneous configuration is faster but no less accurate than the sequential configuration.

ITAM's mechanism for response set was inherited from TVA and GCM. It was presented in Equations 1–4 (i.e., the response set **R** and the βs that it comprises). The mechanism for stimulus set is inherited from TVA. It is built from the same components as the response set mechanism—a subject-independent evidence parameter (η) and a subject-controlled gain parameter (π for priority)—and it involves the same sort of choice process as response set. There is a set **S** of desired stimulus properties and there is a priority parameter π_k for each property k in the stimulus set **S**. The π parameters act as gain controls much like the βs in response set. They multiply η values that represent possible categorizations of stimulus properties of objects in the display. The πs are controlled

completely by the subject and the ηs are determined by the objects in the display and the subjects' history with categorizations of stimulus properties.[1]

An object is chosen by the same mechanism used in response set. The probability of choosing object x with the stimulus set **S** is given by the ratio of the sum of products of η and π that support x to the sum of products of η and π over the whole display:

$$P_\pi(x) = \frac{\sum\limits_{k \in S} \eta(x, k)\pi_k}{\sum\limits_{z \in D}\sum\limits_{k \in S} \eta(z, k)\pi_k}. \tag{14}$$

Equation 14 represents the relative attention weight on object x (Bundesen, 1990). ITAM uses these probabilities to prioritize the objects in the display, multiplying each evidence parameter (η value) in the response set by the probability of selecting the object it refers to in the stimulus set. Thus, processing rate becomes

$$v(x, i) = \eta(x, i)\beta_i P_\pi(x), \tag{15}$$

and the probability of choosing object x and category i becomes

$$P(x \cap i) = \frac{\eta(x, i)\beta_i P_\pi(x)}{\sum\limits_{z \in D}\sum\limits_{j \in R} \eta(z, j)\beta_j P_\pi(z)}. \tag{16}$$

Thus, the greater the likelihood that object x is selected by stimulus set, the greater the likelihood that object x and category i are chosen together.[2]

ITAM assumes that all of these choices are simultaneous. Processing rates defined by Equation 15 can be substituted directly into Equations 4–13 to make exactly the same points about the merits of early, late, and simultaneous selection. Nevertheless, it is instructive to consider how β and π contribute to object selection and to categorization of the selected object. The probability of choosing object x is given by

$$P(x) = \frac{\sum\limits_{j \in R} \eta(x, j)\beta_j P_\pi(x)}{\sum\limits_{z \in D}\sum\limits_{j \in R} \eta(z, j)\beta_j P_\pi(z)}, \tag{17}$$

and the probability of choosing category i given object x is given by

---

[1] Bundesen (1990) interpreted priority learning in terms of changes in π, so that π depends on the stimulus display as well as the subject's homunculus. I prefer to localize learning in the η parameters, so that π depends only on the homunculus.

[2] TVA, CTVA, and ECTVA talk about stimulus set in a different manner, following Bundesen's (1990) original description. Bundesen (1990) defined stimulus set in terms of attention weights, w, assigned to different objects. The weight w_x on object x is defined as $w_x = \sum_{k \in S}\eta(x, k)\pi_k$. When the attention weight on object x enters the TVA equations, it is divided by the sum of the attention weights over all objects in the display, creating a relative attention weight that is equivalent to $P_\pi(x)$. TVA's expression for processing rate is $v(x, i) = \eta(x, i)\beta_i(w_x/\sum_{z \in D}w_z)$, which is equivalent to Equation 15. I prefer to talk about stimulus set in terms of $P_\pi(x)$ because it seems to make the contribution of stimulus set easier to understand.

$$P(i \mid x) = \frac{\eta(x, i)\beta_i P_\pi(x)}{\sum_{j \in R} \eta(x, j)\beta_j P_\pi(x)} = \frac{\eta(x, i)\beta_i}{\sum_{j \in R} \eta(x, j)\beta_j}. \qquad (18)$$

The stimulus set parameter, π, influences the choice of object x, through $P_\pi(x)$, Equation 17, but it does not influence the probability of choosing category i given object x (the far right-hand expression in Equation 18). In this sense, stimulus set is independent of response set. Increasing the likelihood of choosing object x by manipulating stimulus set increases the likelihood of all possible response-set categorizations of object x (Equation 17; also see Bundesen, 1990). Knowing your rental car is red makes it more likely to be chosen (Equation 17), and that increases the likelihood that you will find the matching license plate (Equation 18). However, knowing the car is red has no effect on the likelihood that you will confuse the license number with similar numbers (Equation 18).

The response set parameter, β, influences both the likelihood of choosing object x (Equation 17) and the likelihood of choosing categorization i given object x (Equation 18). Contrary to classical analyses, ITAM allows response set to participate in object selection as well as category selection. Stimulus set and response set both contribute to object selection; response set alone contributes to category selection.

The complementary roles of stimulus set and response set can be seen more clearly by defining the probability $P_\beta(x)$ that object x is selected by response set alone:

$$P_\beta(x) = \frac{\sum_{j \in R} \eta(x, j)\beta_j}{\sum_{z \in D} \sum_{j \in R} \eta(z, j)\beta_j}. \qquad (19)$$

Note that Equation 19 is essentially the same as Equation 6. The probability of selecting object x can be expressed in terms of P_π and P_β by dividing the numerator and denominator of Equation 17 by the denominator of Equation 19:

$$P(x) = \frac{\frac{\sum_{j \in R} \eta(x, j)\beta_j}{\sum_{z \in D} \sum_{j \in R} \eta(z, j)\beta_j} P_\pi(x)}{\sum_{z \in D} \frac{\sum_{j \in R} \eta(z, j)\beta_j}{\sum_{z \in D} \sum_{j \in R} \eta(z, j)\beta_j} P_\pi(z)} = \frac{P_\beta(x) P_\pi(x)}{\sum_{z \in D} P_\beta(z) P_\pi(z)}. \qquad (20)$$

Equation 20 shows that object selection depends symmetrically on stimulus set and response set. The probability of selecting object x increases as $P_\pi(x)$ and $P_\beta(x)$ increase. The sums of products in the denominator of Equation 20 represent the covariation between $P_\pi(z)$ and $P_\beta(z)$ over the display $\mathbf{D}$ (i.e., the tendency for stimulus set and response set to select the same objects). Equation 20 suggests that a correlation between stimulus set and response set decreases the probability of object selection. Stimulus set increases the probability of object selection only when it provides new information that is not already available in response set.

ITAM improves accuracy relative to selection by response set alone because both stimulus set and response set contribute to object selection. The machinery developed here so far allows quantification of the improvement in the accuracy of object selection. The advantage can be construed as a simple multiplier, A, applied to $P_\beta(x)$, so that

$$P(x) = A \cdot P_\beta(x). \qquad (21)$$

If there is no advantage of stimulus set in object selection, then $A = 1.0$. If there is an advantage, $A > 1.0$. It is instructive to solve Equation 21 for A:

$$A = \frac{P(x)}{P_\beta(x)} = \frac{P_\pi(x)}{\sum_{z \in D} P_\pi(z) P_\beta(z)}. \qquad (22)$$

Like Equation 20, Equation 22 suggests that the advantage of TVA over object selection by response set increases as the efficiency of selection by stimulus set increases, that is, as $P_\pi(x)$ increases, and decreases as the correlation between stimulus set and response set increases.

The contribution of stimulus set to overall performance (object and category selection) can be calculated in a similar manner. To do so, we need to introduce another probability, the probability $P_\beta(x \cap i)$ that object x and category i are chosen by response set alone.

$$P_\beta(x \cap i) = \frac{\eta(x, i)\beta_j}{\sum_{z \in D} \sum_{j \in R} \eta(z, j)\beta_j}. \qquad (23)$$

What advantage does ITAM provide over Equation 23? Starting with Equation 16, which defines $P(x \cap i)$, we see that

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

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

$$= \frac{\eta(x, i)\beta_i}{\sum_{z \in D} \sum_{j \in R} \eta(z, j)\beta_j} \cdot \frac{P_\pi(x)}{\sum_{z \in D} P_\beta(z) P_\pi(z)}$$

$$= P_\beta(x \cap i) \cdot A. \qquad (24)$$

The advantage A of stimulus set in selecting objects and categorizations is the same as the advantage of stimulus set in selecting objects (Equation 22). Stimulus set improves overall performance by improving object selection, not category selection. This is another expression of the independence of stimulus set and response set.

Cue Type and Cue Validity

ITAM's distinction between stimulus set and response set provides some insight into two contrasting procedures that are com-

monly used to study attentional cuing. Both procedures involve presenting a display and a cue that indicates the location of the target in the display. In one procedure, most closely associated with Eriksen (e.g., C. W. Eriksen & Hoffman, 1972), each item in the display is a member of the response set **R**, and the cue indicates which item to report. The Eriksen procedure requires the subject to process the cue; otherwise, performance would be at chance. In the other procedure, most closely associated with Posner (e.g., Posner & Cohen, 1984), only one item in the display is a member of the response set **R**, so the subject can choose the correct response without processing the cue. The cue provides additional information about the target's location, and that often benefits performance, but the Posner procedure does not require the subject to process the cue. Examples of Eriksen and Posner cues are presented in Figure 3.

ITAM's analysis of object selection suggests that there is a continuum between Eriksen-type and Posner-type cuing situations that reflects the contribution of response set (see Equation 20). Eriksen-type cues are at one end of the continuum, at which response set contributes nothing to object selection. Each object is an equally valid member of the response set, so the probability of object selection by response set would be at chance (i.e., $P_\beta(z) = 1/N_D$ for all $z \in \mathbf{D}$, where N_D is the number of objects in the display), so $P_\beta(x)$ would cancel out of Equation 20. Object selection would depend entirely on stimulus set, that is, $P(x) = P_\pi(x)$. Posner-type cues form the middle of the continuum, in which $P_\beta(x)$ is above chance, and object selection depends on both stimulus set and response set. At the end of the continuum opposite to Eriksen, $P_\beta(x) = P(x)$, and object selection depends entirely on

Eriksen Cues

⇓

H H S H S S

Posner Cues: Valid

⇓

3 6 9 S 2 8 4

Posner Cues: Invalid

⇓

7 3 H 9 4 2 3

Figure 3. The distinction between Eriksen-type cues and Posner-type cues. With Eriksen cues, response set contributes nothing to object selection. With Posner cues, response set works together with stimulus set in object selection.

response set. Researchers wishing to separate stimulus set and response set effects should prefer the Eriksen procedure. Researchers wishing to study their interaction in object selection should prefer the Posner procedure.

Most studies of Posner-type cuing manipulate cue validity. On some proportion of the trials, the cue is valid in that it indicates the correct location of the target. On the remaining proportion of the trials, the cue is invalid in that it indicates an incorrect location of the target (see Figure 3). The main finding is that performance is often faster and more accurate when the cue is valid than when it is invalid. For example, Posner, Snyder, and Davidson (1980, Experiment 1) presented a cue on the left or right side of fixation indicating the target's position. On 80% of the trials, the cue was valid, and the target appeared on the same side as the cue. RT was fast and accuracy was high. On 20% of the trials, the cue was invalid, and the target appeared on the side opposite the cue. RT was slower and accuracy was lower. Bundesen (1990) modeled these data quite successfully with TVA.

ITAM interprets the cue validity effect in terms of the interaction between stimulus set and response set in object selection. In Posner-type cuing situations, the object can be selected by stimulus set or response set, and cue validity pits one against the other. In the stimulus set, π should be set high for the right and left locations, so a cue in one location (e.g., left) should bias object selection toward that object, that is, a cue on the left should increase $P(left)$. In the response set, β should be set high for the identity of the target and low for the identity of the cue. Thus, object selection by response set should be affected by the target and not by the cue. When the cue is valid, stimulus set and response set are biased toward selecting the same object, so $P(x)$ and $P(x \cap i)$ should both increase. When the cue is invalid, stimulus set and response set are biased toward selecting different objects, so $P(x)$ and $P(x \cap i)$ should both decrease.
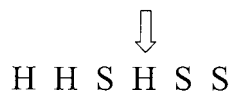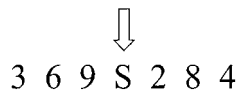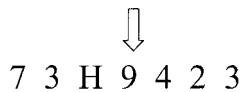
Similarity and Choice

Another fundamental commonality among the members of the family tree in Figure 1 is that they all assume that the choice process is driven by similarity. This commonality is shared with many theories in cognitive psychology. Similarity appears in various guises in numerous contexts, both as an explanation and a thing to be explained (see, e.g., Medin, Goldstone, & Gentner, 1993; Shepard, 1957, 1964; Tversky, 1977). Similarity pervades cognitive theory as much as choice does, so it is no accident that it pervades ITAM's family tree.

The common approach to similarity in the different branches of ITAM's family tree lies in the basic machinery of the choice process. The η values in Equations 2 and 15, which represent the similarity between an object and a category, appear in choice equations of each theory. Similarity is given a more thorough treatment in the memory branch than in the attention branch. In TVA and its relatives, similarity is often a free parameter that is manipulated to optimize fit (e.g., Bundesen, 1987; Shibuya & Bundesen, 1988). The η parameters are treated as "atoms" and not decomposed further. By contrast, in GCM and its relatives, similarity is often an experimental manipulation and the η parameters are broken down into subatomic components that account for a broad variety of data (e.g., Nosofsky, 1984, 1986, 1987, 1988, 1991b; Shin & Nosofsky, 1992).

ITAM adopts GCM's interpretation of similarity in terms of distance in multidimensional similarity space (Shepard, 1957, 1964). Each object is a point in multidimensional similarity space, and the similarity between one object and another (e.g., object x and a member of category i) is an exponential function of the distance between them (Shepard, 1958, 1987):

$$n(x, i) = \exp[-s \cdot d_{xi}], \qquad (25)$$

where s is a sensitivity parameter and d_{xi} is the distance between x and i in similarity space. Distance is defined by

$$d_{xi} = [\sum_{h=1}^{H} w_h |u_{xh} - u_{ih}|^r]^{\frac{1}{r}}. \qquad (26)$$

The parameter r is the Minkowski distance metric. If $r = 2$, distance is Euclidean; if $r = 1$, distance is city-block (see Nosofsky, 1984). The parameter w_h represents the attention weight on dimension h. The attention weights are set by the homunculus (see Logan & Gordon, 2001) but are constrained to sum to 1.0.

Nosofsky and colleagues tested GCM's similarity assumptions quite rigorously (see, e.g., Nosofsky, 1987, 1988, 1991a, 1991b; Shin & Nosofsky, 1992). For example, Nosofsky (1986) generated confusion matrices by requiring subjects to identify briefly presented masked stimuli and then derived multidimensional similarity spaces from the confusion matrices. The distances in this space were used to define similarities following Equations 25 and 26. Subjects then learned to classify the stimuli in several ways, and Nosofsky predicted their classification performance from their similarity spaces. These spaces fixed the distances $|u_{xh} - u_{ih}|$ in Equation 26 so that similarity was not a free parameter. Only the attention weights w and the sensitivity parameter s were allowed to vary. GCM fit the classification data very well.

Representing Stimulus Dimensions

ITAM adopts GCM's assumption that objects can be represented meaningfully as points in multidimensional space, or equivalently, as sets of values of orthogonal dimensions. The assumption gives ITAM a rich representation with considerable power. Part of the flexibility results from the exponent r in the Minkowski distance metric. When $r = 2$, the metric describes similarity between objects that vary on *integral* dimensions (Garner, 1974), like hue and saturation in color or height and width in form, that cannot be perceived independently. When $r = 1$, the metric describes similarity between objects that vary on *separable* dimensions (Garner, 1974), like form and color or shape and texture, that can be perceived independently.

Part of the flexibility results from the variety of dimensional differences that can be represented in Equation 26. It can represent differences in dimensions with continuous values, like tilt or hue (e.g., Nosofsky, 1987), and it can represent differences in discrete dimensions, like closure or type of line junction (T or Y) or presence or absence of features, like curvature in lines or voicing in speech. Typically, stimuli are represented in five dimensions or fewer, but there is no principled constraint on the number of dimensions. The formalism will allow spaces with high dimensionality that are typical of distributed representations in connec-

tionist models and global memory models (e.g., McClelland, Rumelhart, & the PDP research group, 1986; Murdock, 1993).

Attention to Dimensions

TVA and its relatives model attention to categories and attention to objects, but they do not model attention to dimensions, ignoring a whole set of phenomena investigated by Garner and his colleagues (e.g., Garner, 1974; but see Nosofsky & Palmeri, 1997). GCM models attention to dimensions through the attention weights, w_h in Equation 26, and ITAM adopts GCM's model of attention to dimensions. The attention weights are assumed to be controlled by the subject (i.e., the homunculus; see Logan & Gordon, 2001) and adjusted to optimize performance (cf. Kruschke, 1992). Adjusting the attention weights affects performance by "stretching" and "shrinking" multidimensional similarity space. Increasing the weight on a dimension stretches it, increasing the distance between objects on that dimension in multidimensional space. The weights are constrained to add to 1.0, so increasing the weight on one dimension necessarily decreases the weight on the other dimensions; stretching one dimension necessarily shrinks the other ones. The effect, illustrated in Figure 4, is to increase discriminability on the attended dimension and reduce discriminability on the unattended dimensions.

The idea that attention weights are adjusted to optimize performance implies that different weights may be required for different tasks. This insight allowed Nosofsky (1984) to resolve a long-standing paradox in the categorization literature. Shepard, Hov-

Equal Weight to Both Dimensions
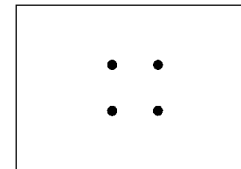
More Weight to Horizontal Dimension
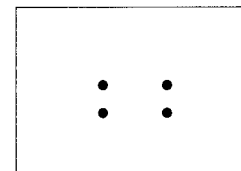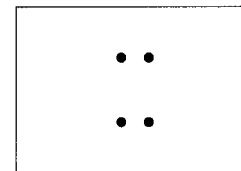
More Weight to Vertical Dimension

Figure 4. A hypothetical two-dimensional similarity space with four points in it. In the top panel, equal weight is given to each dimension; in the middle panel, more weight is given to the horizontal dimension; in the bottom panel, more weight is given to the vertical dimension.

land, and Jenkins (1961) estimated similarities between stimuli in an identification task in which subjects learned a unique response to each stimulus and used those similarities to predict choice probabilities in six classification tasks. The predictions failed dramatically, and the failure replicated several times (e.g., Nosofsky, 1987; Nosofsky, Gluck, Palmeri, McKinley, & Glauthier, 1994). Nosofsky (1984) argued that the optimal division of attention between dimensions is different for identification tasks than for classification tasks and showed that attention weights optimized for identification made the same errors of prediction of classification probabilities that Shepard et al. (1961) observed (also see Nosofsky, 1986, 1987). Nosofsky (1984) showed that GCM can predict classification performance from identification performance if it holds distance constant and lets the attention weights vary between tasks. Furthermore, Nosofsky (1988, 1991b) showed that the relation between episodic recognition ("did this stimulus appear on the training list?") and classification could be accounted for similarly in terms of differences between tasks in the distribution of attention weights to dimensions.

Similarity Is an Exponential Function of Distance

Exponential generalization gradient. The assumption that similarity is an exponential function of distance in multidimensional space (i.e., Equation 25) has several important implications. First, it implies that generalization or discrimination is an exponential function of distance. This implication is consistent with Shepard's (1987) claim that an exponential distance function is a universal law of generalization supported by a wide variety of empirical generalization gradients (e.g., Shepard, 1958, 1987). Shepard (1987) showed mathematically that exponential generalization gradients result from very broad and general assumptions, so they should be ubiquitous in nature.

It may be helpful to think of the exponential generalization gradient as a receptive field that is "tuned" to a target object in multidimensional space like a cortical neuron is tuned to a particular value of some dimension (e.g., orientation; see Shepard, 1987). ITAM interprets distance as similarity and similarity as processing rate (Equation 15), so the analogy is direct. Objects close to the target activate it strongly; objects farther from the target activate it less strongly. The sharpness of the tuning depends on the sensitivity parameter, s (see Equation 25). Large values of s produce steep generalization gradients and narrow receptive fields; small values produce shallow generalization gradients and broad receptive fields.

Interactive similarity. The exponential distance function allows a kind of *interactive similarity* that was a central concept in the mother of all models in the memory branch, Medin and Schaffer's (1978) context model of classification. Interactive similarity allows the context model to represent the particular combinations of features that subjects experienced in classification tasks and it allows combinations of features to influence performance. The alternative to interactive similarity was *additive similarity* (Franks & Bransford, 1971; Hayes-Roth & Hayes-Roth, 1977; Reed, 1972), in which different features contribute independently to similarity between objects. Medin and Schaffer showed that interactive similarity provided a better account of classification learning and transfer than additive similarity because it explained the effects of similar exemplars (i.e., objects sharing combinations of features) on performance.

Medin and Schaffer (1978) adopted a multiplicative version of interactive similarity, arguing that the similarity between objects was the product of the similarities between their features. That is,

$$\Psi(i, j) = \prod_{h=1}^{H} \psi_h(i, j), \tag{27}$$

where $\Psi(i, j)$ is the overall similarity between objects i and j and $\psi_h(i, j)$ is the dimensional similarity between objects i and j on dimension h, which varies from 0 to 1, with 1 representing identity. This multiplicative similarity emphasizes combinations of features because objects have to match on all features in order to have a high degree of similarity. A single mismatching feature with a dimensional similarity near zero will reduce the overall similarity to near zero. Partial matches do not contribute very much to overall similarity.

Nosofsky (1984) generalized the context model by assuming that the dimensional similarities were exponential functions of distance (i.e., $\psi_h(i, j) = \exp[-s \cdot d_{hij}]$, where d_{hij} is the distance between objects i and j on dimension h) in a multidimensional similarity space with a city-block distance metric (i.e., $r = 1$ in Equation 26). Thus,

$$\Psi(i, j) = \prod_{h=1}^{H} \exp[-s \cdot d_{hij}] = \exp\left[-s \cdot \sum_{h=1}^{H} d_{hij}\right] = \eta(i, j). \tag{28}$$

Under these assumptions, the overall similarities in the context model become the ηs in GCM and ITAM. The exponential distance function transforms additive distances in similarity space into multiplicative similarities. It allows ITAM the benefits of interactive similarity that were pointed out by Medin and Schaffer (1978).

Why stimulus set and response set multiply. The exponential distance function explains a puzzling aspect of TVA: why stimulus set and response set multiply (see Equations 15 and 16). Bundesen (personal communication, January 2000) chose multiplication for mathematical reasons: It was the only way to combine stimulus set and response set independently (see Equations 17 and 18). However, if GCM's assumption is adopted that similarity is an exponential function of distance in multidimensional space, then stimulus set and response set have to multiply. With the exponential distance function, TVA's combination of stimulus set and response set becomes another example of multiplicative similarity. Stimulus set adds another dimension on which the target differs from the distractors. The added dimension adds to the distance between them in multidimensional similarity space and decreases the similarity between them, multiplying it by a value less than 1 (see Equation 25). The effect can be seen in Figure 5, in which two identical stimuli are presented in different display locations. When these stimuli are projected onto similarity space, they overlap almost entirely. There is no way to choose between them. Adding a difference between the two objects on a dimension other than identity (i.e., color) separates the two objects in similarity space and provides a basis for choosing between them beyond their identities (see Figure 5).
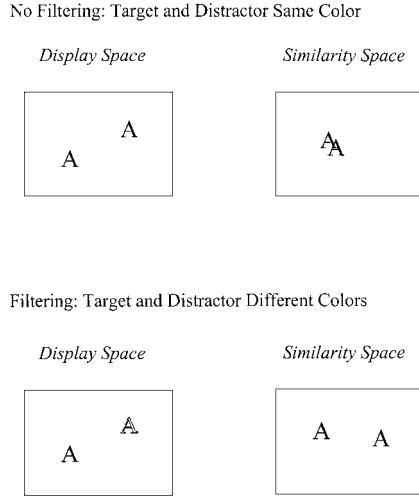
No Filtering: Target and Distractor Same Color

Display Space   *Similarity Space*

Filtering: Target and Distractor Different Colors

Display Space   *Similarity Space*

Figure 5. Interpretation of filtering in terms of similarity. The top panel shows two letters in a display (display space) on the left and the same two letters projected onto similarity space on the right. The letters are identical, so there is no basis for distinguishing them in similarity space. The bottom panel shows two letters in a display when one has been colored differently from the other (left side). The bottom right panel shows the same two letters projected onto similarity space. The difference in color separates them in similarity space and provides a basis for distinguishing them.

In terms of the model, adding a dimension of difference increases the dimensionality of the similarity space and changes the definition of the target. The similarity space becomes at least two-dimensional, with one dimension for stimulus set and at least one dimension for response set, and the target becomes a multidimensional object in that space. The discriminability of the target from the distractors depends on $\eta(x, M)$, which is the similarity between the object x and a multidimensional representation of the target M that reflects the point it occupies in multidimensional similarity space. Stimulus set and response set address separable dimensions (Garner, 1974), so $\eta(x, M)$ is an exponential function of city-block distance ($r = 1$ in Equation 26).

Consider the case in which there is only one category k in the stimulus set and one category i in the response set so the target M is the point i, k in similarity space. The distance d_{xM} from object x to the representation of M is the sum of the distance d_{xi} between x and the target value i on the response set dimension and the distance d_{xk} between x and the target value k for the stimulus set dimension (see Equation 26). Thus, from Equation 25,

$$\eta(x, M) = \exp[-s \cdot d_{xM}]$$
$$= \exp[-s \cdot (d_{xi} + d_{xk})]$$
$$= \exp[-s \cdot d_{xi}] \cdot \exp[-s \cdot d_{xk}]$$
$$= \eta(x, i)\eta(x, k). \qquad (29)$$

Equation 29 suggests that TVA's combination stimulus set and response set results in a new representation defined by the product of η values. This product is a central part of the basic TVA equation for processing rate (Equation 15). If $\beta_i = \pi_k = 1$, then from Equation 15, the expression for processing rate,

$$v(x, i) = \eta(x, i)\beta_i P_\pi(x)$$

$$= \eta(x, i)\beta_i\eta(x, k)\pi_k \frac{1}{\sum_{z \in D} \eta(z, k)\pi_k}$$

$$= \eta(x, i)\eta(x, k)\frac{1}{\sum_{z \in D} \eta(z, k)},$$

contains the product of η values in Equation 29. Thus, the multiplication of stimulus set and response set in TVA is no mystery in ITAM; it is a natural consequence of the exponential distance function inherited from the memory branch.

Spatial Representation of Similarity

The assumption that similarities can be represented as distances in multidimensional space is powerful but controversial. Some researchers point out the limitations of the representation, arguing that it does not capture hierarchical or relational information very well (Medin et al., 1993). Other researchers challenge the metric assumptions underlying the spatial representation (e.g., Tversky, 1977; Tversky & Gati, 1982; Tversky & Hutchinson, 1986). The problems seem more severe with complex, abstract concepts than with the simple perceptual classifications that the attention literature and GCM address, so the challenges may be less critical than they seem.

The assumption that has been challenged most strenuously is the assumption of symmetry: The distance between i and j should be the same as the distance between j and i and so should the similarity. However, similarity judgments are often asymmetrical; subjects say Korea is more similar to China than China is to Korea (Tversky, 1977). Confusions in identification tasks are often asymmetrical; subjects will confuse E for F more often than they confuse F for E (Garner & Haun, 1978). Important asymmetries appear in visual search tasks; subjects find E among Fs more easily than they find F among Es (Treisman & Gormican, 1988; Treisman & Souther, 1985). These effects seem to challenge the assumption that similarity and distance are symmetrical.

Nosofsky (1991a) showed that many of the challenges to symmetrical similarity in identification and classification tasks could be accommodated within GCM's assumption of symmetric similarity by allowing differential bias. If the bias for F were greater than the bias for E, for example, subjects would be more likely to confuse E for F than F for E. Nosofsky (1991a) investigated a generalization of GCM in which there were stimulus biases as well as response biases. He showed that this model was equivalent to an additive version of Tversky's (1977) contrast model that accounts for a variety of asymmetries. Nosofsky (1991a) suggested that stimulus bias could reflect things like salience or familiarity that are independent of the response given to the stimulus. Indeed, TVA's relative attention weight, $P_\pi(x)$, can be viewed as a stimulus bias parameter. An important challenge for future research will be to see whether the asymmetries in visual search reported by Treisman and Souther (1985) and Treisman and Gormican (1988) can also be accounted for by symmetrical similarity and asymmetrical stimulus or response biases (Nosofsky, 1991a). It would be especially challenging to predict asymmetries in search tasks from

asymmetries in identification tasks (see Nosofsky, 1984, 1986, 1987).

Similarity and Visual Search

The problem of finding your car in a crowded parking lot has been studied extensively by attention researchers in the guise of *visual search* tasks. Subjects are shown displays that contain a target and several distractors or just several distractors and their task is to say whether or not the display contains a target. Displays are often exposed until subjects respond, so the main dependent variable is RT rather than accuracy. RT increases linearly with the number of objects in the display (*display size*) for both target-present and target-absent responses, and researchers interpret the slope of the linear function relating RT to display size as a measure of *search efficiency*. Efficient searches produce shallow slopes that range from 0 to 10 ms per object, whereas inefficient searches produce steep slopes that are often larger than 25 ms per object.

A major empirical goal of attention research has been to characterize the conditions under which search is efficient and inefficient (but see Wolfe, 1997). Two main trends have emerged in the last 20 years. One focuses on the role of similarity between targets and distractors and between the distractors themselves. Search is more efficient when targets and distractors are dissimilar and when distractors are similar to each other (Duncan & Humphreys, 1989). The other trend focuses on the nature of the decision required to find a target. Treisman and colleagues contrasted *feature search* with *conjunction search*, finding that search was more efficient when the target differed from the distractors by a single feature than when the target differed by a conjunction of features (Treisman & Gelade, 1980; Treisman & Schmidt, 1982). A major theoretical goal in attention research is to explain why similarity and decision type determine the efficiency of search.

ITAM is concerned with the role of similarity in the interaction between top-down and bottom-up processes, when a cognitive representation of the target meets a perceptual representation of the items in the display. Several researchers have investigated bottom-up similarity effects based on local differences between items that produce texture segregation and efficient search (e.g., Beck, 1982; Julesz & Bergen, 1983; Nothdurft, 1992). Effects like these are interesting and important, but they are dissociable from the kinds of search effects ITAM addresses (Wolfe, 1992; also see Bacon & Egeth, 1991; Duncan, 1989). They are beyond the scope of the current version of ITAM.

ITAM and target–distractor similarity. Search is more efficient the less similar the targets are to the distractors. This generalization holds for several kinds of search tasks, including feature search (Duncan, 1989), conjunction search (Wolfe et al., 1989), and tasks in which the targets are and are not linearly separable from the distractors (Bauer, Jolicoeur, & Cowan, 1996). ITAM accounts for these effects in terms of the exponential relation between similarity and distance in multidimensional similarity space. Distractors that are close to the target's location in similarity space are likely to be confused with the target and chosen instead of it (see Equations 4 and 16). Distractors that are farther from the target are less likely to be confused or chosen. Confusions make search inefficient.

ITAM and distractor–distractor similarity. Search efficiency depends on the similarity of the distractors to each other. Several

investigators have found more efficient search with displays like the one in the top left panel of Figure 6, in which the distractors are homogeneous distractors, than with displays like the one in the bottom left panel of Figure 6, in which the distractors are heterogeneous (Estes, 1974; Farmer & Taylor, 1980; Humphreys & Müller, 1993). Duncan and Humphreys (1989) interpreted these results in terms of grouping by similarity, arguing that similar distractors are grouped together and rejected as a group. The more homogeneous the display, the fewer groups there are to be rejected, so search ends sooner (also see Humphreys & Müller, 1993).

ITAM accounts for the distractor heterogeneity in terms of the attention weights, w, inherited from GCM (see Equation 26), without invoking similarity-based perceptual grouping processes. Homogeneous distractors cluster together in similarity space and so can be separated from the target by assigning a large attention weight to one of the dimensions. The large attention weight stretches the dimension and increases the distance between the target and the distractors, and that improves performance. This effect is illustrated in the top panels of Figure 6. The top left panel shows a display from a homogeneous distractor condition in which the task is to find an *A* among *B*s. The top middle panel shows the display projected onto similarity space with equal attention weights to the horizontal and vertical dimensions. The top right panel shows the effect of stretching the horizontal dimension, which is to increase the separation between the *A* and the cluster of *B*s.

Heterogeneous distractors may not allow such an advantageous distribution of attention. Heterogeneous distractors are spread out further in similarity space and may overlap with the target on different dimensions. Stretching one dimension may move some of
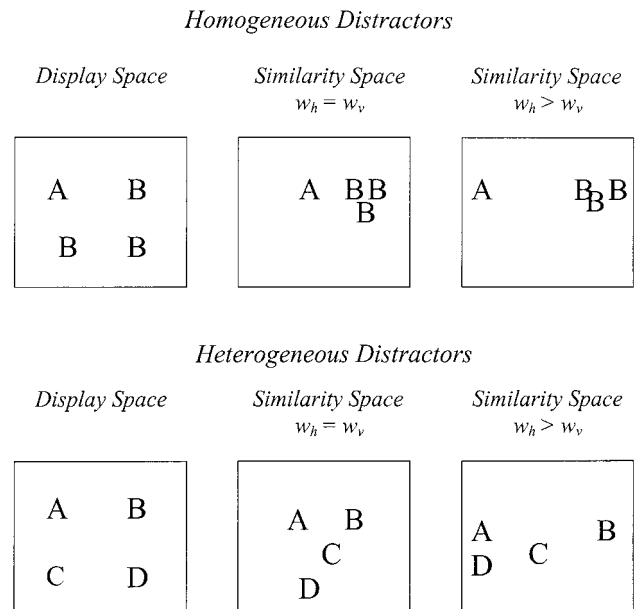
Homogeneous Distractors

Figure 6. Search for a target (A) among homogeneous (top panels) and heterogeneous (bottom panels) displays. The left side shows the displays in display space. The right side shows the displays in similarity space. Homogeneous distractors cluster together in similarity space. Heterogeneous distractors do not.

the distractors away from the target, but it will shrink the other dimensions, and that may bring other distractors closer to the target. Target–distractor distance would remain the same, resulting in no net gain in discriminability. This is illustrated in the bottom panel of Figure 6. The bottom left panel shows a display from a heterogeneous distractor condition in which the task is to find an *A* among *B*s, *C*s, and *D*s. The bottom middle panel shows the display projected onto similarity space with equal attention weights, and the bottom right panel shows the effects of stretching the horizontal dimension and, consequently, shrinking the vertical dimension.

Researchers have found distractor similarity effects when homogeneous and heterogeneous displays are mixed randomly in a block of trials (Humphreys & Müller, 1993; Treisman, 1988, 1991). Subjects are faster at finding an *A* among *B*s or an *A* among *C*s than at finding an *A* among *B*s and *C*s, even when the *A*-among-*B* displays are mixed randomly with the *A*-among-*C* displays. This finding presents a problem for ITAM's attention weight account of distractor similarity effects because ITAM has to know which distractor is present in order to allocate attention weight to the optimal dimension(s).

One solution to this problem is to allocate attention weights equally to all dimensions until the first distractor is categorized and then to reallocate attention weight to those dimensions that distinguish the target from that distractor. The initial categorization with equal attention weights would increase RT relative to conditions in which the distractor was known in advance, but that is consistent with the data. Subjects can be as much as 100 ms slower if the target dimension is not specified in advance (Treisman, 1988, 1991).

The first solution uses machinery that is part of the current version of ITAM. A second solution would be to adopt a method developed by Kruschke and Johansen (1999) that changes attention weights during the course of a single trial. Their RASHNL model uses a gradient descent mechanism to change attention weights on a time scale that may be short enough to be useful in search tasks. Predictions of RT cannot be derived directly from their model, but it offers a promising solution to the problem of finding the dimensions that distinguish the target from the distractors.

Feature search and conjunction search.   Many investigations of the determinants of search efficiency have contrasted feature search and conjunction search (e.g., Duncan & Humphreys, 1989; Treisman & Gelade, 1980; Wolfe, 1994; Wolfe et al., 1989). In feature search, subjects look for a target that has a unique feature that is not shared by any distractor. For example, subjects might look for a *G* among *C*s or a red object among green ones. In conjunction search, subjects look for targets among distractors that share features, and no feature specifies the target uniquely. Targets are defined by a combination of features from two or more separable dimensions, and different distractors will share each of these features with the target. For example, subjects might search for a red *G* in a display of red *C*s and green *G*s. Detecting a single feature is not sufficient to indicate a target is present. Instead, a combination of features must be detected. Feature search is often easy, with slopes ranging from 0 to 10 ms per item, whereas conjunction search is often hard, with slopes of 40 ms per item or more. Feature search is not always easy (Treisman & Gormican,

1988), and conjunction search is not always hard (Wolfe et al., 1989).

The major theories of visual search offer alternative explanations of the difference between feature search and conjunction search. Treisman's *feature integration theory* assumes that single features can be detected preattentively but that attention is required to conjoin separate features. Thus, feature search is parallel and efficient, whereas conjunction search is serial and inefficient (e.g., Treisman & Gelade, 1980). At the other extreme, Duncan and Humphreys (1989) argued that there is no fundamental difference between feature search and conjunction search. Search depends only on target–distractor and distractor–distractor similarity, and these are different in feature and conjunction search. Feature search tends to involve low target–distractor similarity and high distractor–distractor similarity, both of which foster efficient search, whereas conjunction search tends to involve high target–distractor similarity and low distractor–distractor similarity, both of which foster inefficient search (also see Humphreys & Müller, 1993). Wolfe and colleagues proposed an intermediate position, called *guided search*, in which a memorial representation of the target has a top-down influence on preattentive processes that can direct attention to the target when the target is sufficiently different from the distractors (Cave & Wolfe, 1990; Wolfe, 1994; Wolfe et al., 1989).

ITAM can do conjunction search in three different ways. Consider the example of searching for a red *G* in a display of red *C*s and green *G*s. The most natural way to do this search from the perspective of TVA and GCM is to represent the conjunction target as a single multidimensional object, using one β value to represent the target itself (i.e., "red *G*") and different attention weights to represent the relevant dimensions (i.e., w_G and w_{red} would be set high). Feature targets would be represented similarly, with a single β value to represent the target and an attention weight to represent the relevant dimension (e.g., β_{red} and w_{red} would be set high). Conjunction search and feature search would be performed in the same way, both driven by target–distractor and distractor–distractor similarities, as Duncan and Humphreys (1989) suggested. According to the analyses in the section on filtering, the β that represents the target would participate in object selection—directing attention to the target—as well as target detection, so this version would resemble guided search as well (see Wolfe, 1994).

Feature search is still different from conjunction search in this version of ITAM. In feature searches in which the target differs from all the distractors in a single feature, all of the attention weight can be assigned to the dimensions that distinguish the target from the distractors, stretching the dimension and improving discriminability. In conjunction searches, stretching one dimension shrinks the others, so the improvement in discriminability in one dimension is bought at the cost of reduced discriminability in another. This difference in the effects of attention weights on feature and conjunction search is illustrated in Figure 7.

Alternatively, conjunction targets could be represented as logical conjunctions of elementary categorizations with one β for each feature. Thus, searching for a red *G* would involve setting β_{red} and β_G high and responding "target present" if both categorizations "*x* is red" and "*x* is *G*" came out of the choice process. This method of searching would require the system to discriminate cases in which *red* and *G* came from the same object from cases in which
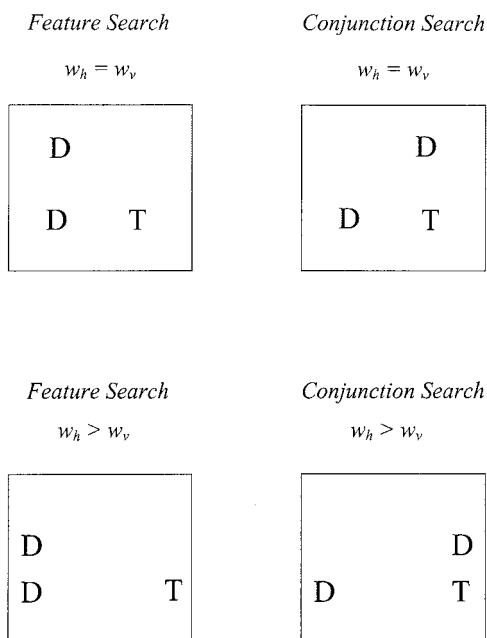
Feature Search

$w_h = w_v$

```
┌─────────────────────┐
│                     │
│   D                 │
│                     │
│      D      T       │
│                     │
└─────────────────────┘
```

Conjunction Search

$w_h = w_v$

```
┌─────────────────────┐
│                     │
│              D      │
│                     │
│      D      T       │
│                     │
└─────────────────────┘
```

Feature Search

$w_h > w_v$

```
┌─────────────────────┐
│                     │
│   D                 │
│   D         T       │
│                     │
└─────────────────────┘
```

Conjunction Search

$w_h > w_v$

```
┌─────────────────────┐
│                     │
│              D      │
│   D         T       │
│                     │
└─────────────────────┘
```

Figure 7. Similarity spaces for feature (left side) and conjunction (right side) search. In the top panels, equal attention weight is given to vertical and horizontal dimensions. In the bottom panels, more attention weight is given to the horizontal dimension. This separates distractors from noise in feature search (bottom left) but brings one distractor closer as it moves the other farther away in conjunction search (bottom right).

red and *G* came from different objects. This discrimination is known as the *binding problem* (Hummel & Biederman, 1992; Pylyshyn, 1989; Ullman, 1984), and feature integration theory was originally proposed as a solution to it (Treisman & Gelade, 1980). ITAM could solve the binding problem by searching through the display serially, just as feature integration theory did, using the π parameter to select different objects in the display (Logan & Gordon, 2001). Again, β_{red} and β_G would participate in object selection, so ITAM would resemble guided search as well as feature integration theory. In this version, feature search would be different from conjunction search. It would require a simpler description of the target (only one β for a single feature target), and it would allow parallel processing.

The third alternative is to distribute the target representation across stimulus set and response set, searching for a *G* among the red items, for example. Thus, π_{red} would be set high to select the red items and β_G would be set high to select *G*. Several studies suggest that human subjects can restrict their attention to one of the features of a conjunction target and search for the other feature(s) in the attended subset (Bacon & Egeth, 1997; Egeth, Virzi, & Garbart, 1984; Kaptein, Theeuwes, & Van der Heijden, 1995; Zohary & Hochstein, 1989). This kind of search represents conjunctions as logical conjunctions, like the previous method, but it is riskier than the previous method because it only requires explicit categorization of one feature (*G* in the example). In ITAM, only categorizations addressed by response set (β) are represented explicitly in working memory. Categorizations addressed by stimulus set (π) are not represented explicitly. Thus, the subject would know that a *G* was present but would have to infer that it was red.

Further research will be required to evaluate the ITAM approaches to conjunction search. The elementary processes must be combined in some way to reach a decision about target presence and target absence, and several combinations are possible. The advantage of framing theoretical alternatives in a common formal language—ITAM—lies in the ability to isolate crucial processes and keep other, less interesting processes the same.

Instance Representation

A fundamental similarity among the members of the memory branch of ITAM's family tree is that they all assume *instance representation*. Each encounter with an object is represented separately as an individual example, or *instance*. It is encoded into memory separately, stored separately, and retrieved separately. Instance representation applies to episodic memory as well as semantic memory. The idea of separate traces for different events accounts naturally for many episodic memory phenomena (Gillund & Shiffrin, 1984; Hintzman, 1988; Jacoby & Brooks, 1984; Raaijmakers & Shiffrin, 1981). The idea of separate traces for different members of the same category seems less intuitive to the uninitiated; nevertheless, it accounts for many phenomena in classification and categorization (Hintzman, 1986; Jacoby & Brooks, 1984; Medin & Schaffer, 1978; Nosofsky, 1988). ITAM adopts the instance representation assumption and inherits the theoretical victories won by its ancestors in the memory branch.

The ancestors in the attention branch contrast with those in the memory branch in that they assume *prototype representation*. Categories are represented as single entities, or *prototypes*, that represent the average or most typical values of the objects' features. Individual experiences with the objects are not represented in the prototype, nor are features that distinguish one instance from another. In my view, prototype representation is an unfortunate assumption because prototype theories have not done well when they were pitted against instance theories (e.g., Medin & Schaffer, 1978). However, prototype representation does not seem to be fundamental to the attention branch in the same way instance representation is fundamental to the memory branch. The prototype representation assumption could be replaced by the instance representation assumption without losing any predictive power. Consequently, ITAM adopts instance representation and eschews prototype representation.

ITAM changes the interpretation of η from the similarity of an object to a prototype, as in TVA, to the similarity of an object to an instance in memory, as in GCM. In many cases, there will be several similar instances in memory. Each of these instances leads to the same categorization, so the instances are interchangeable. The probability of categorizing x as a member of category i is the sum of the probabilities of categorizing x as each of N_i instances of category i. Thus, Equation 4 becomes

$$P(x \cap i) = \frac{v(x, i_1)}{\sum\limits_{z \in D} \sum\limits_{j \in R} v(z, j)} + \frac{v(x, i_2)}{\sum\limits_{z \in D} \sum\limits_{j \in R} v(z, j)} + \cdots + \frac{v(x, i_N)}{\sum\limits_{z \in D} \sum\limits_{j \in R} v(z, j)}$$

$$= \frac{\sum\limits_{m=1}^{N_i} v(x, i_m)}{\sum\limits_{z \in D} \sum\limits_{j \in R} v(z, j)}, \qquad (30)$$

which is the ratio of the sum of processing rates for the instances of category i, designated with subscripts, to the sum of the processing rates for all categorizations of all objects in the display. The sum of processing rates can be reduced further:

$$\frac{\sum_{m=1}^{N_i} v(x, i_m)}{\sum_{z \in D} \sum_{j \in R} v(z, j)} = \frac{\sum_{m=1}^{N_i} \eta(x, i_m)\beta_i P_\pi(x)}{\sum_{z \in D} \sum_{j \in R} v(z, j)} = \frac{(\sum_{m=1}^{N_i} \eta(x, i_m))\beta_i P_\pi(x)}{\sum_{z \in D} \sum_{j \in R} v(z, j)}.$$

The rightmost term suggests that ITAM equations can be constructed directly from TVA equations by replacing TVA's η values for prototypes with ITAM's sums of η values for instances. That is,

$$\eta(x, i)_{\text{TVA}} \Rightarrow \sum_{m=1}^{N_i} \eta(x, i_m)_{\text{ITAM}}. \tag{31}$$

Thus, Equation 19, which defines the processing rate, becomes

$$v(x, i) = (\sum_{m=1}^{N_i} \eta(x, i_m))\beta_i P_\pi(x). \tag{32}$$

These equations can be used to generate choice probabilities, with Equation 16, and categorization times, with Equation 5. The substitution in Equation 31 applies to ηs for stimulus set as well as to ηs for response set. In ITAM, all knowledge is represented as collections of instances.

A major legacy from GCM is the ability to use a single set of instances to support performance in identification, classification, and recognition memory tasks (Nosofsky, 1984, 1986, 1987, 1988). The tasks differ primarily in terms of the biases involved, the objects whose similarities are summed together in generating choice probabilities, and the attention weights given to the dimensions of the objects in similarity space. Consider the situation depicted in Figure 8, in which a subject has seen six members of Category A and six members of Category B and is then presented with a probe, 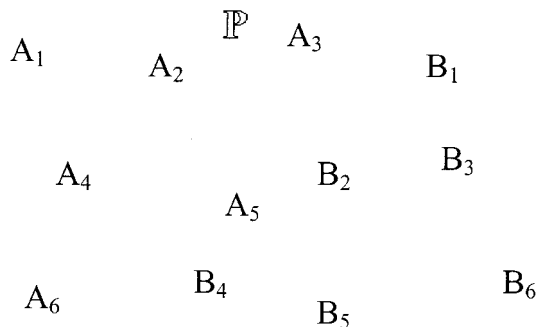P. If the task is item identification—to identify P as one of the 12 items seen in training—there is a bias for each item (e.g., β_{A1}, β_{A2}, β_{A3}, etc.), and the numerator of the choice equation contains a single processing rate, the product of the bias and the similarity between P and the item (i.e., $\eta(P, A_1)\beta_{A1}$; see Equation 4). The subject would be advised to give more weight to the vertical dimension, to increase the separation between P and the incorrect alternatives (e.g., A_1, A_2, A_4, and A_5; B_1, B_2, and B_3).

If the task is categorization—deciding whether P is a member of Category A or Category B—there is one bias for each category (i.e., β_A and β_B) and the numerator of the choice equation is the sum of similarities over all of the objects in one of the categories (e.g., A_1–A_6 for Category A; see Equation 30). The subject would be advised to give more weight to the horizontal dimension to increase the separation between the categories.

If the task is recognition memory—deciding whether or not P was one of the training exemplars—then all items would receive the same bias and similarity would be summed over all the items in both categories (Nosofsky, 1988). Attention weight should be divided equally between dimensions. Summed similarity would be greater for items that were presented before (and so are part of the representation) than for items that were not presented. The recognition judgment would be made by applying a signal-detection–type criterion to the summed similarity (Gillund & Shiffrin, 1984; Raaijmakers & Shiffrin, 1981).

The instance representation assumption is powerful and simple. Different tasks can be performed on instance representations by varying pieces of machinery—biases and attention weights—that are already part of GCM. Every theory of categorization and memory requires the ability to represent individual items. Instance theories require nothing more; they take a "minimalist" approach to modeling cognition. By contrast, prototype theories require some additional mechanisms to extract the prototype and update it from trial to trial. An attention theorist shopping for a theory of categorization may prefer instance theories because they do not require these extra processes. ITAM accounts for categorization and memory phenomena using little more than the mechanisms that are already part of TVA.

Learning

The theories in the attention branch of the family tree live in a universe of eternal "now." They address immediate performance in a stable environment with a constant history. Nothing changes from trial to trial and each trial is the same as the last. You are always in the same parking lot looking for the same car. The theories in the memory branch add a past and a future to this eternal present, accounting for changes over time in a variety of learning and memory paradigms. Together, they provide three ways to model learning that can be used separately or in combination: accumulation of instances, optimizing attention weights, and increasing sensitivity.

Learning as the Accumulation of Instances

The first method of learning is a direct consequence of the instance representation assumption. Each encounter with an object creates an instance representation that is stored in memory. As time passes and more encounters occur, more instances are added to memory, building up a knowledge base. The instance represen-

Figure 8. A similarity space containing instances of two categories (A and B) and a probe, P. If the task is to identify P, attention weight should be allocated to the vertical dimension rather than the horizontal dimension. If the task is to categorize P, attention weight should be allocated to the horizontal dimension rather than the vertical dimension. If the task is episodic recognition, attention weight should be allocated to both dimensions.

tation assumption applies to cases in which the same object is encountered again as well as to cases in which different objects are encountered. Different encounters with the same object result in different instances. Equations 30 and 32 apply to repetitions of single items as well as to presentations of different category members. Accumulation of instances affects RT and choice probability.

Power law of practice. Much of the research on automatization and skill learning in the last 20 years has focused on the nature of the ubiquitous speedup in RT with practice. Throughout that time, the most popular description of the learning curve was in terms of a power function,

$$RT = a + b \cdot N^{-c}, \tag{33}$$

where a is the asymptotic RT, b is the maximum amount by which RT can be reduced by practice, N is the number of practice trials, and c is the learning rate. The greater the absolute value of c, the sharper the inflection in the learning curve.

In 1981, Newell and Rosenbloom declared the power function to be the fundamental law of learning, and since then, many researchers have promulgated their law (e.g., Anderson, 1982, 1992; Cohen, Dunbar, & McClelland, 1990; MacKay, 1982; Schneider, 1985). A major contribution of the instance theory of automaticity was to predict power-function learning from a race between instances. The more practice, the more instances there are stored in memory. A familiar stimulus retrieves all the traces associated with it throughout practice, and the first trace to be retrieved determines performance. Mathematically, it can be shown that, under general conditions, the finishing time for a race between N runners decreases as a power function of N (Logan, 1988, 1992; also see Colonius, 1995; Cousineau, Goodman, & Shiffrin, in press; Logan, 1995b).[3] Moreover, instance theory predicted a power function reduction in the entire RT distribution, whereas previous theories addressed only mean RT. The power law for distributions has been tested and confirmed in some cases (e.g., Anderson, 1992; Cohen et al., 1990; Logan, 1988, 1992; Logan & Etherton, 1994) and disconfirmed in others (e.g., Delaney, Reder, Straszewski, & Ritter, 1998; Rickard, 1997). The power law itself has been challenged. Heathcote, Brown, and Mewhort (2000) argued that it is an artifact of averaging item-level exponential-function learning curves (but see Myung, Kim, & Pitt, 2000).

ITAM places less emphasis on the power law than instance theory did. ITAM assumes a random-walk response selection process, which cumulates the results of several individual races. RT is the sum of the durations of those races plus some "book-keeping" time required for the random walk, and the combination of effects makes it difficult to predict the effect of practice on RT mathematically. Simulations show good power function fits, however (Nosofsky & Palmeri, 1997; Palmeri, 1997).

ITAM does predict that the mean finishing time for a single run of the race will decrease with the number of instances. This effect can be seen clearly in ITAM's expression for finishing time:

$$FT = \frac{1}{\sum_{j \in R} \sum_{m=1}^{N_j} v(x, j_m)} \approx \frac{1}{\sum_{j \in R} N_j v(x, j)} \approx \frac{1}{N} \frac{1}{\sum_{j \in R} v(x, j)}.$$

If the processing rates for all categorizations are the same, and there are N instances in memory for each category, then RT decreases as a power function of N with an exponent $c = 1$.

Accuracy and number of instances. Choice probability also depends on the number of instances in memory, but the effects are less dramatic than the effects on RT. Consider the choice between two categories, A and B, after N_A trials with identical instances of Category A and N_B trials with identical instances of Category B. The probability of choosing Category A for object x is given by

$$P(\text{"}x \text{ is A"}) = \frac{N_A v(x, A)}{N_A v(x, A) + N_B v(x, B)}.$$

If the number of instances in the two categories is the same (so $N_A = N_B$), choice probability is not affected by the number of instances in memory. The probability of choosing Category A for object x would remain the same throughout practice.

To account for changes in choice probability with practice, Nosofsky and Alphonso-Reese (1999) introduced the idea of *background elements* that reflect preexperimental associations. The background elements tend to be retrieved along with instances from the experiment and add a chance component to choice probabilities. From this perspective, the expression for choice probabilities becomes

$$P(\text{"}x \text{ is A"}) = \frac{[N_A \eta(x, A) + b]\beta_A}{[N_A \eta(x, A) + b]\beta_A + [N_B \eta(x, B) + b]\beta_B},$$

where b is a constant representing the number of background elements associated with x. In this expression, the probability of choosing Category A for object x increases with practice because N_A increases with practice and b does not.

Optimizing Attention Weights

Nosofsky (1984, 1986, 1987) assumed that attention weights were assigned to dimensions to optimize classification accuracy. This assumption allows GCM to adopt different attention weights for different tasks (e.g., identification and classification; recognition and classification), and that allows GCM to account for performance in one task in terms of similarities derived from another. Krushke (1992) extended GCM by assuming that optimal attention weights are learned over the course of training. Nosofsky,

---

[3] The distribution function $F_1(x)$ for the minimum of N samples from the same distribution, $F(x)$, is $F_1(x) = 1 - [1 - F(x)]^N$. The distribution function for the Weibull is $F(x) = 1 - \exp[-a \cdot x^c]$. Note that the Weibull distribution is a generalization of the exponential distribution with the variable x raised to a power c. Put differently, the exponential distribution is a special case of the Weibull distribution with $c = 1$. Substituting the distribution function for the Weibull into the distribution function for minima yields $F_1(x) = 1 - (\exp[-a \cdot x^c])^N = 1 - \exp[N \cdot (-a \cdot x^c)] = 1 - \exp[-a \cdot (N^{1/c}x)^c] = N^{-1/c}F(x)$. Thus, the distribution of minima sampled from a Weibull distribution is itself a Weibull distribution with its scale reduced by a factor of $N^{1/c}$. The distribution of minima decreases as a power function of N with an exponent $1/c$ that is the reciprocal of the exponent c of the parent Weibull distribution. This relation leads to the prediction that the shape of the retrieval time distribution (i.e., c) determines the shape of the learning curve (i.e., the exponent $1/c$ of the power-function speedup). I tested this prediction in several data sets and confirmed it in most of them (Logan, 1992).

Palmeri, and colleagues adopted Krushke's assumption in recent applications of GCM to classification learning (Nosofsky et al., 1994; Nosofsky & Palmeri, 1996). ITAM could profit by adopting similar assumptions.

Increasing Sensitivity

Nosofsky (1988) allowed the sensitivity parameter (s in Equation 25) to increase over practice and found that increasing sensitivity provided a better fit to learning data than constant sensitivity. The effect of increasing the sensitivity parameter is to increase psychological distance between nonmatching objects and, consequently, to reduce the probability of confusion between them. Many theories of perceptual learning have assumed something like increasing sensitivity over practice (e.g., Gibson & Gibson, 1955), so it is a promising avenue for future research. However, Nosofsky (1988) did not explain how or why sensitivity might increase with practice, so the idea needs further development.

Attention and Learning

The idea that attention determines what is learned has been important throughout the history of the memory branch of the family tree. Medin and Schaffer (1978) argued that subjects learned the dimensions they attended to in proportion to the amount of attention they paid to them during training. My instance theory of automaticity assumed that attention is essential for learning. Attention at encoding determines what is put into an instance, and attention at retrieval determines which instances are retrieved (Logan, 1988, 1990). However, the memory branch provided no theory of attention that would allow investigators to predict what was learned and what was retrieved. ITAM's combination of TVA and instance theory is intended to provide such a theory.

ITAM assumes that response-set categorizations (driven by β) are represented explicitly in working memory, but stimulus-set categorizations (driven by π) are not (see Equation 16; also see Bundesen, 1990; Logan, 1996; Logan & Bundesen, 1996). A natural consequence of this assumption is that explicit learning depends only on what is contained in working memory, and thus the instances that are learned reflect the contents of working memory (Boronat & Logan, 1997; Logan & Etherton, 1994). Thus, ITAM would seem to predict that subjects learn response-set categorizations of the stimuli they encounter, but they do not learn (much) about stimulus-set categorizations. This prediction is supported by several data sets. Logan (1990) found that subjects learned particular response-set categorizations of display items. When the response set changed so that different categorizations of the same items were now relevant, there were no savings from previous presentations (also see Gorfein & Bubka, 1997; Logan, 1988). By contrast, Logan, Taylor, and Etherton (1996, 1999) found complete savings when subjects were transferred from one stimulus-set categorization to another, as if the stimulus-set categories were not associated with the items at all (also see C. B. Cave, Bost, & Cobb, 1996; Logan, 1998).

Learning in multi-item displays presents an important challenge to ITAM. Subjects appear to learn associations between objects they attend to and learn little about objects they ignore, even though they are presented in central vision and influence immediate performance to some degree (Boronat & Logan, 1997; Logan

& Etherton, 1994). Subjects are able to learn where to attend in multi-item displays, screening out distracting stimuli (Hillstrom & Logan, 1998), focusing attention on likely display positions (Chun & Jiang, 1998), and ignoring irrelevant positions (Haider & Frensch, 1996, 1999). Modeling these kinds of learning is an important direction for future research.

Response Selection

ITAM assumes that the categorizations that TVA and GCM produce are accumulated in *response counters*, and there is one counter for each response. The values in the counters are compared with a decision rule, and the response associated with the first counter that satisfies the decision rule is selected and passed on to the motor system to be executed. There are three distinct decision rules in ITAM, a race model, a counter model, and a random-walk model. Response selection and the categorization processes that precede it are *cascaded* rather than discrete stages (McClelland, 1979).

Race Model

The race model has been discussed throughout this article. The decision rule is to stop when the first categorization arrives in a counter, choosing the response associated with the counter in which the categorization appears. Race models are fast but are prone to error, particularly in situations in which there is conflict (e.g., categorizing *bat* as mammal vs. bird). In conflict situations, RT usually increases, and accuracy remains relatively stable (see e.g., Logan & Etherton, 1994; Palmeri, 1997). The race model predicts difficulty in conflict situations, but it places the effect in the wrong dependent variable. It predicts no difference in RT between conflict and nonconflict situations (e.g., congruent and incongruent Stroop stimuli) as long as the number of runners in the race is the same and they race at the same speed. This is not in accord with the data (e.g., Logan & Etherton, 1994; Palmeri, 1997). The race model predicts a large effect of conflict in the accuracy data. Accuracy depends on the probability that the first categorization is correct and that probability can be quite low in conflict situations (e.g., 0.7). Error rate is 1.0 minus the probability that the first categorization is correct, so in the example just given, the error rate would be 0.3 (i.e., $1.0 - 0.7 = 0.3$). This, too, is not in accord with the data (e.g., Logan & Etherton, 1994; Palmeri, 1997).

Counter Model

The counter model is a generalization of the race model. The counter model's decision rule is to respond with the first counter to accumulate K traces (see Logan, 1996; Townsend & Ashby, 1983). If K is 1.0, then the counter model is equivalent to a race model. It stops with the first categorization and produces the same RT and accuracy. If K is greater than 1.0, the counter model is slower and more accurate than a race model. If the probability that the winning categorization was correct were 0.7, the race model would produce an accuracy of 0.7. The counter model would produce an accuracy of 0.784 if K were 2.0 and 0.837 if K were 3.0. These increases in accuracy are bought at the cost of an increase in RT. The counter

model requires at least K categorizations to finish before it responds.

The counter model maintains accuracy in the face of conflict. In some conditions, RT will increase with conflict, which is consistent with the data. For example, categorizations may occur at the same rate in conflict and nonconflict situations, so the amount of time to accumulate a count would remain the same. However, the probability of accumulating the correct count would decrease, and more counts would be required overall to accumulate K counts in the correct counter. Logan (1996) found that CTVA and a counter model provided a reasonable account of RT and accuracy in B. A. Eriksen and Eriksen's (1974) flanker task.

Random-Walk Model

The random-walk model is the third decision rule that can be implemented on ITAM's counters, inherited from EBRW (Nosofsky & Palmeri, 1997; Palmeri, 1997). The random-walk model waits until there are K more categorizations in the correct counter than in any other counter. The random-walk decision criterion is relative, whereas the counter decision criterion is absolute. Like the counter model, the random walk reduces to the race model if K equals 1.0, responding when the first categorization finishes.

The random-walk model accounts for accuracy and RT data in conflict situations (Nosofsky & Palmeri, 1997; Palmeri, 1997; Ratcliff, 1978, 1988). Accuracy is higher with a random-walk than with a race model because several categorizations must accumulate before a response is chosen. RT is longer in conflict situations because more categorizations accumulate in the other, incorrect counters, and the correct counter cannot be chosen until it has K more categorizations than any other. Given the same processing rates, the random-walk model will take more steps to terminate than the counter model. At most, the counter model will stop after $K + (N_R - 1)(K - 1)$ steps, where N_R is the number of alternatives in the response set **R**. By contrast, the number of steps a random walk may take is not limited so sharply (Feller, 1968).

Nosofsky and Palmeri (1997) adopted the random-walk decision rule for EBRW because of its ability to account for conflict RT and accuracy and because its duration was not limited as sharply as the counter model's. Consequently, the random-walk model is the preferred decision rule in ITAM. The simple race and the counter model remain strategies in the arsenal of ITAM's homunculus (Logan & Gordon, 2001). A beast equipped with EBRW could instantiate a simple race by setting K to 1.0.

The mathematics of the process of object selection and classification (i.e., TVA and GCM) do not have to be changed at all to generalize ITAM from a simple race to a counter or random-walk model. This follows from ITAM's interpretation of the processing rates, v, as rate parameters for exponential distributions. As long as the rate parameters stay constant, choice probabilities will stay the same and the expected interval between one runner finishing and the next one finishing (i.e., mean finishing time) will stay constant over time (Townsend & Ashby, 1983). Formally, object selection and classification constitute a Poisson process that continues to feed categorizations to the response selection process at the same rate as long as the parameters remain constant. The object selection and classification processes are the same whether the response selection process chooses to respond when the first categorization becomes available (a simple race), when K categorizations accu-

mulate for one alternative (a counter model), or when K more categorizations accumulate for one alternative than for any other (a random walk).

Following EBRW, the time for the random walk to finish is the sum of the durations of each step it takes. The duration of a step in the random walk, ST, is determined by two factors: the time required to select an object and a categorization, given in Equation 5, and a parameter α that represents the time required to increment the random-walk counters and test the threshold (Nosofsky & Palmeri, 1997). That is,

$$ST = \alpha + \frac{1}{\sum_{z \in D} \sum_{j \in R} v(z, j)}. \tag{34}$$

RT is computed by multiplying step time by the expected number of steps in the random walk, $E(T)$:

$$RT = E(T) \cdot ST = E(T) \left(\alpha + \frac{1}{\sum_{z \in D} \sum_{j \in R} v(z, j)} \right). \tag{35}$$

If there are only two responses, the expected number of steps and the accuracy of the random walk overall can be calculated from analytic equations provided by Busemeyer (1982) and Nosofsky and Palmeri (1997). If there are more than two responses, the number of steps and accuracy have to be simulated (see Logan & Gordon, 2001; Palmeri, 1997). Logan (1996) and Townsend and Ashby (1983) provide analytic equations for RT and accuracy for a counter model with two response alternatives. Models with more than two alternatives remain be worked out mathematically or must be simulated.

Remaining Issues

Constraints From Perception

ITAM begins with an "encoded" stimulus and ends with a symbolic description of the response to be executed. In this context, the process of encoding a stimulus involves computing a set of η values that represent the similarity between each stimulus and each instance in memory. In theory, there will be many such η values; in practice, only those for which β and π are high need to be computed to generate ITAM's predictions. Recent theorists have suggested two ways in which the effective η values are constrained further by perception. One, CTVA, is consistent with ITAM, but the other, EGCM, may not be.

CTVA. Logan (1996) and Logan and Bundesen (1996) proposed a theory that combined Bundesen's (1990) TVA and van Oeffelen and Vos's (1982, 1983) COntour DEtector (CODE) theory of perceptual grouping by proximity. The CODE theory assumes that information about objects is not represented as a point in (cortical) space, but rather as a distribution over space (Figure 9, Panel A; also see Ashby, Prinzmetal, Ivry, & Maddox, 1996; Maddox, Prinzmetal, Ivry, & Ashby, 1994). The distributions for the various objects in the display are used to generate a surface that can be operated on to produce regions that represent perceptual groups (Figure 9, Panel B; also see Compton & Logan, 1993, 1999). In CTVA, TVA samples information from one or more of the regions defined by CODE (Figure 9, Panel C). Object repre-
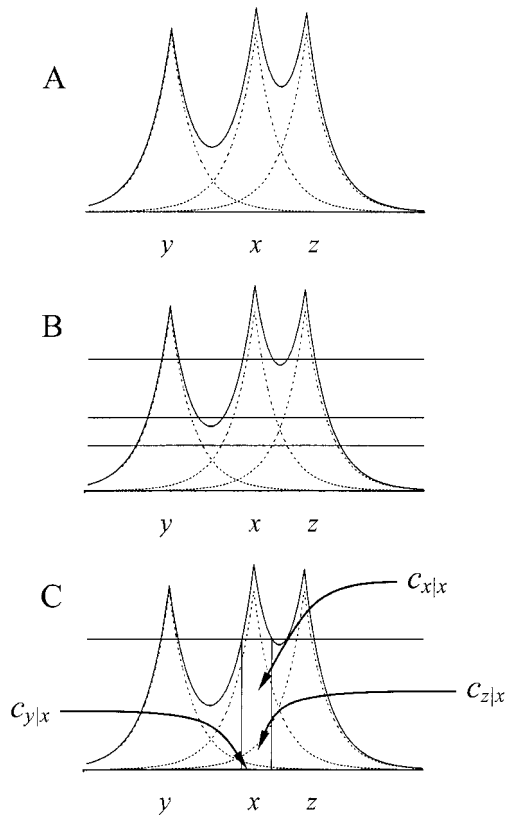
Figure 9. Illustration of the COntour DEtection (CODE) theory of visual attention (CTVA). There are three objects, y, x, and z, distributed along one dimension in space. Panel A represents the distribution of features of each object over representational space (dotted lines). The solid line in Panel A is a *CODE surface* formed by summing the feature distributions at each point. Panel B represents alternative groupings of the display (by proximity). The horizontal lines drawn across the CODE surface are *thresholds*, and each threshold represents a different perceptual organization. According to CODE, a perceptual group is an above-threshold region of the CODE surface. The top threshold produces three above-threshold regions, so each object is grouped separately. There are three groups, $\{y\}$, $\{x\}$, and $\{z\}$. The middle threshold produces two above-threshold regions and thus two groups, $\{y\}$ and $\{x, z\}$. The bottom threshold produces one large above-threshold region, and thus one group $\{y, x, z\}$. Panel C shows how TVA samples from an above-threshold region. The solid horizontal line represents the threshold, and the two solid vertical lines dropping down from it to the ordinate represent the limits of the region in which sampling takes place. The probability of sampling features from an object depends on the area of its feature distribution that falls within the sampled region. This probability is higher for object x than for objects y and z, and it is higher for object z than for object y (i.e., $c_{x|x} > c_{z|x} > c_{y|x}$).

sentations are distributed over space, so it is unlikely that all of the features of object x are available when TVA samples from the region containing object x. Instead, CTVA proposes that some proportion of the features, $c_{x|x}$, are available to drive TVA (Figure 9, Panel C). The c parameter is called the *feature catch* because it represents the proportion of features of object x that are "caught" in the current perceptual organization. This limiting of object features is represented in the model by multiplying η parameters by the c parameters from CODE. Thus,

$$\eta(x, i)_{\text{TVA}} \Rightarrow c_{x|x}\eta(x, i)_{\text{CTVA}} \Rightarrow c_{x|x} \sum_{m=1}^{N_i} \eta(x, i_m)_{\text{ITAM}}. \quad (36)$$

This substitution applies to ηs for stimulus set as well as to ηs for response set.

The idea that object representations are distributed over space implies that features of another object, for example, y, may be sampled from the region containing x. This, too, is represented in CTVA as a feature catch parameter, $c_{y|x}$, which multiplies all the η values representing object y. The closer object y is to object x (in display space), the greater the overlap in their distributions (in cortical space) and so the greater the likelihood that features from object y will be sampled from the region containing object x. That is, the greater the feature catch $c_{y|x}$ from object y. Also, the closer object x is to object y, the smaller the region around object x that is sampled, so the smaller the feature catch $c_{x|x}$. These ideas are responsible for CTVA's account of various distance and grouping effects in the attention literature (Logan, 1996; Logan & Bundesen, 1996).

The idea that perceptual processes constrain the η values in the way proposed in CTVA is quite consistent with ITAM. The substitution in Equation 36 could be performed in all of the ITAM equations and still preserve the flavor of the theory. This is one reason why CTVA is on the path from TVA to ITAM in the family tree.

CTVA is on the path from TVA to ITAM for another reason: It fixes a problem in TVA. A large amount of research shows that location is special in visual attention. People seem to select objects by location rather than by other properties, like color and identity (e.g., Nissen, 1985; Van der Heijden, 1992). However, location is not special in TVA. Location is just another stimulus property or just another categorization; π and β work in the same way for location as they do for any other property. CTVA makes location special by parsing the display into regions that correspond to perceptual groups and by having TVA sample information from those regions. The representation of location is more complex, but the complexity of the representation allows more sophisticated processing (see Logan, 1996).

EGCM. In GCM and in ITAM, the η values are set once at the beginning of the trial (as a result of the encoding stage) and remain the same throughout the trial. Lamberts (1998) challenged this idea in experiments that used separable-dimension stimuli in which one dimension was much easier to discriminate than the other. Subjects were required to respond quickly, before a deadline expired. When the deadline was short and the easy-to-discriminate dimension favored the wrong categorization, subjects often made errors. However, when the deadline was extended so they could process both dimensions before responding, subjects responded accurately to these same stimuli. Lamberts (1998) argued that the similarity representation evolves throughout the trial (also see Lamberts, 1995). Early on, only one dimension contributes to the representation. As time passes and more dimensions become available, the similarity representation becomes more complex, including more dimensions. Lamberts (2000) proposed an extended generalized context model (EGCM) to account for these results. It resembles GCM but it changes the assumption that the η values are set once and for all at the beginning of the trial. Instead, the η values develop over time as different dimensions become available.

The constancy of η values over time is an important assumption in GCM because it simplifies the mathematics considerably. It is a key assumption in ITAM, because the η values represent hazard functions of exponential distributions, and the proofs of equivalence between choice models and race models (Bundesen, 1993; Marley & Colonius, 1992) require that the hazard functions remain mutually proportional over time. Thus, EGCM may be viewed as a serious challenge to GCM and ITAM. On the other hand, EGCM's assumption that different dimensions become available at different times seems more realistic than GCM's simultaneity assumption, and it does appear to be in accord with empirical facts (e.g., Lamberts, 1998).

Executive Control

Recently, researchers have become interested in the executive processes that control cognition and allow a person to switch between tasks and between strategies within tasks (Logan, 1985; Meyer & Keiras, 1997; Norman & Shallice, 1986). One approach has been empirical, trying to discover the regularities underlying experiments on task switching (Allport, Styles, & Hsieh, 1994; Meiran, 1996; Rogers & Monsell, 1995). Another approach has been theoretical, trying to characterize executive operations in terms of a cognitive architecture (Meyer & Keiras, 1997; Norman & Shallice, 1986). ITAM allows a different approach, inherited from ECTVA, specifying executive processing in terms of control over TVA. In ITAM, a task set is a set of TVA parameters necessary to configure it to perform a particular task. Switching set involves changing TVA's parameters (Logan & Gordon, 2001).

ITAM has eight types of parameters, η, β, π, w, s, K, α, and c. Of the eight, four are controlled entirely by the homunculus (i.e., β, π, w, and K); another, c, is controlled partly by the homunculus (which chooses among alternative perceptual groupings) and partly by the spatial layout of the display (which determines proximities between objects). The rest (η, s, and α) are beyond the homunculus' control; η depends on the perceptual qualities of the display and the person's history with the categories in the stimulus set and the response set, and s and α reflect system limitations within the person. A task set, then, is a set of β, π, w, and K values that are sufficient to perform a particular task.

Logan and Gordon (2001) modeled executive control over TVA in a dual-task procedure, in which subjects made separate responses to two stimuli that were presented simultaneously or in rapid succession. In order to get TVA to perform both tasks, Logan and Gordon had to run it twice, once for each stimulus. To make it run twice, they had to specify the executive control processes that caused parameter values to change. For example, to respond to the first of the two stimuli, π had to be set to select properties that distinguished the first stimulus from the second, and β had to be set to select the appropriate classification for the first stimulus. These parameters remained in effect until the first response occurred, whereupon the executive process switched set for the second task. At minimum, π had to change to select the second stimulus rather than the first. If the task set was the same for the two stimuli (which is rare in dual-task studies but a common feature of Logan and Gordon's experiments), then the β values remained the same. If the task set was different, the β values had to be changed. With these assumptions, Logan and Gordon were able to account for set-switching effects in their dual-task experiments. They argued that their account could be generalized to other paradigms, such as that of Allport et al. (1994).

Discussion

ITAM assumes that attention and categorization are choice processes instantiated as a race between competing alternatives. There is only one race. Attention and categorization are the same thing. An object and a categorization for that object are selected in a single step. The choice processes are driven by similarities between display objects and memory representations of category exemplars. Categories are represented as collections of instances, and learning occurs through the accumulation of instances with practice. The output of object and category selection is input to a random-walk response selection process.

Formally, choice probabilities can be computed from the Shepard-Luce choice rule:

$$P(x \cap i) = \frac{v(x, i)}{\displaystyle\sum_{z \in D} \sum_{j \in R} v(z, j)}. \tag{37}$$

The time required for the choice (i.e., to run the race) is

$$FT = \frac{1}{\displaystyle\sum_{z \in D} \sum_{j \in R} v(z, j)}. \tag{38}$$

The expression for processing rate is

$$v(x, i) = c_{x|x} \sum_{m=1}^{N_i} \eta(x, i_m)\beta_i P_\pi(x), \tag{39}$$

where

$$P_\pi(x) = \frac{c_{x|x} \displaystyle\sum_{k \in S} \sum_{m=1}^{N_k} \eta(x, k_m)\pi_k}{\displaystyle\sum_{z \in D} c_{z|z} \sum_{k \in S} \sum_{m=1}^{N_k} \eta(z, k_m)\pi_k}. \tag{40}$$

The η values are defined in terms of distance in multidimensional similarity space:

$$\eta(x, i) = \exp[-s \cdot d_{xi}], \tag{41}$$

and distance is defined by

$$d_{xi} = \left[\sum_{h=1}^{H} a_h |u_{xh} - u_{ih}|^r\right]^{\frac{1}{r}}. \tag{42}$$

The expressions for the time taken in each step of the random walk and the total duration of the random walk are given in Equations 34 and 35.

Equations 37–42 express the formal structure of ITAM. They can be substituted for the equations of its ancestors to extend the ancestors to deal with phenomena that currently elude them (i.e., TVA, CTVA, ECTVA on the attention branch and the context model, GCM, instance theory, and EBRW on the memory branch;

these theories are connected with bidirectional arrows in Figure 1). The general theory can apply in each case.

From another perspective, each of ITAM's ancestors can be construed as a special case. In the attention branch, if there is only one instance in memory or if prototype theory happens to be true, then ITAM becomes CTVA and ECTVA. In terms of processing rates, Equation 39 becomes

$$v(x, i) = c_{x|x}\eta(x, i)\beta_i P_\pi(x),$$

which is the expression for processing rate in CTVA and ECTVA (Logan, 1996; Logan & Gordon, 2001). If there is complete parallel processing over the display, then $c_{x|x} = 1.0$ and CTVA becomes TVA:

$$v(x, i) = \eta(x, i)\beta_i P_\pi(x),$$

which is the processing rate for TVA (Bundesen, 1990).

In the memory branch, if there is only one object in the display, $c_{x|x} = 1.0$ and $P_\pi(x) = 1.0$, so ITAM becomes EBRW and GCM:

$$v(x, i) = \sum_{m=1}^{N_i} \eta(x, i_m)\beta_i.$$

If all instances are equally similar to the target, then EBRW becomes a special (exponentially distributed) case of instance theory:

$$v(x, i) = N_i\eta(x, i)\beta_i.$$

Thus, ITAM can be configured to replicate each of its ancestors exactly. When it takes the form of the ancestor, it mimics the predictions of the ancestor exactly. In this sense, ITAM inherits the successes of its ancestors, and those successes have been considerable. ITAM is built from proven components.

Implications

The elementary psychological event is a conjunction. ITAM has two fundamental implications for research in both branches of the family tree: that the elementary psychological event is a conjunction, and that everything is conditional on attention. First, the elementary psychological event is as follows: The stochastic processes in ITAM are built around joint choices of objects and categorizations. An object and a categorization are chosen simultaneously in each run of the race. Probabilities of nonconjunctive choices, such as the choice of an object, are built by summing probabilities of conjunctive choices (see, e.g., Equation 40). Thus, ITAM assumes binding of objects and categorizations as an axiom. It does not explain how binding occurs. The idea that the elementary psychological event is a conjunction is a prediction that can be tested at various levels, from behavior to event-related brain potentials to single-cell firing rates. If it turns out to be false, the mathematical basis for ITAM will be undermined.

ITAM's idea that the elementary psychological event is a conjunction is consistent with Duncan and colleagues' biased competition theory of attention (Chelazzi, Duncan, Miller, & Desimone, 1998; Desimone & Duncan, 1995; Duncan, 1996; Duncan, Humphreys, & Ward, 1997). Duncan and colleagues assumed that each display object activates a number of modules in the brain that represent different properties of the objects. One module might

process color. Another might process orientation. The objects fight it out within modules until one object dominates. There is communication among the modules, so that when an object gains an advantage in one module, that advantage propagates to other modules and gives the object an extra edge in the other competitions. Ultimately, one object dominates all modules, and at that moment, the person has attended to the object. That moment may be the elementary conjunction event defined in ITAM: the simultaneous choice of an object and a categorization. Thus, the evidence that supports the biased competition theory may also support ITAM. Indeed, Duncan and colleagues often cite Bundesen's (1990) TVA theory as similar to their view.

Everything is conditional on attention. The second fundamental implication of ITAM is that object selection, categorization, memory retrieval, and automatic processing are all conditional on attention (see Bargh, 1992; De Jong, Liang, & Lauber, 1994; Logan, 1988). In ITAM, β and π determine which objects get selected, and β determines which categorizations are made of the selected object. They turn up the gain on desired objects and categorizations, and they turn down the gain on the undesired ones. This dependence on attention is also a prediction that can be tested in many ways. It has been tested extensively in the literature, and it has been supported many times.

The Stroop effect is conditional on stimulus set. It can be reduced or eliminated by placing the conflicting color word outside the focus of attention (Kahneman & Henik, 1981) or by narrowing the focus of attention to a single letter in the color word (Besner & Stolz, 1999; Besner, Stolz, & Boutilier, 1997). ITAM would model these effects in terms of π. The Stroop effect is also conditional on response set. Color words outside the response set produce smaller Stroop effects. Tasks other than naming the color often produce smaller Stroop effects (Bauer & Besner, 1997). ITAM would interpret these effects in terms of β.

Semantic priming also appears to be conditional on response set. It is reduced substantially if the prime is not treated as a word. For example, if subjects search the prime word *butter* for the letter *t*, *butter* will not prime *bread* (e.g., Chiappe, Smith, & Besner, 1996; Henik, Friedrich, & Kellogg, 1983; McKoon & Ratcliff, 1995; Smith, 1979; Smith, Theodor, & Franklin, 1983; Stolz & Besner, 1996). ITAM interprets this conditionality of priming in terms of β. Only those categorizations for which β is high have a chance to be selected. If the β required to process the prime does not address the word as a whole, then categorizations of the word will not be selected.

Episodic memory is also conditional on stimulus and response set. People learn about the things they attend to, and what they learn depends on how they paid attention. This idea is expressed in the literature on levels of processing (Craik & Lockhart, 1972), transfer-appropriate processing (Morris, Bransford, & Franks, 1977; also see Roediger, 1990), and encoding specificity (Tulving & Thompson, 1973). ITAM interprets these effects in terms of the task set in the study phase and the task set in the test phase. Memory performance will be good to the extent that β and π values match at study and test.

Falsifiability

Theories as broad as ITAM are notoriously difficult to falsify. ITAM has several components that are conceptually separate from

each other. It may be possible to falsify each of the components separately. The CODE assumptions about early perceptual processing are likely to be false, for example, but that need not compromise the rest of the theory. Any theory of perceptual processing that provides a value of c between 0.0 and 1.0 could take the place of CODE relatively transparently. This modularity makes the theory hard to falsify as a whole. I view that as a virtue, however. It allows me to improve faulty parts of the theory without harming the other parts. Following Newell (1990), I think it is more important to build theories than to tear them down.

From one perspective, ITAM is simply an assertion that certain theories of categorization and attention are formally related. That assertion can be falsified directly by finding errors in the mathematics. If the mathematics are correct, then the theory can be tested by seeing whether phenomena in the two branches of the family tree can be accounted for by the same parameterization of ITAM. If the range of parameters required to fit attentional phenomena does not overlap much with the range of parameters required to fit memory phenomena, then ITAM will be undermined. It would not be reasonable (or fruitful) to regard attention and memory as different perspectives on the same choice process. More generally, the claim that attention and memory are essentially the same may prompt researchers to look for more parallels between them. The fruitfulness of that enterprise is another test of the theory.

It is important to remember that ITAM is built of components that have been very successful in their own areas of research. Several have passed falsifiability tests in their own domains. ITAM inherits its ancestors' predictions, so it also passes those falsifiability tests as well.

Conclusions

ITAM suggests that attention, categorization, and memory are different aspects of the same fundamental choice process, and it shows how interactions between attention, categorization, and memory can be accounted for by a single set of simple equations (i.e., Equations 37–42). ITAM promises a unified account of human cognition and suggests essential similarities between cognitive phenomena that are traditionally viewed as separate and different. The next time you find yourself driving home reflecting on your colleagues' interpretations of the simple act of cognition by which you found your car, think about ITAM's claim that they all are studying the same phenomena and saying the same things. As John Lennon said, "I am he as you are he as you are me and we are all together" (Lennon & McCartney, 1967).

References

Allport, A., Styles, E. A., & Hsieh, S. (1994). Shifting intentional set: Exploring the dynamic control of tasks. In C. Umilta & M. Moscovitch (Eds.), *Attention and performance XV* (pp. 421–452). Cambridge, MA: MIT Press.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89,* 369–406.

Anderson, J. R. (1992). Automaticity and the ACT* theory. *American Journal of Psychology, 105,* 165–180.

Ashby, F. G., Prinzmetal, W., Ivry, R., & Maddox, W. T. (1996). A formal theory of feature binding in object perception. *Psychological Review, 103,* 165–192.

Bacon, W. F., & Egeth, H. E. (1991). Local processes in preattentive feature detection. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 77–90.

Bacon, W. F., & Egeth, H. E. (1997). Goal-directed guidance of attention: Evidence from conjunctive visual search. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 948–961.

Bargh, J. A. (1992). The ecology of automaticity: Toward establishing the conditions needed to produce automatic processing effects. *American Journal of Psychology, 105,* 181–193.

Bauer, B., & Besner, D. (1997). Processing in the Stroop task: Mental set as a determinant of performance. *Canadian Journal of Experimental Psychology, 51,* 61–68.

Bauer, B., Jolicoeur, P., & Cowan, W. B. (1996). Visual search for colour targets that are or are not linearly-separable from distractors. *Vision Research, 36,* 1439–1466.

Beck, J. (1982). Textural segmentation. In J. Beck (Ed.), *Organization and representation in perception* (pp. 285–317). Hillsdale, NJ: Erlbaum.

Besner, D., & Stolz, J. A. (1999). What kind of attention modulates the Stroop effect? *Psychonomic Bulletin & Review, 6,* 99–104.

Besner, D., Stolz, J. A., & Boutilier, C. (1997). The Stroop effect and the myth of automaticity. *Psychonomic Bulletin & Review, 4,* 221–225.

Boronat, C. B., & Logan, G. D. (1997). The role of attention in automatization: Does attention operate at encoding, or retrieval, or both? *Memory & Cognition, 25,* 36–46.

Broadbent, D. E. (1958). *Perception and communication.* Elmsford, NY: Pergamon Press.

Broadbent, D. E. (1971). *Decision and stress.* London: Academic Press.

Brown, R. (1958). How shall a thing be called? *Psychological Review, 65,* 14–21.

Bundesen, C. (1987). Visual attention: Race models for selection from multielement displays. *Psychological Research, 49,* 113–121.

Bundesen, C. (1990). A theory of visual attention. *Psychological Review, 97,* 523–547.

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

Bundesen, C. (1998a). A computational theory of visual attention. *Philosophical Transactions of the Royal Society: Series B, 353,* 1271–1281.

Bundesen, C. (1998b). Visual selective attention: Outlines of a choice model, a race model, and a computational theory. *Visual Cognition, 5,* 287–309.

Bundesen, C., & Harms, L. (1999). Single-letter recognition as a function of exposure duration. *Psychological Research, 62,* 275–279.

Bundesen, C., Pedersen, L. F., & Larsen, A. (1984). Measuring the efficiency of selection from briefly exposed displays: A model for partial report. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 329–339.

Bundesen, C., Shibuya, H., & Larsen, A. (1985). Visual selection from multielement displays: A model for partial report. In M. I. Posner & O. S. Marin (Eds.), *Attention and Performance XI* (pp. 631–649). Hillsdale, NJ: Erlbaum.

Busemeyer, J. R. (1982). Choice behavior in a sequential decision-making task. *Organizational Behavior and Human Performance, 29,* 175–207.

Cave, C. B., Bost, P. R., & Cobb, R. E. (1996). Effects of color and pattern on implicit and explicit picture memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 639–653.

Cave, K. R., & Wolfe, J. M. (1990). Modeling the role of parallel processing in visual search. *Cognitive Psychology, 22,* 225–271.

Chelazzi, L., Duncan, J., Miller, E. K., & Desimone, R. (1998). Responses to neurons in inferior temporal cortex during memory-guided visual search. *Journal of Neurophysiology, 80,* 2918–2940.

Chiappe, P. R., Smith, M. C., & Besner, D. (1996). Semantic priming in visual word recognition: Activation blocking and domains of processing. *Psychonomic Bulletin & Review, 3,* 249–253.

Chun, M., & Jiang, Y. (1998). Contextual cuing: Implicit learning and

memory of visual context guides spatial attention. *Cognitive Psychology, 36,* 28–71.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review, 97,* 332–361.

Colonius, H. (1995). The instance theory of automaticity: Why the Weibull? *Psychological Review, 102,* 744–750.

Compton, B. J., & Logan, G. D. (1993). Evaluating a computational model of perceptual grouping by proximity. *Perception & Psychophysics, 53,* 403–421.

Compton, B. J., & Logan, G. D. (1999). Judgments of perceptual groups: Reliability and sensitivity to stimulus transformation. *Perception & Psychophysics, 61,* 1320–1335.

Cousineau, D., Goodman, V., & Shiffrin, R. M. (in press). Extending statistics of extreme values to distributions varying in position and scale and the implications for race models. *Journal of Mathematical Psychology.*

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671–676.

De Jong, R., Liang, C.-C., & Lauber, E. (1994). Conditional and unconditional automaticity: A dual-process model of stimulus–response correspondence. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 731–750.

Delaney, P. F., Reder, L. M., Straszewski, J. J., & Ritter, F. E. (1998). The strategy-specific nature of improvement: The power law applies by strategy within task. *Psychological Science, 9,* 1–7.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective attention. *Annual Review of Neuroscience, 18,* 193–222.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review, 70,* 80–90.

Duncan, J. (1989). Boundary conditions on parallel processing in human vision. *Perception, 18,* 457–469.

Duncan, J. (1996). Cooperating brain systems in selective perception and action. In T. Inui & J. L. McClelland (Eds.), *Attention and Performance XVI* (pp. 549–578). Cambridge, MA: MIT Press.

Duncan, J., & Humphreys, G. (1989). A resemblance theory of visual search. *Psychological Review, 96,* 433–458.

Duncan, J., Humphreys, G., & Ward, R. (1997). Competitive brain activity in visual attention. *Current Opinion in Neurobiology, 7,* 255–261.

Egeth, H. E., Virzi, R. A., & Garbart, H. (1984). Searching for conjunctively defined targets. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 332–39.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16,* 143–149.

Eriksen, C. W., & Hoffman, J. E. (1972). Temporal and spatial characteristics of selective encoding from visual displays. *Perception & Psychophysics, 12,* 201–204.

Estes, W. (1974). Redundancy of noise elements in visual detection of letters. *Perception & Psychophysics, 16,* 53–60.

Farmer, E. W., & Taylor, R. M. (1980). Visual search through color displays: Effects of target–background similarity and background uniformity. *Perception & Psychophysics, 27,* 267–272.

Feller, W. (1968). *An introduction to probability theory and its applications.* (Vol. 1). New York: Wiley.

Franks, J. J., & Bransford, J. D. (1971). Abstraction of visual patterns. *Journal of Experimental Psychology, 90,* 65–74.

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

Garner, W. R., & Haun, F. (1978). Letter identification as a function of type of perceptual limitation and type of attribute. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 199–209.

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review, 62,* 32–41.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1–67.

Gorfein, D. S., & Bubka, A. (1997). A transfer analysis of the repetition effect in the lexical and ambiguity decision tasks. *Psychonomic Bulletin & Review, 4,* 232–236.

Haider, H., & Frensch, P. A. (1996). The role of information reduction in skill acquisition. *Cognitive Psychology, 30,* 304–337.

Haider, H., & Frensch, P. A. (1999). Eye movement during skill acquisition: More evidence for the information reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 172–190.

Hayes-Roth, B., & Hayes-Roth, F. (1977). Concept learning and the classification of exemplars. *Journal of Verbal Learning and Verbal Behavior, 16,* 321–338.

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review, 7,* 185–207.

Henik, A., Friedrich, F. J., & Kellogg, W. A. (1983). The dependence of semantic relatedness effects on prime processing. *Memory & Cognition, 11,* 366–373.

Hillstrom, A. P., & Logan, G. D. (1998). Decomposing visual search: Evidence for multiple item-specific skills. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1385–1398.

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

Hintzman, D. L. (1988). Judgments of frequency and recognition in a multiple-trace memory model. *Psychological Review, 95,* 528–551.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review, 99,* 480–517.

Humphreys, G. W., & Müller, H. J. (1993). SEarch via Recursive Rejection (SERR): A connectionist model of visual search. *Cognitive Psychology, 25,* 43–110.

Jacoby, L. L., & Brooks, L. R. (1984). Nonanalytic cognition: Memory, perception, and concept learning. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 18, pp. 1–47). New York: Academic Press.

James, W. (1890). *Principles of psychology.* New York: Holt.

Julesz, B., & Bergen, J. R. (1983). Textons, the fundamental elements in preattentive vision and perception of textures. *Bell System Technical Journal, 62,* 1619–1645.

Kahneman, D., & Henik, A. (1981). Perceptual organization and attention. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 307–332). Hillsdale, NJ: Erlbaum.

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

Kaptein, N. A., Theeuwes, J., & Van der Heijden, A. H. C. (1995). Search for a conjunctively defined target can be selectively limited to a color-defined subset of elements. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 1053–1069.

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

Kruschke, J. K., & Johansen, M. K. (1999). A model of probabilistic category learning. *Journal of Experimental Psychology: Learning, Memory and Cognition, 25,* 1083–1119.

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

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

Lamberts, K. (2000). Information-accumulation theory of speeded categorization. *Psychological Review, 107,* 227–260.

Lennon, J., & McCartney, P. (1967). I am the walrus. On *Magical mystery tour* [phonograph record]. London: Apple Records.

Logan, G. D. (1985). Executive control of thought and action. *Acta Psychologica, 60,* 193–210.

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.

Logan, G. D. (1995a). Linguistic and conceptual control of visual spatial attention. *Cognitive Psychology, 28,* 103–174.

Logan, G. D. (1995b). The Weibull distribution, the power law, and the instance theory of automaticity. *Psychological Review, 102,* 751–756.

Logan, G. D. (1996). The CODE theory of visual attention: An integration of space-based and object-based attention. *Psychological Review, 103,* 603–649.

Logan, G. D. (1998). What is learned during automatization? II. Obligatory encoding of location information. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1720–1736.

Logan, G. D., & Bundesen, C. (1996). Spatial effects in the partial report paradigm: A challenge for theories of visual–spatial attention. In D. L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 35, pp. 243–282). San Diego, CA: Academic Press.

Logan, G. D. & Etherton, J. L. (1994). What is learned in automatization? The role of attention in constructing an instance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1022–1050.

Logan, G. D., & Gordon, R. D. (2001). Executive control of visual attention in dual-task situations. *Psychological Review, 108,* 393–434.

Logan, G. D., Taylor, S. E., & Etherton, J. L. (1996). Attention in the acquisition and expression of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 620–638.

Logan, G. D., Taylor, S. E., & Etherton, J. L. (1999). Attention and automaticity: Toward a theoretical integration. *Psychological Research, 62,* 165–181.

Logan, G. D., & Zbrodoff, N. J. (1999). Selection for cognition: Cognitive constraints on visual spatial attention. *Visual Cognition, 6,* 55–81.

Luce, R. D. (1959). *Individual choice behavior.* New York: Wiley.

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.

MacKay, D. G. (1982). The problem of flexibility, fluency, and speed–accuracy tradeoff in skilled behavior. *Psychological Review, 89,* 483–506.

Maddox, W. T., Prinzmetal, W., Ivry, R., & Ashby, F. G. (1994). A probabilistic multidimensional model of location information. *Psychological Research, 56,* 66–77.

Marley, A. A. J., & Colonius, H. (1992). The "horse race" random utility model for choice probabilities and reaction times, and its competing risks interpretation. *Journal of Mathematical Psychology, 36,* 1–20.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86,* 287–330.

McClelland, J. L., Rumelhart, D. E., & the PDP Research Group (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition.* Cambridge, MA: MIT Press.

McKoon, G., & Ratcliff, R. (1995). Conceptual combinations and relational contexts in free association and in priming in lexical decision and naming. *Psychonomic Bulletin & Review, 2,* 527–533.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100,* 254–278.

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

Meiran, N. (1996). Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1423–1442.

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review, 104,* 3–65.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16,* 519–533.

Murdock, B. B. (1993). TODAM2: A model for the storage and retrieval of item, associative, and serial-order information. *Psychological Review, 100,* 183–203.

Myung, I. J., Kim, C., & Pitt, M. A. (2000). Toward an explanation of the power law artifact: Insights from response surface analysis. *Memory & Cognition, 28,* 832–840.

Newell, A. (1990). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

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.

Nissen, M. J. (1985). Accessing features and objects: Is location special? In M. I. Posner & O. S. M. Marin (Eds.), *Attention and Performance XI* (pp. 205–219). Hillsdale, NJ: Erlbaum.

Norman, D. A. (1968). Toward a theory of memory and attention. *Psychological Review, 75,* 522–536.

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behaviour. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation* (Vol. 4, pp. 1–18). New York: Plenum.

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,* 87–108.

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

Nosofsky, R. M. (1991a). Stimulus bias, asymmetric similarity, and classification. *Cognitive Psychology, 23,* 94–140.

Nosofsky, R. M. (1991b). 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., & Alphonso-Reese, L. A. (1999). Effects of similarity and practice on speeded classification response times and accuracies: Further tests of an exemplar–retrieval model. *Memory & Cognition, 27,* 78–93.

Nosofsky, R. M., Gluck, M. A., 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). (1994). *Memory & Cognition, 22,* 352–369.

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. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review, 104,* 266–300.

Nothdurft, H. C. (1992). Feature analysis and the role of similarity in preattentive vision. *Perception & Psychophysics, 52,* 355–375.

Palmeri, T. J. (1997). Exemplar similarity and the development of auto-

maticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 324–354.

Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma & D. Bouwhuis (Eds.), *Attention and Performance X* (pp. 531–556). Hillsdale, NJ: Erlbaum.

Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109,* 160–174.

Pylyshyn, Z. (1989). The role of location indices in spatial perception: A sketch of the FINST spatial-index model. *Cognition, 32,* 65–97.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 552–572.

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

Ratcliff, R. (1988). Continuous versus discrete information processing: Modeling the accumulation of partial information. *Psychological Review, 95,* 238–255.

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology, 3,* 382–407.

Rickard, T. C. (1997). Bending the power law: A CMPL theory of strategy shifts and the automatization of cognitive skills. *Journal of Experimental Psychology: General, 126,* 288–311.

Roediger, H. L., III. (1990). Implicit memory: Retention without remembering. *American Psychologist, 45,* 1043–1056.

Rogers, R. D., & Monsell, S. (1995). The cost of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124,* 207–231.

Schneider, W. (1985). Toward a model of attention and the development of automatic processing. In M. I. Posner & O. S. Marin (Eds.), *Attention and performance XI* (pp. 475–492). Hillsdale, NJ: Erlbaum.

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 generalization: Deduction of the generalization gradient from a 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. (1987). Toward a universal law of generalization for psychological space. *Science, 237,* 1317–1323.

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

Shibuya, H., & Bundesen, C. (1988). Visual selection from multielement displays: Measuring and modeling the effects of exposure duration. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 591–600.

Shin, H. J., & Nosofsky, R. M. (1992). Similarity-scaling studies of dot-pattern classification and recognition. *Journal of Experimental Psychology: General, 121,* 278–304.

Smith, M. C. (1979). Contextual facilitation in a letter search task depends on how the prime is processed. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 239–251.

Smith, M. C., Theodor, L., & Franklin, P. (1983). The relationship between contextual facilitation and depth of processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 697–712.

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

Stolz, J. A., & Besner, D. (1996). Role of set in visual word recognition: Activation and activation blocking as a nonautomatic process. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 1166–1177.

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

Treisman, A. (1969). Strategies and models of selective attention. *Psychological Review, 76,* 282–299.

Treisman, A. (1988). Features and objects: The Fourteenth Bartlett Memorial Lecture. *Quarterly Journal of Experimental Psychology, 40A,* 201–237.

Treisman, A. (1991). Search, similarity, and integration of features between and within dimensions. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 652–676.

Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12,* 97–136.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review, 95,* 14–48.

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology, 14,* 107–141.

Treisman, A., & Souther, J. (1985). Search asymmetry: A diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: General, 114,* 285–310.

Tulving, E., & Thompson, D. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80,* 352–373.

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

Tversky, A., & Gati, I. (1982). Similarity, separability, and the triangle inequality. *Psychological Review, 89,* 123–154.

Tversky, A., & Hutchinson, J. W. (1986). Nearest neighbor analysis of psychological spaces. *Psychological Review, 93,* 3–22.

Ullman, S. (1984). Visual routines. *Cognition, 18,* 97–159.

Van der Heijden, A. H. C. (1992). *Selective attention in vision.* New York: Routledge.

Van Oeffelen, M. P., & Vos, P. G. (1982). Configurational effects on the enumeration of dots: Counting by groups. *Memory & Cognition, 10,* 396–404.

Van Oeffelen, M. P., & Vos, P. G. (1983). An algorithm for pattern description on the level of relative proximity. *Pattern Recognition, 16,* 341–348.

Wolfe, J. M. (1992). "Effortless" texture segregation and "parallel" visual search are not the same thing. *Vision Research, 32,* 757–763.

Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review, 1,* 202–238.

Wolfe, J. M. (1997). What can 1 million trials tell us about visual search? *Psychological Science, 9,* 33–39.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 419–433.

Zohary, E., & Hochstein, S. (1989). How serial is serial processing in vision? *Perception, 18,* 191–200.