

The CODE Theory of Visual Attention: An Integration of Space-Based and Object-Based Attention

Gordon D. Logan
University of Illinois

This article presents a theory that integrates space-based and object-based approaches to visual attention. The theory puts together M. P. van Oeffelen and P. G. Vos's (1982, 1983) COntour DE-tector (CODE) theory of perceptual grouping by proximity with C. Bundesen's (1990) theory of visual attention (TVA). CODE provides input to TVA, accounting for spatially based between-object selection, and TVA converts the input to output, accounting for feature- and category-based within-object selection. CODE clusters nearby items into perceptual groups that are both perceptual objects and regions of space, thereby integrating object-based and space-based approaches to attention. The combined theory provides a quantitative account of the effects of grouping by proximity and distance between items on reaction time and accuracy data in 7 empirical situations that shaped the current literature on visual spatial attention.

For the last decade the attention literature has been embroiled in a debate over the nature of visual spatial attention that focuses on the "thing" that attention selects (e.g., Baylis & Driver, 1993; Driver & Baylis, 1989; Duncan, 1984; Egly, Driver, & Rafal, 1994; Kramer & Jacobson, 1991; Vecera & Farah, 1994). Advocates of *space-based attention* argue that attention selects regions of space independent of the objects they contain. Attention is like a spotlight illuminating a region of space. Objects that fall within the beam are processed; objects that fall outside it are not (Eriksen & Eriksen, 1974; Eriksen & St. James, 1986; Posner, 1980; Posner & Cohen, 1984; Treisman & Gelade, 1980; Treisman & Gormican, 1988). Advocates of *object-based attention* argue that attention selects objects rather than regions of space. Selection is spatial because objects necessarily occupy regions of space, but objects rather than the regions themselves are the things that are selected (Kahneman & Henik, 1981; Kahneman & Treisman, 1984; Kahneman, Treisman, & Gibbs, 1992; Pylyshyn & Storm, 1988). Object-based theories assume that attention only selects regions of space that are occupied by objects, whereas space-based theories assume that attention can select empty regions of space (cf. Yantis, 1992).

The purpose of this article is to propose a theory of visual spatial attention that integrates space-based and object-based views. The theory takes a computational approach to the problem, characterizing attention in terms of representations and the processes that operate on them. It differs from most ap-

proaches to attention by being concerned with the representation of space and the representation of objects, incorporating a theory of perceptual organization and a theory of selection. The resolution of the controversy derives from the theory's assumptions about representation.

The article begins by describing five important questions that face any theory of visual spatial attention. The answers proposed by the new theory are presented by way of describing the theory. The theory is applied to seven important paradigms that shaped the current literature on visual spatial attention. Finally, the benefits and limitations of the theory are discussed, and fruitful directions for future research are pointed out.

Five Key Questions

How Is Space Represented?

A key question for both space-based and object-based theories of attention is how space is represented. Despite the importance of space in theories of attention for the last decade or two, very little has been said explicitly about the representation of space, perhaps because it seems that little needs to be said: Objects are arrayed in space in the world. Optics preserve the spatial array as the world is projected on the retinae. Retinotopic projection from retinae to cortex preserves the spatial arrangement in visual cortex, which is interpreted (by theorists) as a representation of space.

Space-based theories of attention appear to assume that space is represented by a two- or three-dimensional (2-D or 3-D) map of locations, with objects represented as points in space (Cave & Wolfe, 1990; Treisman, 1990; Treisman & Gelade, 1980; Treisman & Gormican, 1988; Wolfe, 1994). Theorists appear to assume that distance between objects is represented by a Euclidean metric, because Euclidean distance is an important variable in studies of space-based attention (e.g., Eriksen & Hoffman, 1973; Shulman, Remington, & McLean, 1979; Tsai, 1983).

Object-based theories have been even less explicit about the

This research was supported by National Science Foundation Grant SBR 94-10406. I am grateful to Claus Bundesen for help in extending his theory, to Jim Townsend and especially Brian Ross for help with the mathematics, to Asher Cohen and Claus Bundesen for helpful comments on the article, and to Brian Compton for preparing the figures.

Correspondence concerning this article should be addressed to Gordon D. Logan, Department of Psychology, University of Illinois, 603 East Daniel Street, Champaign, Illinois 61820. Electronic mail may be sent via Internet to glogan@psych.uiuc.edu.

representation of space. They appear to assume that space is represented as a 2-D or 3-D array of objects, organized by Gestalt grouping principles. The distance metric is not clear, but Euclidean distance is not especially important. Researchers often interpret data as evidence for object-based attention when grouping factors counteract Euclidean distance (Baylis & Driver, 1993; Driver & Baylis, 1989; Kramer & Jacobson, 1991). This abandonment of Euclidean distance and granting it to the opposition is a peculiar tactic for object-based theorists because grouping by proximity is a powerful and important Gestalt principle.

A theory of space-based or object-based attention must be explicit in its assumptions about how space is represented. Otherwise, the theories cannot be tested adequately. Moreover, comparisons between classes of theories can be made only if their assumptions about spatial representation are explicit. Otherwise, it is difficult to derive contrasting predictions.

What Is an Object?

The definition of an object is a central issue in object-based theories. Nevertheless, there is no commonly agreed upon definition. The most common tactic is to rely on intuition, as if William James had said, "Everyone knows what an object is." Some researchers rely on ratings of "goodness" of objects (e.g., Kramer & Jacobson, 1991), democratizing intuition. Others rely on Gestalt grouping principles, like similarity (Kahneman & Henik, 1977), common fate (Driver & Baylis, 1989), and proximity (Banks & Prinzmetal, 1976; Prinzmetal, 1981). Still others rely on spatial contiguity: Objects are conjunctions of properties that occur at a common location (Kahneman & Treisman, 1984; Kahneman, Treisman, & Gibbs, 1992). Despite the lack of consensual definition, most researchers agree that objects are hierarchical; objects can be decomposed into parts, and each part can be treated as a single object (Baylis & Driver, 1993; Biederman, 1987; Marr & Nishihara, 1978; Navon, 1977; Palmer, 1977; Palmer & Rock, 1994). A theory of object-based attention should say what an object is and should account for hierarchical organization in the definition it provides.

What Determines Shape of the Spotlight?

A great deal of space-based research has addressed the shape of the region that attention selects. The default assumption seems to be that the region is round, like a spotlight beam, but some researchers have suggested different shapes, from ovals (Eriksen, Pan, & Botella, 1994) to doughnuts (Juola, Bouwhuis, Cooper, & Warner, 1991). In most of these approaches, the shape is determined by "endogenous factors" or "higher-level processes" that are outside the scope of the theory. If that is the case, then the shape of the region becomes a free parameter (or set of parameters) that the theorist can set without constraint to accommodate whatever data may appear. The theory I am proposing constrains the shape of the spotlight, reducing the need to invoke a homunculus to explain selection.

LaBerge and Brown (1989) took a more principled approach to the shape of the spotlight in their gradient theory of attention. They assumed that the spotlight adjusts to the shape of the se-

lected object, opening an aperture the size and shape of the object through which perceptual features are sampled. The main empirical thrust of their assumption focused on aftereffects of selection, providing data that suggested that an aperture the size and shape of the target remained open for a short time after the selected object disappeared.

A theory of space-based attention must be explicit about what determines the shape of the spotlight. Better theories will be more specific about the factors that determine it and leave less work for an omnipotent homunculus to do.

How Does Selection Occur Within the Focus of Attention?

Space-based and object-based theories both assume that everything within the focus of attention is processed. Space-based theories assume that everything within the spotlight is processed (e.g., Eriksen & St. James, 1986; Treisman & Gormican, 1988), and object-based theories assume that every property of the selected object is processed (e.g., Kahneman & Henik, 1981; Kahneman & Treisman, 1984; Kahneman et al., 1992). These assumptions, by themselves, cannot account for cases in which selection occurs within the focus of attention (see Kahneman, 1973; Posner & Boies, 1971; Treisman, 1969).

The Stroop (1935) task provides a compelling example of selection within the spatial focus of attention. The task requires subjects to name the color in which a word is written and ignore the name of the word. They manage to do so with great success. Reaction times may be slower when the word names a different color than the target (*incompatible* displays, e.g., GREEN in red) than when the word names a noncolor (*neutral* displays, e.g., MOST in red) or the same color as the target (*compatible* displays, e.g., RED in red), but accuracy is high. Subjects rarely report the word instead of the color (for a review, see MacLeod, 1991).¹

The basic Stroop results are difficult to accommodate with the assumption that everything in the spatial focus of attention is processed. If everything in the spotlight or every property of the selected object was processed, then the word should have been processed as well as the color. Moreover, the word is usually processed faster and more accurately than the color (Cohen, Dunbar, & McClelland, 1990; Logan, 1980). So why doesn't the word determine performance? Because there is more to attention than spatial selection (Broadbent, 1971; Kahneman, 1973; Posner & Boies, 1971; Treisman, 1969). Theories of visual-spatial attention must interface with theories of other kinds of selection to account for the basic phenomena in visual-spatial attention and to provide a realistic account of attention in general (Phaf, van der Heijden, & Hudson, 1990).

How Does Selection Between Objects Occur?

Most theories agree that visual attention is sometimes serial, focusing on one item or one set of items at a time, moving from

¹ Subjects rarely make mistakes; error rate is typically lower than 10%. When they do make errors, they tend to report the word (Hillstrom & Logan, in press, found that subjects reported the word on 83% of the error trials), but they do not make errors very often.

one to the other. Serial shifting of the focus is an important issue in search tasks (e.g., Duncan & Humphreys, 1989; Treisman & Gelade, 1980; Treisman & Gormican, 1988; Wolfe, 1994). Serial search raises important questions for between-object selection: How does attention know which item to choose next? How does the spotlight know where to go? There are very few theories of the processes that govern the movements of attention (but see Cave & Wolfe, 1990; Koch & Ullman, 1985; Wolfe, 1994; Wolfe, Cave, & Franzel, 1989). Most often, the government is left to "higher order processes" or homunculi.

Selection between objects is a prominent feature of attentional cuing tasks, which many theories address (e.g., Eriksen & St. James, 1986; Eriksen & Yeh, 1985; Posner, 1980; Posner & Cohen, 1984). Cues are presented that indicate which item is the target or which location is likely to contain the target, and subjects benefit from using the cues. Cuing also raises important questions about between-object selection: How does attention know where to go? It has to go to the cue first and then from the cue to the target. How does it know what to do? The computational problem is more complex than with search, because attention has to move in a specific direction. Few theories address the problem even though it is a major issue in cuing tasks (but see Logan, 1995).

The problem for space-based and object-based theories is that they must interface with other theories of cognition in order to account for basic phenomena like serial visual search or moving attention from cue to target. The other theories may be able to explain some of the things that are currently left to homunculi (Attneave, 1960).

Theory of Attention

In this article, I propose the *CODE theory of visual attention* (CTVA) that integrates space-based and object-based approaches to attention and interfaces visual spatial attention with other kinds of attentional selection and with higher level processes that apprehend relations between objects. The theory is a wedding of the *C*Ontour *D*ETector (CODE) theory of perceptual grouping by proximity (Compton & Logan, 1993; van Oeffelen & Vos, 1982, 1983) and Bundesen's (1990) *t*heory of *v*isual *a*ttention (TVA). As with most weddings, each theory retains its fundamental identity but compromises on details in order to work with the other. This section of the article describes the fundamental assumptions of the theories before and after the wedding and describes the compromises and developments that were necessary to join the theories together.

Basic Architecture

The basic architecture of the theory is illustrated in the top panel of Figure 1. Many theories of visual spatial attention use the same architecture (e.g., Milner, 1974; Mozer, 1991; Treisman & Gormican, 1988; van der Heijden, 1992): There are early visual processes and late visual processes. The early visual processes, often identified with V1 in striate cortex, represent location and identity together. Later processes distinguish between location and identity and represent them separately. The late location system is identified with processes in the magnocellular pathway leading through V2, V3, and V5 to posterior

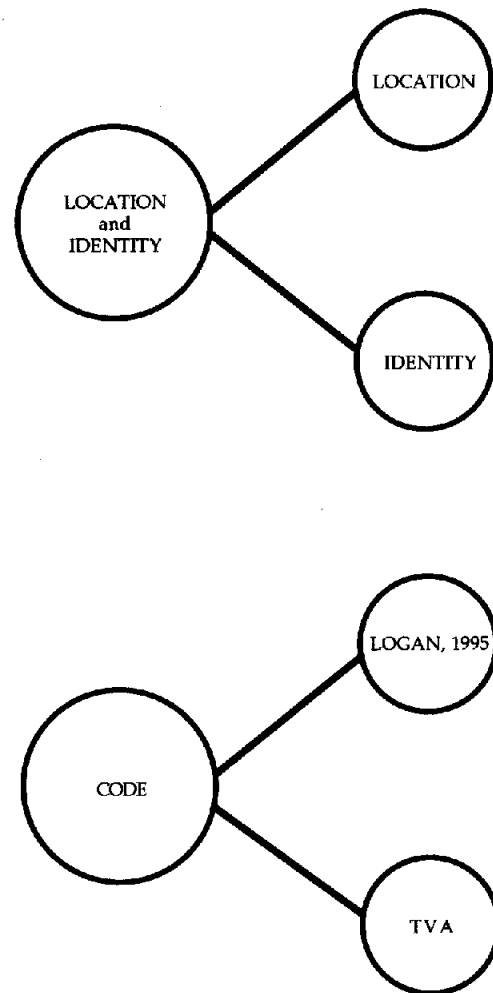


Figure 1. Architecture of the CODE Theory of Visual Attention, including an early system in which location and identity information are combined, represented by the CODE theory, a late system that processes identity information, represented by Bundesen's (1990) Theory of Visual Attention (TVA), and a late location system, represented by Logan's (1995) spatial relation theory. Top panel: schematic representation of components; bottom panel: theories associated with the components.

parietal cortex, whereas the late identity system is identified with processes in the parvocellular pathway leading through V2, V3, and V4 to inferotemporal cortex (Ungerleider & Mishkin, 1982; van der Heijden, 1992).

The CODE theory of visual attention adopts the same architecture but fleshes out the details. As illustrated in the bottom panel of Figure 1, the early visual processes are represented by van Oeffelen and Vos's (1982, 1983) and Compton and Logan's (1993) CODE model of perceptual grouping by proximity and the late identity processes are represented by Bundesen's (1990) TVA model of parallel selection in vision. The late location system is less well developed because less is known about conceptual representation of location. The CODE theory of visual attention adopts a preliminary theory proposed by Logan

(1995) that accounts for some instances of conceptually guided selection between perceptual objects.

Representing Space and Defining Objects

CODE theory of grouping by proximity. Representation of space and objects is the key to the theoretical integration of space-based and object-based approaches to attention. The new theory's representation derives from the CODE theory of perceptual grouping by proximity proposed originally by van Oeffelen and Vos (1982, 1983) and extended by Compton and Logan (1993). The CODE theory provides two representations of space, an analog representation of the locations of items and a quasi-analog, quasi-discrete representation of objects and groups of objects. The analog representation of location is produced by bottom-up processes that depend entirely on the proximities of the various items in the display. The representation of objects and groups is produced by an interaction between top-down processes that apply a threshold to the analog representation of locations and the bottom-up processes that generated the analog representation in the first place.

Locations of items are distributed in space. A key assumption of the CODE theory, which contrasts with the implicit assumption in most theories of attention, is that the representation of location is distributed across space. Locations are not points but distributions in 1-D, 2-D, and 3-D space (also see Ashby, Prinzmetal, Ivry, & Maddox, 1996; Maddox, Prinzmetal, Ivry, & Ashby, 1994). The form of the distribution may not matter much, as long as it is roughly symmetrical and peaked in the center. Van Oeffelen and Vos (1982, 1983) originally assumed that the distribution was normal, but Compton and Logan (1993) showed that Laplace distributions worked just as well in accounting for subjects' grouping judgments. I chose the Laplace distribution for the current theory because it is easier to work with than the normal. Like the normal, the Laplace distribution can be defined in more than one dimension. One- and two-dimensional definitions are sufficient for the examples of perceptual organization considered in this article. The probability density function for the 1-D Laplace distribution is:

$$f(x) = \frac{1}{2} \lambda \exp[-\lambda |x - \theta|]. \quad (1)$$

The mean is θ and the standard deviation is $\sqrt{2}\lambda^{-1}$. The mean represents the center of the item in the x dimension, and the standard deviation determines the spread of the distribution over the x dimension.

The representation of locations as distributions is illustrated in Figure 2. The points x , y , and z represent the locations of the items in the x dimension in the display. The dotted lines above each of the points are the distributions that represent the location of the items in the CODE representation. The points x , y , and z are the means of those distributions. The spread of the distributions is determined by the standard deviation, which is the same for all three items in this example (cf. Compton & Logan, 1993).

Representation of spatial array is a CODE surface. CODE assumes that the location of each item in space is represented by its own distribution. Bottom-up processes sum the distribu-

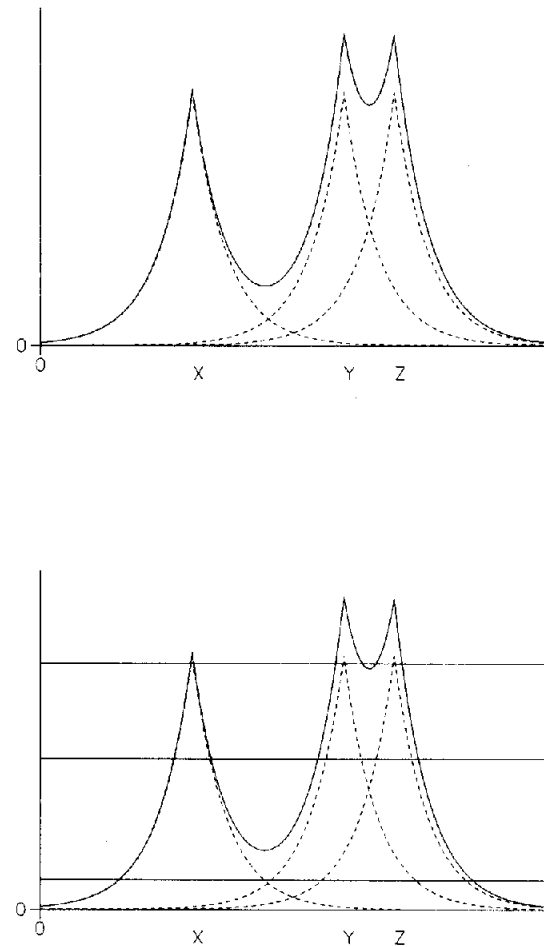


Figure 2. Feature distributions and the CODE surface representing three items (X , Y , and Z) arrayed in one dimension. The top panel shows the feature distributions and the CODE surface; the bottom panel shows three thresholds applied to the CODE surface that parse the display into a three (high threshold), two (intermediate threshold), or one (low threshold) group.

tions for the different items producing a *CODE surface*. The top panel of Figure 2 illustrates how a 1-D CODE surface is generated from items whose locations vary in one dimension. The dotted lines represent the distributions for each item, and the solid line represents the CODE surface, which is the sum of the distributions of locations of the individual items. To formalize this notion, the height of the CODE surface at point x , $h(x)$, is

$$h(x) = \sum_{i=1}^N \frac{1}{2} \lambda_i \exp[-\lambda_i |x - \theta_i|]. \quad (2)$$

for a display of N items.

Figures 3A and 3B illustrate how a 2-D CODE surface is generated from items whose locations vary in two dimensions, such as an array of letters presented in a visual search task. Figure 3A shows the distribution of items (points) in 2-D space, and Figure 3B shows the CODE surface that represents their loca-

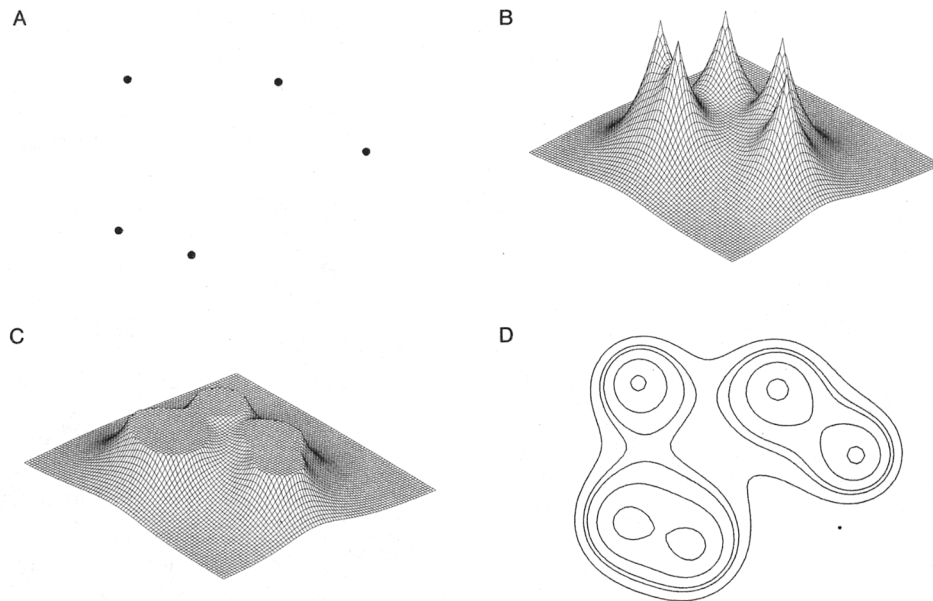


Figure 3. A dot pattern arrayed in two dimensions (3A), the corresponding CODE surface (3B) with a threshold applied to it (3C), and a contour map of the CODE surface (3D) representing all possible groupings of the dots in the pattern.

tions. As in the 1-D case, the distributions for individual items are summed to produce the CODE surface. In this case, the distributions of individual items are 2-D, so the resulting CODE surface is 2-D.

Perceptual groups depend on a threshold applied to CODE surface. Bottom-up processes produce the CODE surface and make it available to top-down processes. Perceptual groups are produced by applying a *threshold* to the CODE surface. The threshold cuts off peaks in the CODE surface, and items residing in the same above-threshold region of the CODE surface belong to the same perceptual group. Items that reside in different above-threshold regions are part of different perceptual groups.

The operation of the threshold is illustrated in the bottom panel of Figure 2. The threshold is a y value that intersects the CODE surface at particular x - y points. An above-threshold region is a range of x values for which the y value of the CODE surface is greater than the y value for the threshold. Items that fall within the range of x values for a given above-threshold region are part of the same perceptual group.

Perceptual grouping is hierarchical. Hierarchical grouping is an inherent property of CODE. It is produced by varying the threshold. The lower the threshold, the larger the groups (i.e., the more items they contain). As the threshold is raised, large groups break up into smaller ones, but the relationship is hierarchical in that smaller groups are always nested within the larger ones. Hierarchical grouping is illustrated in the bottom panel of Figure 2. The lowest threshold value groups all of the items together. The intermediate value breaks the large group into two smaller ones, and the highest value groups each item separately.

The operation of the threshold in the 2-D case is illustrated in Figures 3C and 3D. Figure 3C shows the same surface displayed

in Figure 3B with the peaks "sliced off" the CODE surface by a threshold. Figure 3D shows a contour map of alternative groupings, generated by applying several different thresholds to the CODE surface. As in the 1-D case, grouping is hierarchical, with smaller groups nested in larger ones.

Van Oeffelen and Vos (1982) showed that CODE could account for the subjects' judgments about the appearance of groups in the sorts of stimuli that appear in textbook demonstrations (e.g., a matrix of x 's organized in rows or columns by manipulating proximity). Compton and Logan (1993) showed that several different parametric variations of CODE could account for subjects' judgments of grouping in random dot patterns. The variation used in the current theory—Laplace distributions with equal standard deviations—accounted for grouping judgments as well as any other. Compton and Logan (1996) examined the invariance of grouping judgments over transformations of size and orientation, which CODE predicts, and found that subjects' judgments were not invariant. They were close, however, and CODE provided a reasonable description of the data.

Extension of CODE to Attention

CODE distributions are distributions of item features. The application of CODE to attention involves a straightforward extension of the assumption about representation of location: Location is distributed in the sense that information about the features of the items is distributed over space. The distributions that make up the CODE surface are distributions of item features. The height of a distribution at any point in space represents probability (density) of sampling the features of the item it represents. Given the shape of the assumed distribution (Laplace), the probability of sampling features will be highest

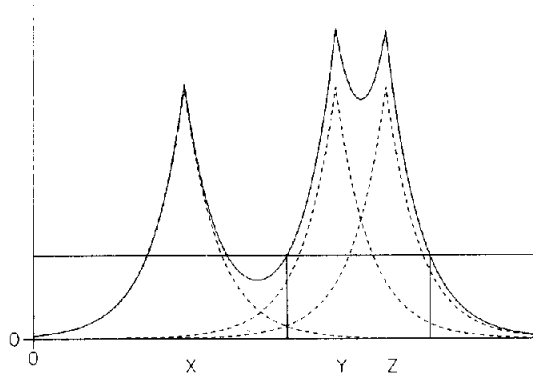


Figure 4. Illustration of the feature catch produced by applying a threshold to a CODE surface representing three items (X , Y , and Z). (Note that part of the feature distribution for item X is included in the feature catch for items Y and Z .)

near the center of the item. It will drop off exponentially as distance from the center of the item increases.²

The assumption that information about features are distributed over space is similar to assumptions made by Wolford (1975), Ratcliff (1981), Maddox et al. (1994), and Ashby et al. (1996) to account for spatial factors in visual tasks. The assumption can be articulated in terms of the receptive fields of feature detectors in visual cortex: If an item falls in the center of a receptive field, the detector will respond strongly to it. If the item falls near the edge or on the edge of a receptive field, the detector will respond less strongly. A given item will stimulate several feature detectors, some in the center and some near the periphery. The representation of the item's features is distributed over space in the sense that the detectors that respond to them are distributed over space.

The CODE surface also represents distributions of features. It represents the sum of the distributions of the features of all the items in the display. The height of CODE surface at any point in space represents probability (density) of sampling features of all of the items whose distributions intersect at that point.

Attention selects CODE-defined objects. The theory assumes that attention chooses among perceptual objects in the sense that it chooses among above-threshold regions. It assumes that attention samples the features that are available within the above-threshold region. The features of different items falling within the above-threshold region are sampled with a probability equal to the area of the distribution of the item that falls within the above-threshold region; this probability of sampling features is called the *feature catch*. Figure 4 depicts a three-item display with a threshold set so that two of the items are grouped together, in that they both fall within the above-threshold region. The probability of sampling features from those items (the feature catch for those items) is high because large parts of the distributions that represent them fall in the above-threshold region. Note, however, that part of the distribution for the third item that is not grouped together with the other two nevertheless falls within the above-threshold region. Features of that item will be sampled along with features of the items within the

group with a probability proportional to the area of the distribution in the above-threshold region (i.e., the feature catch for the third item is greater than zero).

To formalize this idea in the 1-D case, the feature catch for item z , $c_{z|T}$, for a given threshold, T , may be defined as

$$c_{z|T} = \int_{l_0}^{h_i} \frac{1}{2} \lambda_z \exp[-\lambda_z |x - \theta_z|] dx, \quad (3)$$

where l_0 and h_i represent the limits of the above-threshold region. The sample that is taken from the above-threshold region and subjected to later processing (e.g., TVA), is the sum of the N individual feature catches.

$$\begin{aligned} \text{Sample} &= \sum_{z=1}^N c_{z|T} \\ &= \sum_{z=1}^N \left(\int_{l_0}^{h_i} \frac{1}{2} \lambda_z \exp[-\lambda_z |x - \theta_z|] dx \right). \quad (4) \end{aligned}$$

The idea of the feature catch makes CTVA like object-based theories of attention. Object-based theories assume that all properties of a selected object are processed (Kahneman & Treisman, 1984; Treisman, 1969); in CTVA, the above-threshold region corresponds to the selected object, and all features available in that region are sampled. However, CTVA is unlike object-based theories in that features of items outside the selected perceptual group are also sampled with some nonzero probability. The sample that is subjected to later processing contains features of all of the items in the display, not just those in the selected group.

The CODE theory of visual attention is like space-based theories in that it assumes that features are sampled from items other than the one that is the current focus of attention. Both assume a kind of "fuzziness" in the processing, so that features of unattended items intrude in the processing of attended items. However, the fuzziness lies in different parts of the system. Space-based theories of attention assume that the boundary of the sampled region—the edge of the spotlight—is fuzzy (e.g., Eriksen & Eriksen, 1974; Eriksen & Hoffman, 1973). This idea is explicit in LaBerge and Brown's (1989) gradient theory. By contrast, in CTVA, the boundary of the attended region is sharp and the rep-

² At this point in the development of the theory, I do not wish to draw a strong distinction between items and features, so I will treat the feature distributions in CODE as distributions of individual features and as distributions of the entire set of features that belong to an item. Many current theories of visual search propose separate spatial maps for the individual features of an item—one map for redness, one for vertical lines, and so on—with a master map of item locations that can be used to address the individual features of an item (e.g., Cave & Wolfe, 1990; Treisman & Gelade, 1980; Treisman & Sato, 1990; Wolfe, 1994; Wolfe, Cave, & Franzel, 1989). In principle, CTVA could be applied to the individual feature maps or to the master location map. The mathematics would be the same in either case, and as long as the (spatial) variability of the distributions was the same for different features, the predictions would be essentially the same. However, there is nothing in CTVA that forces the same variability on distributions for different features and it could be fruitful to use CTVA to explore the idea of multiple maps and multiple CODE surfaces. That exploration is beyond the scope of this article.

representation of items in space is fuzzy. Unattended items intrude on the processing of attended ones because their representations are distributed across space and fall within the (sharply circumscribed) above-threshold region that attention samples.

Thresholds, variability, and the feature catch. CTVA provides later processes with a sample of features to process. The probabilities of sampling features from particular items (the feature catches for those items) depend on the proximities of the items in the display, the variability of the feature distributions, and the threshold applied to the CODE surface. The proximities are determined outside the theory by the experimenter or the external world. The variability scales the proximities. CTVA assumes that the variability of the feature distribution is the same for all items in the display. Variability is manipulated as a parameter of the model. Increasing variability has two effects on the feature catch: It decreases the contribution of items within the group to the feature catch (by decreasing the area of their feature distributions that falls within the limits of the above-threshold region), and it increases the contribution of items outside the group but nearby (by increasing the area of their feature distributions that falls within the limits of the above-threshold region).

The threshold is manipulated as another parameter of the model. Increasing the threshold decreases the magnitude of the feature catch, decreasing the contribution of items inside and outside the above-threshold region. The effects can be seen in Equations 3 and 4. Increasing the threshold amounts to decreasing the range of the limits of integration, including less of the distribution in the sample.

The CODE surface and the feature catch. The local minima or "saddle-points" on the CODE surface are important because they represent the boundary between grouping and separating sets of items. If the threshold is higher than the local minimum, the items will break into two (or more) groups. If the threshold is lower, the items will cluster into one group (Compton & Logan, 1993). For CTVA, this represents a boundary between serial and parallel processing: If the threshold is higher than the local minimum, groups of items can be processed one at a time. If the threshold is lower than the local minimum, the items are grouped together and must be processed together.

The effect of threshold variation on the feature catch is illustrated in Figure 5. The top panel represents the feature distributions and CODE surfaces for three items—a central target and two flanking distractors (e.g., Eriksen & Eriksen, 1974). The left panel represents items placed closer together than the right panel. The middle panel plots the area in the feature catch for the central target and the sum of the areas in the feature catch from the two flanking distractors as a function of threshold setting, going from low on the left to high on the right. The total volume of the feature catch decreases as the threshold increases. At low threshold values, below the local minimum, information is sampled from the whole display and the contribution from the distractors outweighs the contribution from the target by a substantial margin. At high threshold values, above the local minimum, attention is focused on the central target item and information from the target outweighs information from the distractors.

The impact of these effects on the feature catch can be seen in the bottom panel of Figure 5, which plots the ratio of the

feature catch for the central target to the feature catch for the sum of the distractors—a signal-to-noise ratio. The signal-to-noise ratio is less than 1.0 and approximately invariant for low thresholds smaller than the local minimum, but it jumps abruptly at the local minimum to a value above 1.0 and grows substantially as the threshold increases further.

There is a tradeoff between the magnitude of the signal and the quality of the signal: Bigger signals have lower signal-to-noise ratios; signal-to-noise ratio can be increased only by decreasing signal magnitude. In CTVA, the tradeoff is masked because signal magnitude and signal-to-noise ratio are both positively related to speed and response probability. Thus, different combinations of signal magnitude and signal-to-noise ratio can produce the same reaction time and accuracy. The tradeoff is bounded by a sharp discontinuity (in the curve in the bottom panel of Figure 5) at the point at which the threshold equals the local minimum. In fitting the theory to data, I found that the model performed similarly at all threshold values below the local minimum and similarly (but differently) at all threshold values above the local minimum. The largest difference occurred when the threshold crossed the local minimum.

Spatial indexing. The local minima between items or groups is lower the further apart the items or groups are. Conversely, the local minimum increases as items or groups get closer together. This can be confirmed by inspecting Figures 2 and 3. The local minimum represents the lowest threshold value at which an item or group can be separated from the rest, and that threshold value is different in different parts of a display, increasing with the density of items in the display. The question for theory is whether the system maintains a single threshold for all of the items in the display or takes advantage of these differences and allows several different thresholds to operate at once.

There should be no problem with multiple thresholds if the different items or groups are processed serially. The system could reset the threshold before each serial inspection. Still, the system must keep track of which items or groups have been processed, and in doing so, it might also keep track of the threshold level associated with each item or group. If items or groups are processed in parallel, there must be some way of keeping track of thresholds and keeping track of which item or group goes with which threshold. Between-object parallel processing has the same problem with multiple groups even if the threshold value is the same for each group. The system must keep track of which group is which. Keeping track of which threshold went with which group should not be much more difficult.

Spatial indices are often proposed as a solution to the problem of keeping track (Pylyshyn, 1984, 1989; Trick & Pylyshyn, 1994; Ullman, 1984). Spatial indices provide an identity-neutral way of referring to perceptual objects, and it seems reasonable to attach things like threshold values to the spatial indices. The idea is similar to Kahneman and Treisman's (1984; Kahneman et al., 1992) idea of an *object file*: I am proposing a temporary episodic representation of an object that includes an index to the perceptual representation of the object and information about the threshold value at which the object was defined. The object file serves as a referent to which other information can be attached, such as the identity of the object or some other categorization.

Ultimately, the answer to the question of whether the system can support multiple thresholds is "yes, but with a cost." The

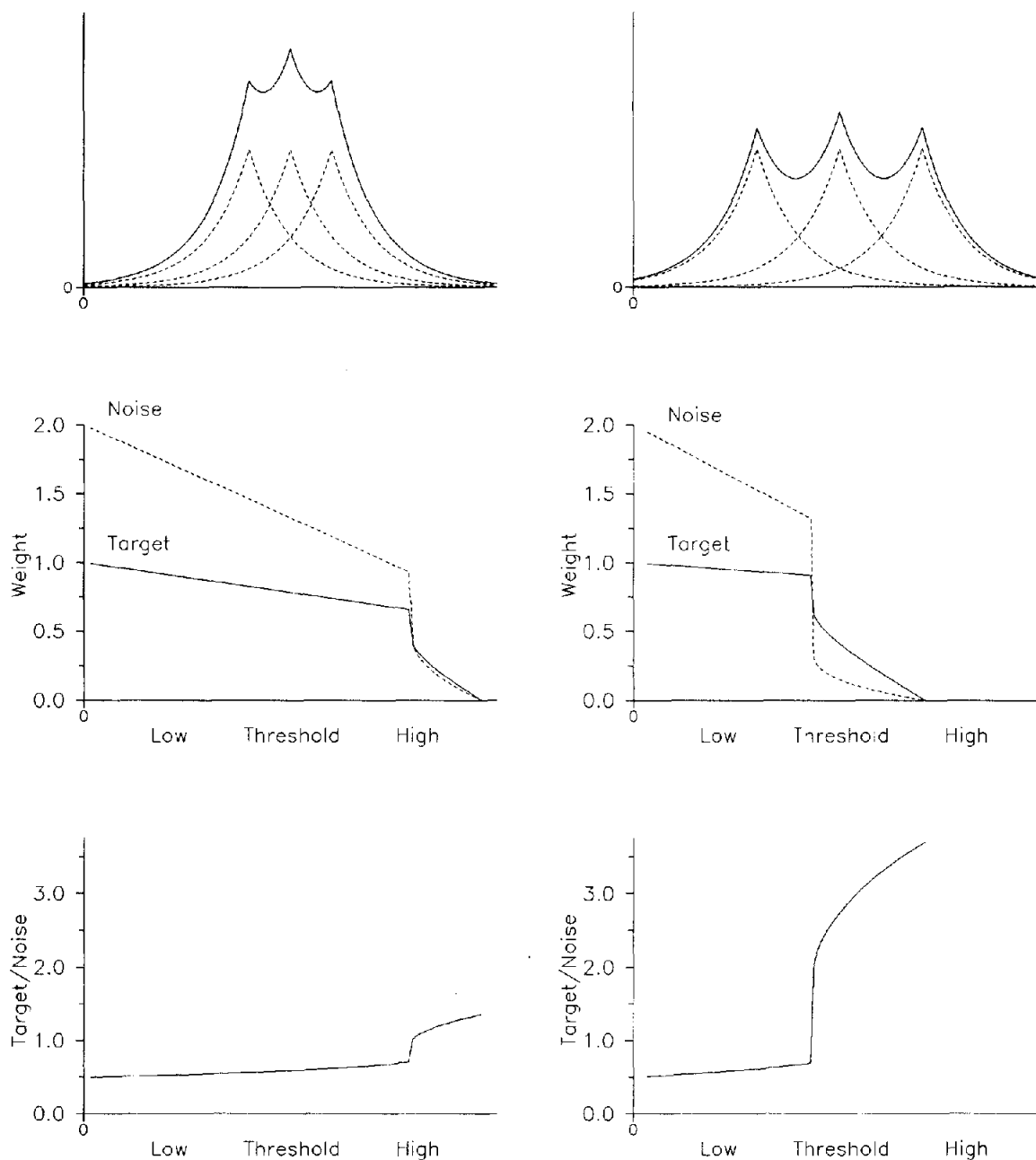


Figure 5. CODE surfaces representing a central target item and two flanking distractors (top panels) placed near (left panels) or far (right panels) from the target; magnitudes of the feature catches (areas under the above-threshold regions) for targets and distractors as a function of threshold (middle panels); and signal-to-noise ratios reflecting the ratio of the target feature catch to the distractor feature catch as a function of threshold (bottom panels). (Note the discontinuity in the bottom and middle panels when the threshold increases above the local minima that separate the target from the distractors. Signal-to-noise ratio increases markedly after the discontinuity.)

cost is that it must have some way of implementing a spatial indexing process and it must implement an episodic memory that keeps track of objects' locations and the spatial resolution (threshold value) at which they were seen. There are many reasons for proposing a system with a capacity for spatial indexing and episodic storage. Keeping track of different threshold values in CTVA is another one.

Bundesen's (1990) Theory of Visual Attention

The CODE theory of visual attention still does not deal with within-object or within-region selection. CODE provides the input to subsequent selection mechanisms. CODE says that the input consists of the sum of the feature catches from all of the items whose distributions fall in the above-threshold region, but

CODE does not say how that input is processed. I adopted Bundesen's (1990) theory of visual attention (TVA) as the recipient of the input that CODE provides.

In many ways, Bundesen's (1990) TVA model is an ideal match for CODE. They are both formal theories, and their mathematics are compatible. They address phenomena at the same level of abstractness, focusing, for example, on the idea that items are composed of features without specifying the nature or the number of the features. Most importantly, they are compatible in that CODE provides as output what TVA takes as input—a sampling of visual features. The original CODE changed for the wedding, extending its assumptions about the distribution of location to include the idea that the distributions were distributions of feature values. TVA must change for the wedding as well by altering its assumption about the representation of location.

Basic TVA. Bundesen (1990) conceived of TVA as a model of selection, intended to explain the process by which people choose among the inputs confronting them. TVA evolved over several years, beginning with an attempt to model selection in partial report tasks (Bundesen, Pedersen, & Larsen, 1984; Bundesen, Shibuya, & Larsen, 1985; Shibuya & Bundesen, 1988). Bundesen (1987) generalized the model of partial report to a fixed-capacity and independent race model of selection that was the direct ancestor of TVA. Since 1990, Bundesen generalized the model further, showing that fixed- and unlimited-capacity versions of TVA correspond to Luce's (1959) choice model of selection (Bundesen, 1993). The 1990 version addresses attention most directly and most generally, so that was the version I married to CODE. The mathematical details of TVA are described and explained in Appendixes A and B.

Basically, TVA chooses among categorizations of perceptual inputs. TVA assumes two levels of representation: (a) a perceptual level that consists of features of display items; and (b) a conceptual level that consists of categorizations of display items and display features. The two representations are linked by a parameter $\eta(x, i)$, which represents the amount of sensory evidence for membership in category i that comes from item x . The greater the $\eta(x, i)$, the more likely x is to belong to category i . The $\eta(x, i)$ reflects the bottom-up component of TVA. They are determined entirely by the quality of the data and the set of categories.

Variable x is an index for a display item, representing one member of a set, S , of display items. Variable x is a symbol at the categorical level of representation that stands for a sample of information from the perceptual representation (i.e., a perceptual item). Variable x does not represent the location of the item. In Bundesen's (1990) theory, location is just another categorizable feature of the item, like color or form.

Variable i represents a particular categorization for x . It could be "red," "square," or "located in the top left corner." The variable i represents one member of a set, R , of possible categorizations. There is an η value for each combination of item and categorization, reflecting the strength of perceptual evidence that each x belongs to each i .

TVA selects among perceptual items and categorizations by choosing a particular categorization for a particular item (or particular categorizations for K items). The choice is determined by the outcome of a race between the alternative categorizations,

with the first one to finish being the one that is selected (or the first K to finish). It is important to note that two things are selected simultaneously by the outcome of the race: (a) a perceptual item and (b) a categorization for it. Thus, TVA is both an early selection theory and a late selection theory. It is an early selection theory in that items are not identified before they are selected; it is a late selection theory in that items are selected on the basis of their identities, on the basis of the categorization that wins the race (Bundesen, 1990).

Strength of perceptual evidence. The η values are important determinants of the outcome of the race. η values, modified by two attentional parameters, determine the rate at which the categorizations that correspond to them are processed. Thus, $\eta(x, i)$ determines the rate at which x is categorized as an i . The larger the η value, the faster the process. Other things equal, the categorization with the largest η value is most likely to win the race. It is likely to be fastest and thus finish first. However, the race is stochastic. Ultimately, η values represent the rate parameters in exponential distributions of finishing times for the different categorizations (see Appendix A). The race is between the exponential distributions, and no one of them is guaranteed to finish first.

Bundesen (1990) formalized the race model by specifying the rate of categorization, $v(x, i)$ in terms of $\eta(x, i)$ and two attentional parameters in the following equation:

$$v(x, i) = \eta(x, i) \beta_i \frac{w_x}{\sum_{z \in S} w_z} \quad (5)$$

Perceptual bias. In Equation 5, η values are modified by two kinds of attentional weight, β s and w s. β_i reflects the person's bias to categorize the display as i . β_i is a bias because it raises the probability that the first categorization will be i but it does not change the likelihood that any given item (e.g., x) will win the race (i.e., be the first item categorized as i ; see Appendix B). Note that the β_i values are under the person's control (i.e., the homunculus). They can be varied to control the categorization process, to determine which way the display is categorized. The display is likely to be categorized as i if β_i is high, so desired analyses can be selected by raising β_i values. Moreover, the display is unlikely to be categorized as j if β_j is low, so undesired or irrelevant categorizations can be turned off by setting their β_j values low. (The default assumption is that β_i is low unless category i is relevant.)

Attentional weights and priority. The variable w_x reflects the attentional weight on item x . It is an attentional weight because increasing its value makes it more likely that item x will be categorized, but it does not change the likelihood of any particular categorization of x (see Appendix B). Thus, it provides a way of focusing in on item x in the display. According to Bundesen (1990), the attentional weight, w_x , is determined by the following equation:

$$w_x = \sum_{j \in R} \eta(x, j) \pi_j \quad (6)$$

The new term in Equation 6 is π_j , which represents the priority of attending to items that belong to category j . Like β , π can be set by the "person" (homunculus). The variable π works

together with η to determine the attentional weight. The item with the largest combination of η and π receives the greatest weight. That item is the most likely to contain relevant information: A high value of η suggests that it contains much information, and a high value of π suggests that the information is pertinent. Thus, the homunculus can use TVA to focus in on relevant items by controlling the π values. Setting π_i high makes items in category i more likely to be selected. Setting π_i low makes items in category i less likely to be selected. Combining the π values with η values focuses attention on the most informative items.

Predicting accuracy and reaction time. The ease with which predictions can be derived from TVA is one of its nicest features. The predictions follow from Bundesen's (1990) interpretation of the $v(x, i)$ values as rate parameters in exponential distributions. The exponential distributions are functions of time, and the rate parameter determines the mean finishing time. Moreover, Bundesen (1990) assumed that the different exponential distributions race against each other, and the race model allows predictions of response probability. Accuracy depends on the relative magnitudes of the rate parameters. The connections between the $v(x, i)$ values and exponential distributions are given in Appendix A.

TVA can predict reaction time and accuracy in several ways. The simplest, which Bundesen (1990) used most often, involves a simple race in which the first categorization is the one that is selected. Accuracy is the probability of choosing the appropriate categorization first. The probability that categorization " x belongs to i " finishes first is computed by taking the ratio $v(x, i)$ to all of the v 's in the display:

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

Mean reaction time is simply the mean finishing time of the winner of the race plus some additive constant, b , that represents stimulus and response processing. The rate parameter of the distribution of finishing times for the winner of the race is the sum of the $v(x, j)$ values for each x in S and each j in R , and the mean of the distribution is the reciprocal of its rate parameter (see Appendix A); that is,

$$RT_{(x,i)} = \frac{1}{\sum_{z \in S} \sum_{j \in R} v(z, j)} + b. \quad (8)$$

Predictions about accuracy and reaction time depend on the three parameters that determine the v values, expressed in Equations 5 and 6: η , the strength of sensory evidence, β , perceptual bias, and π , pertinence. In wedding TVA to CODE, one more parameter is added, that represents the proportion of the feature catch corresponding to each item. Adding that parameter to the model requires some changes in TVA's assumptions about the representation of location.

These predictions about reaction time and accuracy assume that the response is determined by the first categorization—a simple race between the alternative categorizations. However, TVA can support more than a simple race. TVA can be config-

ured as a *counter model*, in which K_i categorizations of type i must be made before the person responds with " i ." Reaction time would depend on an additive constant, b , and the time required to make K categorizations. The time to make K categorizations can be computed using a standard Poisson counter model (Townsend & Ashby, 1983). A counter model would be useful in a situation in which a race model produced less than ideal accuracy (say 80%). The imperfect accuracy could be improved by sampling repeatedly and accumulating the results of the sampling.³

To make the predictions concrete, consider a case in which a person discriminates between an H and an S . The counting model has two counters, one for H and one for S . There is a criterion number of counts for each counter, K_H and K_S , and the process terminates when the criterion number of counts accumulates in one counter or the other. The probability of responding correctly (i.e., responding " H " when the target was an H) is

$$P(R_H|S_H) = \sum_{j=0}^{K_S-1} \binom{K_H+j-1}{j} \left(\frac{v(x, H)}{v(x, H) + v(x, S)} \right)^{K_H} \left(\frac{v(x, S)}{v(x, H) + v(x, S)} \right)^j, \quad (9)$$

and mean reaction time for correct responses is

$$\bar{T}_{H|H} = \left\{ \left[\sum_{j=0}^{K_S-1} \binom{K_H+j-1}{j} \left(\frac{v(x, H)}{v(x, H) + v(x, S)} \right)^{K_H} \left(\frac{v(x, S)}{v(x, H) + v(x, S)} \right)^j \left(\frac{K_H+j}{v(x, H) + v(x, S)} \right) \right] \frac{1}{P(R_H|S_H)} \right\} + b. \quad (10)$$

The counter model is a straightforward generalization of the original race model in TVA. The counter model involves a race between the H counter and the S counter. Accuracy depends on the probability that the H counter finishes first, given that " H " was presented, and the reaction time depends on the time taken to accumulate K_H counts. The H counter can finish first if K_H counts accumulate in it before K_S counts accumulate in the S counter. The S counter can accumulate $j = 0$ to $K_S - 1$ counts before the H counter accumulates K_H counts, and the H counter will still win the race. The j counts can accumulate in the S counter in many ways. The first S count could occur before the first H count, before the second H count and so on. The actual number of ways j counts could accumulate is given by the binomial expression in the first term of Equation 9. The probability that the H counter will increment is given by the ratio $v(x, H)/$

³ Bundesen (personal communication, February 1995) considered implementing TVA as a counter model in his original conception of the model to account for speed-accuracy tradeoff effects and the like. However, the main focus of his theorizing was on simple detection tasks, in which one look at the stimulus would suffice, and on partial report tasks, in which the main focus was on the probability that the first K categorizations to finish were part of the cued subset, so he left the counter model for future development.

$[v(x, H) + v(x, S)]$ in the second term of Equation 9. The probability that that K_H counts will accumulate in the H counter is given by raising $v(x, H)/[v(x, H) + v(x, S)]$ to the K_H th power, as is shown in the second term in Equation 9. The probability that the S counter will increment is given by the ratio $v(x, S)/[v(x, H) + v(x, S)]$ in the third term of Equation 9. The probability that $j = 0$ to $K_S - 1$ counts accumulate in the S counter is given in the third term of Equation 9, by raising $v(x, S)/[v(x, H) + v(x, S)]$ to the j th power. The three terms in Equation 9 combine to produce the probability that the H counter will finish first given that "H" was presented, which measures response accuracy. Equation 10 includes these three terms plus a fourth that is the mean of a Gamma distribution for the time required to reach $K_H + j$ counts. The four terms in the numerator of Equation 10 are divided by $P(R_H | S_H)$ to yield mean counting time conditional on making a correct response, and an intercept constant b is added to yield mean reaction time. For further details on Poisson counting models, see Townsend and Ashby (1983, pp. 272–280).⁴

The race model is a special case of the counter model in which the counting process terminates when the first runner finishes (i.e., $K_H = K_S = 1$, so only the first runner is counted). The counter model adds two more parameters to TVA— K_H and K_S —for a total of five. In many applications it is reasonable to set the criteria equal to each other, so that only one more parameter is required beyond the three in the original TVA.

The predictions for the counter interpretation of TVA are more complicated than the predictions for the simple race model, but ultimately, they still depend on the $v(x, i)$ values, and those values depend on η , β , and π , which are at the heart of TVA. The predictions depend on the rate of counting, and distribution of intervals between counts is the same exponential distribution that governs the simple race model. Predicted reaction times will be longer in a counter model than in a simple race, because the process in the simple race has to iterate at least K times. Predicted accuracy will be higher as well, because of the repeated sampling. But reaction time and accuracy depend on the same factors—the $v(x, i)$ —in both interpretations.

CTVA is largely agnostic with respect to the process by which TVA determines reaction time and accuracy. Its main purpose is to describe the processes that give input to TVA and how that input is modulated by the perceived spatial organization of the display. Response-related processing is important, because the input, modulated and processed, has to produce a response to be measured, but it is not a central factor in the theorizing. Either the race or the counter interpretation could serve my purposes.

The Wedding of CODE and TVA

The wedding of CODE and TVA is straightforward: CODE provides the input to TVA. CODE's feature catch provides the sensory data that defines the η values in TVA. TVA provides the β and π values that allow selection of an appropriate response. The feature catch is run through Equations 5–10 to provide predictions of reaction time and accuracy. CODE and TVA become CTVA.

Feature catch weights sensory evidence. The feature catch modifies the strength of sensory evidence from the various

items in the display. Items that fall within the perceptual group from which the feature catch is sampled will contribute a great deal of sensory evidence. Items that fall outside but nearby will contribute some sensory evidence, but less than the amount contributed by items within the group. Items far from the group will contribute very little sensory evidence. Thus, attention is focused primarily on the members of the selected group and to a lesser extent on their near neighbors.

From a formal perspective, the feature catch from item x , defined in Equation 3 modifies the $\eta(x, i)$ values, multiplying them by a number, c_x , between 0 and 1.0 that depends on the area of the distribution of x that falls within the above-threshold region, that is, $\eta(x, i)c_x$. Thus, the attentional weights, w_x , become

$$w_x = \sum_{j \in R} \eta(x, j) \pi_j c_x, \quad (11)$$

and $v(x, i)$ becomes

$$\begin{aligned} v(x, i) &= c_x \eta(x, i) \beta_i \frac{w_x}{\sum_{z \in S} w_z} \\ &= c_x \eta(x, i) \beta_i \frac{\sum_{j \in R} \eta(x, j) \pi_j c_x}{\sum_{z \in S} \sum_{j \in R} \eta(z, j) \pi_j c_z}. \end{aligned} \quad (12)$$

If all of the η s, β s, and π s are equal to 1, Equation 12 reduces to

$$v(x, i) = c_x \frac{c_x}{\sum_{z \in S} c_z}. \quad (13)$$

According to Equation 13, $v(x, i)$ depends on the ratio of the feature catch for x to the sum of the feature catches from all of the items in and nearby the selected perceptual group (i.e., the sum of the feature catches within the selected above-threshold region). Inserting Equation 13 into Equations 7–10 shows that reaction time and accuracy also depend on the ratio of the feature catches. Control of the feature catches—by controlling the threshold—is an important function in the new theory.

How did TVA change? Items play a different role in CTVA than they do in TVA. TVA treats items as discrete units. Items can be selected individually, and sensory evidence and attentional weights are attached to items. By contrast, CTVA treats items as spatial distributions and attaches sensory evidence and attentional weights to parts of those distributions. CTVA selects perceptual objects, not items. Perceptual objects may be made of several items, and a given perceptual object may contain information from adjacent items that do not belong to it. The

⁴ Nosofsky and Palmeri (in press) present an exponential random walk model as an alternative to the Poisson counter model. The main difference is that, in the random walk model, evidence for one alternative is evidence against the other, whereas in the counter model, evidence for the different alternatives accumulates independently. There is nothing inherent in CTVA that would lead one to choose a Poisson counter model over an exponential random walk model; CTVA could be configured either way. I present the counter version here because the mathematics were easy to derive.

information in a perceptual object is a blend of the information about items inside and nearby the object.

The representation of location information in CTVA is much more complex than it is in TVA. TVA treats location as an attribute of an item, just like color and shape. Location is a category, like color and shape are categories, and items can be selected by location by increasing the priority for the desired location. CTVA represents location in several ways. First, it retains TVA's notion of location categories but it relies much less heavily on them, generally not distinguishing item locations within a perceptual group. Second, in CTVA, location is a factor in the perceptual representation of the display, in that item locations are represented as distributions over space (cf. Ashby et al., 1996; Maddox et al., 1994). Third, the locations of perceptual objects are represented in the set of groups constructed by applying a threshold to a CODE surface. And fourth, locations of groups relative to each other may be represented conceptually by predicates like *above* (x, y) that express categorical spatial relations (Logan & Sadler, 1996). Location can be selected in the third and fourth senses by applying visual routines that are outside the current model (Cave & Wolfe, 1990; Koch & Ullman, 1985; Logan, 1995).

Finally, CTVA can process displays in parallel or in series, whereas TVA processes only in parallel. In CTVA, processing within perceptual groups is parallel. Processing between perceptual groups can be serial or parallel, depending on the task and the situation. In this respect, CTVA is midway between theories like TVA that process all items at once and theories like Treisman's feature integration theory (Treisman & Gelade, 1980) that process items one at a time. CTVA is like the theories of Treisman and Gormican (1988), Duncan and Humphreys (1989), Humphreys and Müller (1993), and Grossberg, Mingolla, and Ross (1994) in that it processes parts of the display in parallel and parts in series, but it differs from those theories in how it defines the parts.

From a formal perspective, the difference can be understood in terms of weights on the items, c_x : TVA assumes that c_x equals 1.0 for all the items in the display, and serial processing theories assume that c_x equals 1.0 for the currently selected item and 0.0 for all other items. CTVA assumes that c_x is distributed unevenly between 0.0 and 1.0 over all the items in the display, with the value depending on the area of the item's feature distribution that falls within the sampled region.

Parallel and serial processing are both possible in CTVA, depending on the threshold applied to the CODE surface. As depicted in the bottom panel of Figure 2 and in Figure 3D, a low threshold includes all the items in one group. The areas of the different items' feature distributions are approximately equal, so the weight on each item is approximately equal, as in parallel processing. A high threshold picks off the peak of one of the items in the display. The feature catch under that threshold weights the central item heavily and adjacent items lightly, approximating the all-or-none distribution of weights in serial processing. Raising the threshold from low to high changes the emphasis from parallel to serial processing.⁵

Thus, the original TVA model is a special case of CTVA, in which the entire distribution for every item in the display enters into the sample. Each item has equal (and maximum) weight, so CODE drops out of the picture and performance depends

entirely on TVA. In principle, CTVA can be compared against the special case of TVA to see the extent to which modulating the input with CODE improves prediction. One could ask whether TVA needs CODE, and this comparison will answer the question: To the extent that CODE improves the predictions, TVA needs CODE.

One could also ask whether CODE needs TVA, and the answer is clearly "yes," if CODE is to account for attention. TVA provides CODE with the capacity for within-object selection and for response generation, which CODE lacks. TVA does a lot of the work in the fits of the model to data, with three or four parameters to CODE's two. TVA could improve other models of visual spatial attention that do not specify means of within-object selection or response selection (e.g., Eriksen & St. James, 1986; Kahneman et al., 1992). Thus, when comparing CTVA's ability to account for data against other theories of visual spatial attention, we must be careful to distinguish between what CODE predicts uniquely and what TVA would do for any other theory it interfaced with. To facilitate the distinction, I focused on tasks that emphasized spatial factors that CODE accounts for rather than TVA.

Between-group selection. CODE provides TVA with several different perceptual groups to sample, at intermediate threshold levels. I assume that in some cases, TVA is applied to one perceptual group at a time, processing the display serially, focusing on one above-threshold region and then another. The processes that govern the selection of above-threshold regions for processing are outside the scope of CODE and TVA. Not much is known about them (but see Cave & Wolfe, 1990; Koch & Ullman, 1985; Logan, 1995). Nevertheless, they must take time and that time must contribute to the effects that appear in human performance. They complicate the predictions of the theory but perhaps not enough to make it intractable. I will attempt to separate effects due to within-object selection from effects due to between-object selection in the CTVA analysis of attentional phenomena.

I assume that in other cases, TVA is applied to all of the perceptual groups simultaneously. The different groups race against each other, and the winner is selected. The winner has two components, a categorization, i , into one of the R response categories, and an index, x , that distinguishes the winning group from the other ones in the display. The race selects both the response and the perceptual group that gave rise to the response. The index, x , is important because it can be interpreted computationally as a *spatial index* (Pylyshyn, 1984, 1989; Trick & Pylyshyn, 1994; Ullman, 1984). Spatial indexing is required

⁵ Note, however, that two kinds of parallel processing are possible in CTVA—within objects and between objects. Within-object parallel processing is a necessary consequence of CTVA's assumptions about the sampling of features. All items that fall within the above-threshold region will be processed in parallel. This is true regardless of the value of the threshold, whether it is low enough to include all of the feature distributions of all of the items or high enough to emphasize one item over the rest. Between-object parallel processing depends on assumptions about between-object selection that lie mostly outside of CTVA. Different perceptual groups can be processed in parallel by assigning a separate spatial index to each group (cf. Pylyshyn, 1989; Trick & Pylyshyn, 1994).

in the serial processing version of CTVA. The spatial index, x , keeps track of the current object of attention.

Spatial indices are discussed extensively in the literature on visual cognition (e.g., Pinker, 1984), where, among other things, they are proposed as a solution to the *binding problem* (see Pylyshyn, 1984; Treisman & Gelade, 1980; Trick & Pylyshyn, 1994; Ullman, 1984). A spatial index is a symbol (e.g., x or g) that corresponds to a perceptual object. The symbol acts as an address for the perceptual object, in that it provides a means by which processes that operate on the symbol can access perceptual information about the object to which it refers. The spatial index distinguishes its referent from the alternatives without conferring a particular identity on it. The system simply knows "a thing is there" without knowing what the thing is. Once it knows this, it can ask other questions about it (e.g., "Is it red?" "Is it a T?" "Is it above that other thing?") by accessing the perceptual information it contains (i.e., by applying TVA).

In principle, CTVA should interface nicely with theories in which spatial indexing is an important process. The CODE part of the theory defines the objects that can be spatially indexed, and the TVA part of the theory defines the processing that is done on the indexed objects. Theories of spatial indexing have to explain the processes that choose among perceptual objects.

CTVA may not explain selection between groups, but it interfaces nicely with Logan's (1995) theory of linguistic and conceptual control of attention that accounts for the direction of attention from cues to targets. CODE interfaces with Logan's (1995) theory in the same way it interfaced with TVA: It provides the input that Logan's (1995) theory needs to process.

The inputs to Logan's (1995) theory are schematic representations of objects as points, lines, surfaces, and regions, and these are what CODE provides. The perceptual objects defined by applying a threshold to a CODE surface serve nicely as inputs to Logan's theory. Logan's choice of schematic representations for the input to his theory was motivated by linguistic analyses of the semantics of spatial relations, on which Logan's theory relies heavily. According to linguistic analyses, the spatial relations expressed in language (e.g., those expressed by prepositions in English) schematize the objects they take as arguments, so that a small number of relations (roughly 80 in English) can apply to an indefinitely large number of objects (Clark, 1973; Herskovits, 1986; Jackendoff & Landau, 1991; Talmy, 1983; Vandaloise, 1991).

Logan's (1995) theory involves two representations—a *perceptual representation* of the layout of objects and surfaces and a *conceptual representation* of propositions that express spatial relations between objects. Directing attention from one perceptual object to another involves apprehending the spatial relations between the objects, and that involves coordinating the two representations. Coordination requires two more representations and the processes that operate on them: (a) a *reference frame* that defines an origin, orientation, direction, and scale in perceptual space and (b) a *spatial template* that represents the different regions of acceptability associated with the relation.

Apprehending a relation like *above*(x , y) involves the following steps: (a) finding the perceptual object corresponding to y , (b) imposing the reference frame relevant to the relation (*above*) on the perceptual object corresponding to y , (c) aligning the spatial template for *above* with the reference frame cen-

tered on y , and (d) determine whether x falls in a good or bad region of acceptability relative to the template centered on y .

Cuing attention—directing attention from one object to another—involves the same four steps that were just described. The cue is y , the target is x , and the relation is (typically) *next_to*(*target*, *cue*). Attention is directed to the cue (step a) and then from the cue to the target (steps b–d). Once attention is on the target, the target itself can be processed (i.e., with TVA).

Basic Architecture Revisited

Figure 6 represents the complete version of the sketch of the basic architecture. In the early visual processes, location and identity are bound together in the feature distributions and the CODE surface. The locations of items are given by the environment and the spread of features from the items is determined by the CODE λ parameter. The threshold parameter parses the display into perceptual groups that serve as input to the later processes. From the perspective of the late identity system, the threshold defines the feature catch from each item in the display. From the perspective of the late location system, the threshold defines a perceptual organization for the display.

TVA is the late identity system. It takes the feature catch as input and computes the strength of sensory evidence for the categories relevant to the response alternatives—the η values. The η values, modulated by bias (β) and pertinence (π), determine the probability and the latency with which different categorizations—identities—are selected.

The late location system is represented less completely and much less formally by Logan's (1995) theory of conceptual direction of attention. It takes as input the perceptual groups defined by CODE. It takes two perceptual objects and outputs a relation between them. It takes one perceptual object and a linguistic direction (e.g., *above*) as input and outputs a perceptual object that stands in that direction with respect to the first object.

Individually, the components of this architecture can account for a considerable range of attentional behavior. In combination, as specified by the architecture, they should be able to account for an even broader range of phenomena. It should be possible to make quantitative predictions in each case. The architecture affords a lot of flexibility, so it should be easy to provide quantitative accounts for different phenomena. The remainder of the paper applies CTVA to seven phenomena that have had an important impact on research and theory in visual spatial attention.

Applying the Theory to Data

The theory can be understood best by applying it to data. In the remainder of the article, CTVA is applied to reaction time and accuracy data from seven empirical situations in which grouping by proximity and distance between items have important effects. None of the current models of attention deal with these effects very adequately, including TVA, so they provide a good arena in which to investigate CTVA, to assess the benefits of marriage CODE to TVA. The fits to the data illustrate how

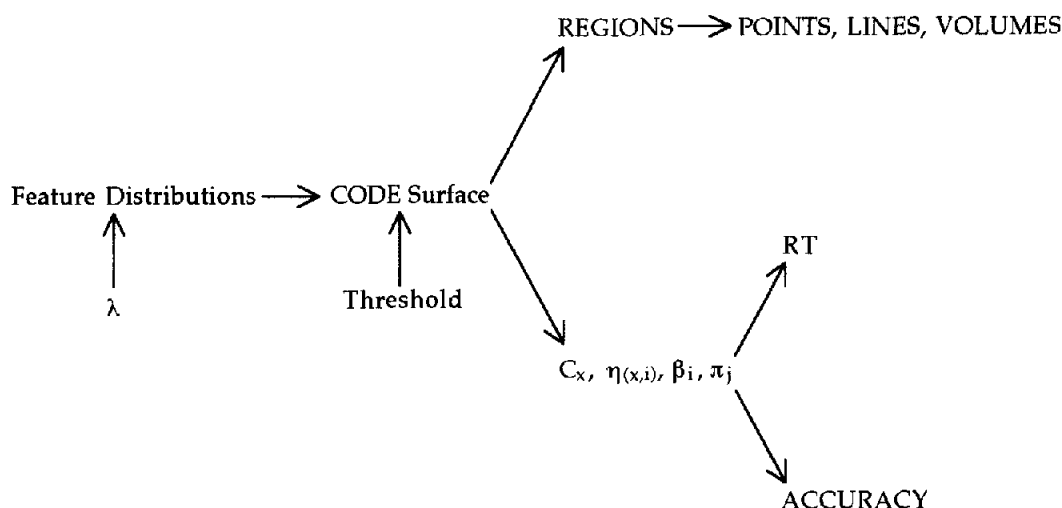


Figure 6. Architecture of the CODE theory of visual attention indicating the parameters and representations associated with the early identity and location system, the late identity system, and the late location system (cf. Figure 1). RT = reaction time.

different parts of the theory work and show which parameters are important.

The fits use the 1-D version of CODE because it is more tractable than the 2-D version. The feature catch for the 1-D case is defined by limits of integration that are single points on the one dimension. The points can be selected to group the display in various ways, as Figure 2 illustrates. The feature catch for the 2-D case is more complicated because the limits of integration extend irregularly in two dimensions, as illustrated by the contour lines defining the different threshold levels in Figure 3. Thus, the feature catch is much harder to compute in the 2-D case than in the 1-D case. Fortunately, the 1-D case provides a reasonable approximation to the situations I chose to fit.

The model involves a minimum of five parameters, and many of the phenomena to be modeled involve fewer than five conditions. For example, the flanker paradigm introduced by Eriksen and Eriksen (1974) involves three main conditions: compatible, incompatible, and neutral flankers. However, in many cases, several of the parameters can be held constant, so that (many) fewer than five actually predict performance. I tried to keep the spatial parameters close to the same values across the different paradigms. In most cases, I set the threshold equal to the local minimum between the target item and the nearest distractor. In most cases, the stimuli were roughly the same size— $\frac{3}{4}$ to 1° of visual angle—so I set 1° equal to 100 units of distance in the model and fixed the standard deviation of the feature distributions (i.e., $\sqrt{2}\lambda^{-1}$) at 50. Details of the fits to individual data sets can be found in Appendix C.

Prinzmetal (1981): Grouping Effects in Conjunction Search

For the last 15 years, much of the research on visual spatial attention has been driven by Treisman's *feature integration theory* (Treisman & Gelade, 1980; Treisman & Sato, 1990; Treisman & Schmidt, 1982). Feature integration theory argues

that attention is necessary to conjoin features that are processed separately. Attentional limitations on the conjunction process led to two predictions that were readily confirmed in the initial research and remain the focus of research today. First, visual search for targets that are conjunctions of separable features (such as a red *T* in a display of green *T*s and red *L*s) should be difficult, compared to search for the features themselves (such as a red *T* in a display of green *T*s; Treisman & Gelade, 1980). Second, when attention is stressed or overloaded, people should erroneously combine features from different objects. These errors, known as *illusory conjunctions*, appear as false alarms in search tasks or false reports of feature combinations in identification tasks (Treisman & Schmidt, 1982).

Prinzmetal (1981) demonstrated an important effect of grouping by proximity on illusory conjunctions. He found that illusory conjunctions were more likely if the features to be conjoined belonged to the same perceptual group than if they belonged to different groups. This result is well cited, and it is regarded as a strong piece of evidence for object-based attention (Kahneman & Treisman, 1984; Kahneman et al., 1992). It is an interesting test for CTVA because it can be accounted for entirely in terms of between-object selection, alternating between perceptual organizations produced at different threshold values. Within-object effects are not very important.

Method and results. Prinzmetal (1981) showed his subjects displays like those in Figure 7. Their task was to indicate whether or not the display contained a "plus": a vertical line superimposed on a horizontal line. The displays were preceded and followed by a noise mask and exposure duration was varied so that mean accuracy was approximately 85%. Exposure durations ranged from 30–150 ms between subjects. Mean exposure duration across Experiments 1–3 was 96.2 ms.

There were two important manipulations: The type of display (target, conjunction, or feature) and the way the features of target and nontarget items were distributed across the percep-

TARGET DISPLAYS

⊕○○○

○○○○

⊕○○○

⊕○○○

CONJUNCTION DISPLAYS

⊕○○○

○○○○

⊕○○○

⊕○○○

FEATURE DISPLAYS

⊕○○○

○○○○

⊕○○○

⊕○○○

Figure 7. Examples of displays from Prinzmetal's (1981) experiments. Top panels = target-present displays; middle panels = target-absent conjunction displays; bottom panels = target-absent feature displays; left panels = same-object displays; right panels = different-object displays.

tual groups formed by the circles. There were three types of display: *Target displays* contained the target "plus" and one other feature (either a horizontal or a vertical line); *nontarget feature displays* contained two examples of one of the features in separate locations but no plus (i.e., two horizontal lines or two vertical lines); *nontarget conjunction displays* contained one example of each feature (i.e., a horizontal line and a vertical line) but no plus. One third of the trials involved target displays, one third involved nontarget feature displays, and one third involved nontarget conjunction displays.

The contrast between conjunction and feature displays was critical: People should find it harder to say "no" to conjunction displays than to feature displays, because conjunction displays contain the two features that are conjoined in the target plus, whereas feature displays contain only one of the features, albeit repeated. Prinzmetal's (1981) subjects produced more than twice as many errors with conjunction displays than with feature displays. Averaging over Experiments 1–3, the probability of saying "yes" was .949 for target displays, .215 for conjunction displays, and .087 for feature displays.

More interesting, the difference between conjunction displays and feature displays was affected strongly by the distribution of features between groups. The features in each display were presented in the context of 8 (Experiments 1 and 2) or 10 (Experiment 3) circles organized by proximity into rows or columns. Figure 7 illustrates the displays from Experiment 1. The features of the conjunction display could either occur in the same group, as illustrated in the left panels of Figure 7, or in different groups, as illustrated in the right panels of Figure 7. Prinzmetal (1981) arranged the displays so that the Euclidean distance between the features was the same whether they appeared in the same or different groups, so that any difference in the probability of falsely conjoining the features of the display

would be due to perceptual organization rather than distance. If people processed all of the features of a perceptual group at once, as object-based theories assume (Kahneman & Henik, 1977, 1981; Kahneman & Treisman, 1984; Kahneman et al., 1992), they should say "yes" to conjunction displays, because one perceptual object—group—possesses both features of the target plus.

The data, averaged over Experiments 1–3 and presented in Table 1, showed a strong effect of perceptual organization. When the features were in the same group, the difference in false-alarm rates between conjunction displays and feature displays was .147; when the features were in different groups, the difference decreased to .110. Prinzmetal (1981) argued that this interaction could not be interpreted without assuming that subjects processed the display in two groups.

CODE. The first step in applying CTVA to the data is to analyse the feature catch provided by CODE. Figure 8A represents the CODE surface that would be produced by the stimuli Prinzmetal (1981) used in Experiment 1. Figures 8B–D illustrate three alternative feature catches available in the display that result from applying thresholds at three different levels. The highest threshold value cuts off the tips of each of the peaks, providing a feature catch that comes predominantly from the item on which the peak is centered.

The intermediate threshold value divides the display into two elongated objects, as Prinzmetal (1981) intended. The feature catch available at this threshold value lumps together all of the features in a perceptual group. Thus, within-group conjunction displays should be hard to distinguish from target displays, because their feature catches both contain the critical horizontal and vertical lines that make up the target plus. Between-group conjunction displays should be indistinguishable from feature displays because each group contains only one feature.

The lowest threshold value includes all of the items in the display in a single group. The feature catch for targets would include three features; the feature catch for nontargets would include two features. Conjunction displays and target displays would both contain the critical horizontal and vertical lines that form the target plus.

Prinzmetal's (1981) results require a mixture of threshold

Table 1
Observed and Predicted Probabilities of "Target-Present" Responses From Prinzmetal's (1981) Experiments on Grouping Effects in Illusory Conjunctions

Group	Target	Conjunction	Feature
Prinzmetal (1981) data			
Same	.956	.247	.100
Different	.942	.183	.073
Middle threshold (two groups)			
Same	.899	.810	.083
Different	.996	.178	.083
High threshold serial (2 items)			
Same	.996	.175	.083
Different	.996	.175	.083
High and middle threshold			
Same	.985	.247	.083
Different	.996	.175	.083

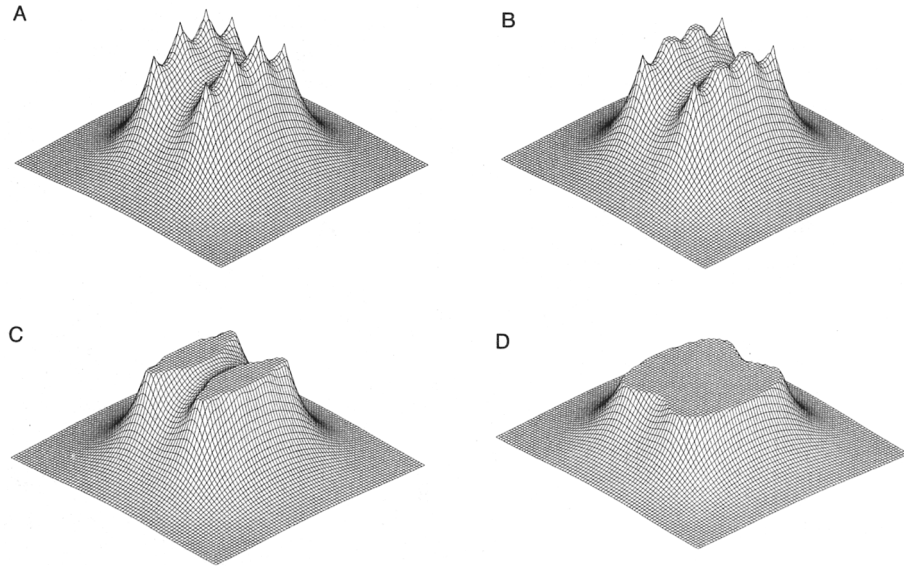


Figure 8. The CODE surface for Prinzmetal's (1981) displays (8A) with a high (8B), intermediate (8C), and low (8D) threshold applied to it (cf. Figure 7).

values. The highest and lowest thresholds predict no difference between same and different groups, and that difference was prominent in Prinzmetal's (1981) results. The intermediate threshold accounts for the difference but goes too far. It predicts a large difference between conjunction and feature displays in the same-group condition but no difference in the different-group condition. In Prinzmetal's (1981) data, the difference between conjunction and feature displays was almost as large in the different-group condition (.110) as in the same-group condition (.147).

The CODE analysis already constrains the interpretation of Prinzmetal's (1981) experiments. The data cannot be accounted for by a single threshold applied to the CODE surface. At least two different thresholds must alternate with one another. Alternation between the lowest and the highest cannot work because neither of them is sensitive to the distribution of features within and between groups. Alternation between the intermediate threshold and either the highest or the lowest may work, if performance with the high or low threshold (or both) is sensitive to the difference between conjunction and feature displays. The purpose of the TVA analysis is to see whether a two-threshold theory can account for the data.

CTVA. In order to fit TVA to Prinzmetal's (1981) data, I had to decide how to represent the features in the display and how to represent conjunctions. I accepted Prinzmetal's assumption that the features were horizontal and vertical lines and that the target cross was detected when the person perceived both a horizontal and a vertical line. To model the detection process, I let the presence of each feature race against the absence of that feature. Thus, horizontal raced with not-horizontal, and vertical raced with not-vertical. The η values for feature absence were 1 minus the η values for feature presence. The fits assumed $\eta = .99$ for feature presence and $\eta = .01$ for feature absence. Feature presence and absence had different β values

(.90 and .10, respectively, in the fits). The w_x values were set to 1.0 for both objects.

The CODE surface was built by placing the centers of the nearest items 125 units apart and setting the standard deviation of the feature distributions to 50. The feature-bearing items were 250 units apart in both the same- and different-group conditions. Two thresholds were applied to the CODE surface, one just above the local minimum between the nearest items and one just below it. The first (high) threshold organized the display into eight groups, as illustrated in Figure 8B, and the second (intermediate) threshold organized the display into two groups, as illustrated in Figure 8C. With the high threshold, if one feature-bearing object was selected, the feature catch for that object was .849 and the feature catch for the other feature-bearing object was .012 in both same-group and different-group conditions. With the intermediate threshold, the feature catch was .855 for both feature-bearing objects in the same-group condition and .855 for the selected object and .012 for the non-selected object in the different-group condition.

I tried two different fits. First, the intermediate threshold was fitted, which divided the display into two groups. Target-present displays contain the critical features necessary for correct detection no matter how the display was grouped. In conjunction displays the critical features necessary for an illusory conjunction were both present in one group in the same-group condition but distributed across groups in the different-group condition. In feature displays, neither grouping contained the critical features.

The predicted results, presented in Table 1, show a difference between conjunction and feature displays in both conditions and an interaction between grouping condition and display type like the one Prinzmetal (1981) observed. However, the false-alarm rate was much too high for the same-group conjunction displays; the difference between target-present displays and con-

junction displays was very small. Clearly, the intermediate threshold by itself cannot account for the data.

Next, the high threshold was fitted, which divides the display into eight objects. Only two of the objects contained features, and I assumed that attention was focused on one of them. The predicted results, presented in Table 1, captured the difference between conjunction and feature displays but not the interaction between display type and grouping. The difference between conjunction and feature displays was the same in the two grouping conditions. Apparently, the high threshold by itself cannot account for the data either.

No single threshold accounted for the data, so I tried combining the intermediate and high thresholds. From the subject's perspective, this amounts to changing between organizations of the display from trial to trial, sometimes seeing it as two rows or columns and sometimes seeing it as eight objects. I looked for a mixture that would give a false-alarm rate of .247 in the same-group conjunction condition and found that a mixture probability of .1135 was sufficient. In other words, subjects saw the display as two groups on 11% of the trials and as eight groups on 89% of the trials. The results, presented in Table 1, capture Prinzmetal's (1981) interaction.

Evaluation. CTVA did a reasonable job of accounting for Prinzmetal's (1981) data. The numbers from the combined-threshold fits in Table 1 are close to Prinzmetal's even though I did not try to optimize the fit formally. More important, the process of fitting was revealing. Prinzmetal (1981) wrote as if subjects always saw the display as two groups and didn't consider the possibility of alternative organizations. By contrast, CTVA could account for the interaction between display type and grouping only if subjects were allowed to group the display in different ways on different trials. In hindsight, subjects might have been expected to alternate between organizations. Prinzmetal drew the circles in blue and the lines and crosses in black. It is possible that on some trials—many trials, by the present analysis—subjects segregated the black objects from the blue ones and saw only the lines and crosses.⁶

Cohen and Ivry (1989): Distance Effects in Illusory Conjunctions

In their original investigation of illusory conjunctions, Treisman and Schmidt (1982) found no effect of distance on the probability of an illusory conjunction. However, since then, several researchers have found distance effects, such that the probability of an illusory conjunction decreases as the distance between the objects that contribute the miscombined features increases (Chastain, 1982; Cohen & Ivry, 1989; Ivry & Prinzmetal, 1991; Lasaga & Hecht, 1991; Prinzmetal & Keysar, 1989; Prinzmetal & Mills-Wright, 1984; Prinzmetal, Treiman, & Rho, 1986; Wolford & Shum, 1980). CTVA provides a straightforward account of this distance effect. The analysis focuses on Cohen and Ivry's (1989) experiments because they were concerned primarily with distance effects.

Method and results: Experiments 1 and 2. Cohen and Ivry (1989) reported four experiments on distance effects in illusory conjunctions. Their experiments were organized in sets of two. The procedures of the first two experiments were straightforward: Subjects were presented with a central digit (Experiment

Table 2

Observed and Predicted Response Probabilities From Cohen and Ivry's (1989) Experiments 1 and 2 on Distance Effects in Illusory Conjunctions

Response	Experiment 1		Experiment 2		Predictions	
	Near	Far	Near	Far	Near	Far
Correct	.535	.679	.608	.759	.558	.673
Color feature	.171	.160	.101	.101	.149	.150
Color conjunction	.135	.061	.130	.041	.189	.076
Letter feature	.063	.050	.074	.055	.065	.075
Letter and color feature	.052	.033	.047	.029	.017	.017
Letter and color conjunction	.044	.017	.039	.016	.022	.009

Note. Correct = probability of reporting letter and color of target object correctly; color feature = probability of reporting letter correctly and color incorrectly; color conjunction = probability of reporting letter correctly and reporting color of nontarget object; letter feature = probability of reporting letter incorrectly and color correctly; letter and color feature = probability of reporting letter and color incorrectly; letter and color conjunction = probability of reporting letter incorrectly and reporting color of nontarget object.

1) or pair of digits (Experiment 2) and a pair of peripheral letters on an imaginary circle about 2.5° from fixation. The displays were exposed briefly and masked. The task was to first name the digit (Experiment 1) or the smaller or larger of the two digits (Experiment 2) and then name the color and identity of one of the letters. One letter was always an *O*. The other was either an *F* or an *X*. The colors were pink, yellow, green, and blue. The letter *O* was a distractor; the task was to name the color and the identity of the letter that was not *O*. The main manipulation was the distance between the letters, which was either .88° (near) or 2.86° (far), center to center.

The main data were the probabilities of reporting combinations of letter identities and colors, which are presented in Table 2. These probabilities came from trials in which the digit was

⁶ Accuracy is better overall in the high-threshold condition than in the middle-threshold condition, which raises the question why subjects would ever adopt the middle threshold instead of relying exclusively on the high threshold. The answer must be that grouping by proximity is compelling: subjects cannot avoid perceiving the display as two groups entirely. The reason for this can be seen in the CODE surfaces depicted in Figure 8: Prinzmetal's (1981) displays are organized in two groups over much of the possible range of threshold variation. Very high thresholds are required to separate the items within groups, and higher thresholds might exclude all the items. The range of threshold variation that parses the display into eight objects is relatively narrow. Note as well that the accuracy for the high-threshold condition is as good as the accuracy for the middle-threshold, different-group condition. This was a consequence of my decision to consider only the two feature-bearing items in the fits. If all eight items were included in the high-threshold fits, accuracy would be lower because of noise from the six featureless items (see Equations 16 and 17). Nevertheless, it would still be higher than accuracy in the middle-threshold, same-group condition, in which the two target features are included in the same group. In that condition, the false-alarm rate for conjunction displays will always be close to the hit rate for target displays because both target features are present in the feature catch.

reported correctly. The requirement to report the digit was intended to focus subjects' attention on the digit, away from the peripheral letters. Treisman and Schmidt (1982) argued that illusory conjunctions occurred primarily when attention was stressed or distracted, and this manipulation was intended to have that effect. The requirement to report the larger or smaller of the two digits in the second experiment was intended to focus attention more stringently than in the first experiment. Reporting a conjunction (of size and identity) should require a sharper focus of attention than reporting a single feature (identity; Treisman & Gelade, 1980).

The results of the two experiments were essentially the same. Subjects made illusory conjunctions in the near-spacing condition but not in the far-spacing condition. Evidence of illusory conjunctions was obtained by comparing the probability of a color-feature error (given that letter identity was reported correctly) with the probability of a color-conjunction error (given that letter identity was reported correctly). There were four colors, one correct and three incorrect. A color-feature error occurred if the reported color was not present in the other item in the display. There were two possible color-feature errors. A color-conjunction error occurred if the reported color was the color of the other item in the display. If color report was at chance, then there should be half as many color-conjunction errors as color-feature errors because there was one nontarget color presented in the display and two not presented (i.e., a ratio of 1:2). The number of color conjunction errors was greater than this chance expectation in the near spacing condition, showing that illusory conjunctions were prevalent when the contributing items were close. The number of color conjunction errors was less than the chance expectation (slightly but significantly) in the far spacing condition, showing that illusory conjunctions were unlikely to occur when the items were far apart. Thus, the probability of illusory conjunction decreases with distance. Cohen and Ivry (1989) interpreted the less-than-chance frequency of illusory conjunctions in the far condition as evidence of an "exclusionary guessing strategy," whereby subjects would detect the color of the far item correctly and exclude it from their guesses.

CODE. The CODE analysis is straightforward. The displays would contain three (Experiment 1) or four distributions (Experiment 2), two of which correspond to the critical colored letters. The analysis focused on the two distributions for the colored letters. I ignored the distributions for the digits because they are far from the colored letters and differ from them categorically. Thus, TVA would set the β and π values for digits close to zero when selecting colored letters, so the digits would have virtually no impact on the race even if they were present in the feature catch.

The feature distributions for the colored letters were set 50 units apart in the near condition and 250 units apart in the far condition. I set the standard deviation of the feature distributions at 50. The threshold was set just above the local minimum between the distributions in order to maximize the feature catch. According to CTVA, illusory conjunctions occur when the feature catch from a selected above-threshold region contains features from different items and the first relevant features to finish come from different items. For example, if a pink X and a green O are both sampled in the feature catch and " X "

and " $green$ " are the first relevant categorizations to finish, the person will report an illusory conjunction. The probability that illusory conjunctions will occur depends on the overlap of the feature distributions from the different items in the feature catch. The further apart the items, the smaller the overlap, and the less likely the illusory conjunctions. With these parameters, the feature catch for the target item and its neighbor were .394 and .192, respectively, in the near condition, and .918 and .041 in the far condition.

CTVA. The TVA analysis involved deciding whether the target item was one of two letters (X or F) and one of four colors (pink, green, yellow, or blue). The β and w_x values for these categorizations were set to 1.0, and the $\eta(x, i)$ values ranged between 0 and 1. The η values for target letters and colors were set to 0.9, and the η values for the nontarget categorizations, given that a target was present, were set to 0.1. Thus, if the target was F , $\eta(x, F)$ was set to 0.9 and $\eta(x, X)$ value for X was set to 0.1. If the color was pink, $\eta(x, pink)$ was set to 0.9 and the $\eta(x, i)$ s for the other colors were set to 0.1. The predicted results appear in Table 2 along with the observed data.

The predicted results capture the main effect observed by Cohen and Ivry (1989): Illusory conjunctions were more prevalent in the near condition than in the far condition. In the near condition, the ratio of color conjunction errors to color feature errors was 1.267 in the simulated data, compared to .789 and 1.287 in Cohen and Ivry's (1989) Experiments 1 and 2, respectively. By contrast, in the far condition, the ratio of color-conjunction errors to color-feature errors was .508 in the simulated data, which is close to chance expectation. Cohen and Ivry (1989) found ratios of .381 and .405 in their far condition, presumably because their subjects used an exclusionary guessing strategy that I did not attempt to model. Nevertheless, the CTVA predictions are reasonably close to their data even though there was no formal attempt to optimize the fit.

Method and results: Experiments 3 and 4. Cohen and Ivry's (1989) third and fourth experiments attempted to test the hypothesis that illusory conjunctions occurred between items that fell within the spotlight beam but not between items that fell outside it. To this end, they had subjects report two digits 3.3° or 6.6° apart (Experiment 3) or 3° or 6° apart (Experiment 4) and then report the color and identity of an X or F that appeared with a colored, nontarget O , as in Experiments 1 and 2. The two colored letters appeared between the digits, one outside and one between the digits, or both outside the digits. According to Cohen and Ivry's hypothesis, subjects would expand the spotlight to encompass both of the digits, so the spotlight would be larger when the digits were more widely spaced. The letters to be reported were presented in two out of six equally spaced locations, which Cohen and Ivry (1989) labeled from left to right as $A-F$. The digits appeared between positions $A-B$ and $E-F$ in the large spotlight condition and between positions $B-C$ and $D-E$ in the small spotlight condition. The two experiments were close replications; the only difference was that visual angle was smaller by 10% in Experiment 4. Examples of their displays are presented in Table 3.

The hypothesis that illusory conjunctions would occur within the spotlight beam but not outside it led to several predictions: First, it predicted illusory conjunctions only in condition CD in the small spotlight condition (narrowly spaced digits), because

Table 3
Examples of Displays and Observed and Predicted Rates
of Illusory Conjunctions in Cohen and Ivry's
(1989) Experiments 3 and 4

Position	Example displays				Exp. 3	Exp. 4	CTVA
Small							
CD	wX Yz		wY Xz		.111	.070	.136
BD-CE	Xw	Yz	Yw	Xz	.008	.022	.071
BE	Xw	zY	Yw	zX	-.013	.032	.019
AD-CF	X w	Yz	Y w	Xz	.016	.017	.032
AE-BF	X w	zY	Y w	zX	-.023	-.009	.009
AF	X w	z Y	Y w	z X	.004	.010	.004
Large							
CD	w	X Y z	w	Y X z	.119	.149	.138
BD-CE	wX	Y z	wY	X z	.103	.081	.084
BE	wX	Yz	wY	Xz	.046	.109	.047
AD-CF	Xw	Y z	Yw	X z	.039	.033	.035
AE-BF	Xw	Yz	Yw	Xz	.021	.065	.022
AF	Xw	zY	Yw	zX	.015	.010	.003

Note. In the example displays, *X* and *Y* represent the locations of the colored letters to be reported in the conjunction task, and *w* and *z* represent the locations of the digits to be reported in the primary task. *Small* and *Large* refer to the distance between the digits. The letters *A-F* represent Cohen and Ivry's (1989) notation for the position of the colored letters, where *A* is the leftmost position, *B* is second from the left, and so on. The rate of illusory conjunction is the probability of a color conjunction error minus half of the probability of a color feature error.

that was the only condition in which both letters fell between the digits and therefore within the beam. Second, it predicted illusory conjunctions in conditions *CD*, *BD*, *CE*, and *BE* in the large spotlight condition (widely spaced digits) because both letters fell between the digits in each of those conditions. Third, it predicted no illusory conjunctions outside the spotlight in any condition (i.e., in conditions *BD*, *CE*, *BE*, *AD*, *CF*, *AE*, *BF*, or *AF* in the small spotlight condition or in conditions *AD*, *CF*, *AE*, *BF*, or *AF* in the large spotlight condition). And fourth, it predicted no effect of distance on the rate of illusory conjunctions when the letters fell within the spotlight.

The illusory conjunction rates, presented as a function of condition in Table 3 and plotted as a function of distance between the letters in Figure 9, provided partial support for their hypothesis. Illusory conjunctions tended to occur within the spotlight but not outside it in the small spotlight condition. However, in the large spotlight condition, illusory conjunctions occurred between letters inside and outside the spotlight (conditions *AD*, *CF*, *AE*, and *BF*) and there were strong distance effects within the spotlight (see Figure 9). Indeed, inspection of Figure 9 reveals no sharp discontinuity in the rate of illusory conjunctions at the boundary of the spotlight.

CODE. The CODE analysis begins by rejecting the idea that the spotlight can be stretched arbitrarily to include the two digits. If the spotlight included the two digits and the space between them, then there should be no basis for conjoining colors and letter identities correctly, so illusory conjunctions should occur as often as correct conjunctions. We saw this effect in the CTVA analysis of the middle-threshold, same-group condition

of Prinzmetal (1981). Furthermore, in order to include both digits in a single spotlight beam, the threshold would have to be set so low that it would include the two colored letters wherever they appeared in the display. Letters outside the digits would be

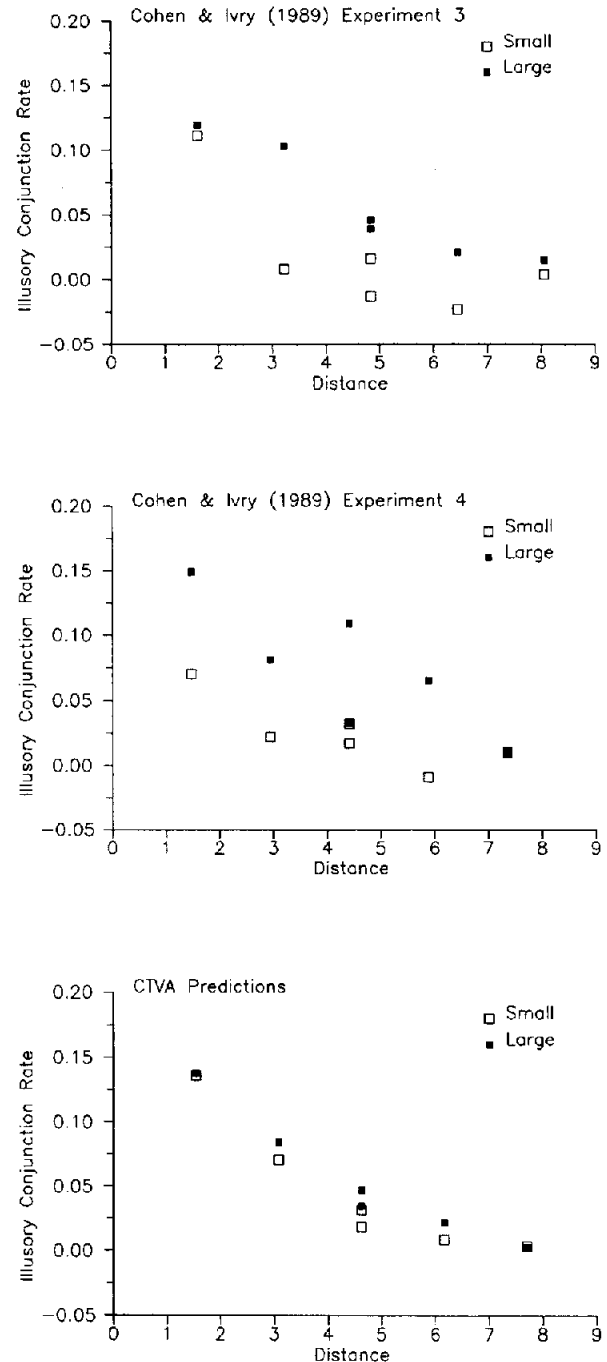


Figure 9. Rates of illusory conjunctions (the probability of a color conjunction error minus half the probability of a color feature error) as a function of distance between the colored letters observed in Cohen and Ivry's (1989) Experiment 3 (top panel) and Experiment 4 (middle panel), and predicted by CTVA (bottom panel). CTVA = CODE theory of visual attention.

Table 4
Feature Catches for Each Condition of Cohen and Ivry's
(1989) Experiments 3 and 4

Position	Small spotlight		Large spotlight	
	Target	Distractor	Target	Distractor
CD	.1490	.1095	.1397	.1038
BD-CE	.1329	.0623	.3119	.1634
BE	.0050	.0011	.4513	.1665
AD-CF	.1779	.0525	.1809	.0564
AE-BF	.0332	.0050	.2549	.0618
AF	.0838	.0079	.0149	.0012

Note. The letters A-F represent Cohen and Ivry's (1989) notation for the position of the colored letters, where A is the leftmost position, B is second from the left, and so on.

just as likely to be included in the above-threshold region as letters between the digits. Because of these problems, the CODE analysis took a different tack.

The CODE analysis assumes that the regions of the display from which features are sampled depend on the shape of the CODE surface and the threshold setting. Cohen and Ivry's (1989) displays were represented as CODE surfaces generated from four feature distributions, with thresholds set at the local minima between the distributions, slicing off four separate regions from which features can be sampled. CODE accounts for the putative effects of the spotlight in terms of the influence of the feature distributions of the digits on the CODE surface.

The smallest distance in the set of displays was set to 25 units (i.e., the distance between *X* and *w* in conditions CD, BD, CE, and BE in the small spotlight condition) and the distance between alternative letter positions was set to 50 units (i.e., the distance between *X* and *Y* in condition CD in both the small and large spotlight conditions). All other distances were multiples of these distances. I set the standard deviation of the feature distribution at 100. The threshold was set differently in each condition at the local minimum between the letters and their nearest neighbors. The feature catches for targets and distractors computed from these parameters are presented in Table 4, averaged over the two positions that targets could have occupied in each display (i.e., the positions corresponding to *X* and *Y* in each row of Table 3). Cohen and Ivry (1989) did not report data separately for the two positions.

CTVA. As in the analysis of Experiments 1 and 2, β and w_x were set to 1.0 for color and letter categorizations. The $\eta(x, i)$ values for target letters and colors were set to 0.825, and the $\eta(x, i)$ values for the nontarget categorizations, given that a target was present, were set to 0.175. The predicted illusory conjunction rates (the probability of a color-conjunction error minus half the probability of a color-feature error) are presented in Table 3 as a function of condition and plotted in Figure 9 as a function of distance between the colored letters. The correlation between observed and predicted values was .888 for Experiment 3 and .729 for Experiment 4. These correlations are high considering that the observed illusory conjunction rates from Experiments 3 and 4 correlated only .805 with each other (i.e., the data were somewhat unreliable).

As in Experiments 1 and 2, CTVA did a good job of capturing the reduction in illusory conjunction rate as the distance between the letters increased. The predicted data in Figure 9 decrease with distance at about the same rate as the observed data in both experiments. Moreover, the predicted data showed illusory conjunctions between letters inside and outside the spotlight and distance effects within the beam in the large spotlight condition, just like the observed data.

CTVA predicted a difference in the right direction between the illusory conjunction rates in the small and large spotlight conditions. Even though the distance between the letters was the same in the two conditions, the model's performance was influenced in the same manner as human subjects' by adding the digits to the display. However, the CTVA fit was not perfect. It tended to overpredict the data in the small spotlight condition and underpredict them in the large spotlight condition; the observed difference was larger than the predicted one. Thus, there may be more going on in these experiments than CTVA can account for.⁷

Evaluation. The CTVA model captured the essential feature of Cohen and Ivry's (1989) experiments, which is a reduction in illusory conjunction errors as the distance between items increased. The model provided a better account of the simple distance effects in Experiments 1 and 2 than the modulation of distance effects by the spacing of the digits in Experiments 3 and 4. Thus, there is room for improvement. Note, however, that the spotlight model proposed by Cohen and Ivry (1989) did not fare very well either, even though it made only qualitative predictions (also see Ashby et al., 1996). Nevertheless, the CTVA fits are encouraging. They suggest that the model could be extended to deal with the other cases in the literature (e.g., Chastain, 1982; Ivry & Prinzmetal, 1991; Lasaga & Hecht, 1991; Prinzmetal & Keysar, 1989; Prinzmetal & Mills-Wright, 1984; Prinzmetal, Treiman, & Rho, 1986; Wolford & Shum, 1980). The model did not deal with the initial digit-report task or the shift of attention from the digits to the target in any of the ex-

⁷ I was able to improve the fit and capture the quantitative difference between the large and small spotlight conditions by allowing the standard deviation of the feature distributions to vary between spotlight conditions, following Ashby et al. (1996) who fitted the same data by allowing larger variance in the large spotlight condition. I set the standard deviation of the feature distributions to 60 units for the small spotlight condition and kept the other parameters the same (i.e., β and $w_x = 1$; $\eta = 0.825$ for color and letter presence; $\eta = 0.175$ for color and letter absence). The large spotlight condition was fitted with the same parameters used for the fits to both conditions in Table 3 and Figure 9 (i.e., the predicted data are the same as those for the large spotlight condition in Table 3). The predicted illusory conjunction rates for the small spotlight condition were much closer to the observed values: .096, .015, .003, .004, .001, and .0002 for Conditions CD, BD-CE, BE, AD-CF, AE-BF, and AF, respectively. The correlations between the predicted and observed illusory conjunction rates (including the new predictions for the small spotlight condition and the old predictions for the large spotlight condition) were much higher than with the previous fits: .986 for Experiment 3 and .869 for Experiment 4. However, these improved fits required me to violate the CTVA assumption that the feature distributions are built by bottom-up processes, which implies that their standard deviations should be independent of how top-down attention is deployed to the displays.

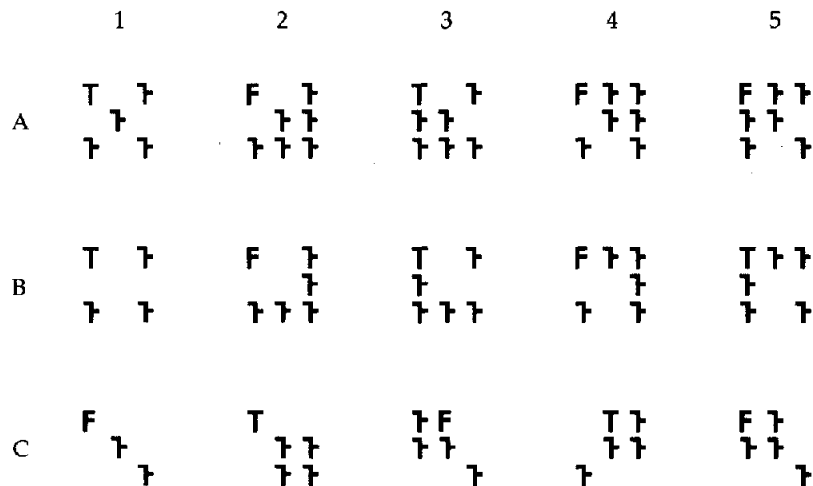


Figure 10. Examples of displays from Banks and Prinzmetal's (1976) experiments. Column 1 = good figure condition; Column 2 = isolated target condition; Columns 3-5 = camouflaged target conditions.

periments. The digit-report task would be easy to model but the shift in attention would require further specification of between-object processing, which is beyond the scope of this article (but see Logan, 1995; Logan & Sadler, 1996).

Banks and Prinzmetal (1976): Grouping Effects in Visual Search

Banks and Prinzmetal (1976) published an important series of experiments that pitted grouping by proximity against the number of items in the display. The results were counterintuitive: Adding items to a display usually impairs performance (e.g., Duncan & Humphreys, 1989; Treisman & Gormican, 1988; Wolfe, 1994), but Banks and Prinzmetal (1976) found that adding items improved performance when those items clustered together with other distractors to isolate the target. These results are well cited and viewed as strong evidence for object-based attention (e.g., Kahneman & Treisman, 1984; Kahneman et al., 1992).

The grouping principle in the Banks and Prinzmetal (1976) experiment is proximity, so CTVA is clearly relevant. In the CTVA analysis, adding items has two effects, one on between-group selection and one on within-group selection. The effect on between-group selection is that adding items makes the isolated target more likely to be selected than in the original displays, and it makes targets that were not isolated but camouflaged by the added items, less likely to be selected than targets in the original displays. The effect on within-group selection is primarily on processing the camouflaged targets: Adding items to the display places one or two more distractors close to the target, close enough to affect a feature catch centered on the target.

Method and results. Banks and Prinzmetal (1976) showed their subjects displays like those in Figure 10. Each display contained a *T* or an *F* and two to six *T-F* hybrids (formed, roughly, by moving the right half of the top bar of the *T* down the stem to the position of lower bar of the *F*). *T*s and *F*s occurred only

in the four corners of the display in Conditions A and B and close to the corners in Condition C. The task was to say whether each display contained a *T* or an *F*.

There were three experiments. Experiment 1 used the full set of displays in Figure 10, exposing them until the subject responded, so reaction time was the main dependent variable. Experiment 2 used a subset of the displays (those in Figure 10A). The displays were exposed briefly (50 ms in the first session, 40 ms in the second) and followed by a blank field with twice the luminance. Accuracy became an important dependent variable as well as reaction time. Experiment 3 used the full set of displays to gather measures of grouping. Subjects looked at pictures of the displays in Figure 10 and drew lines around the groups they saw. The grouping measure, reported in Banks and Prinzmetal's (1976) Table 1, reflected the mean number of group boundaries between the target and the average distractor.

The design of Experiments 1 and 2 compared the five conditions represented in the columns of Figure 10. Condition 1 was the *good figure* condition, in which the target and distractors together formed a simple figure: a diagonal line, a square, or a square with a dot in the middle. Banks and Prinzmetal (1976) expected the target to be embedded in this simple structure.

Condition 2 was the main focus of their research. It was the *isolated target* condition, in which the added distractor grouped together with the other ones to form a cluster separate from the target. Banks and Prinzmetal (1976) expected better performance in the isolated target condition than in the good figure condition, because the target would be easier to extract from the distractors.

Conditions 3-5 were the *camouflaged target* conditions, in which the target appeared in the cluster of items formed by adding two new distractors. Banks and Prinzmetal (1976) expected worse performance in the camouflaged target conditions than in the good figure condition because of a display size effect: There were two more distractors in the displays.

Averaged over the three stimulus sets in Experiment 1, mean reaction times were 576, 553, 632, 670, and 726 ms for Condi-

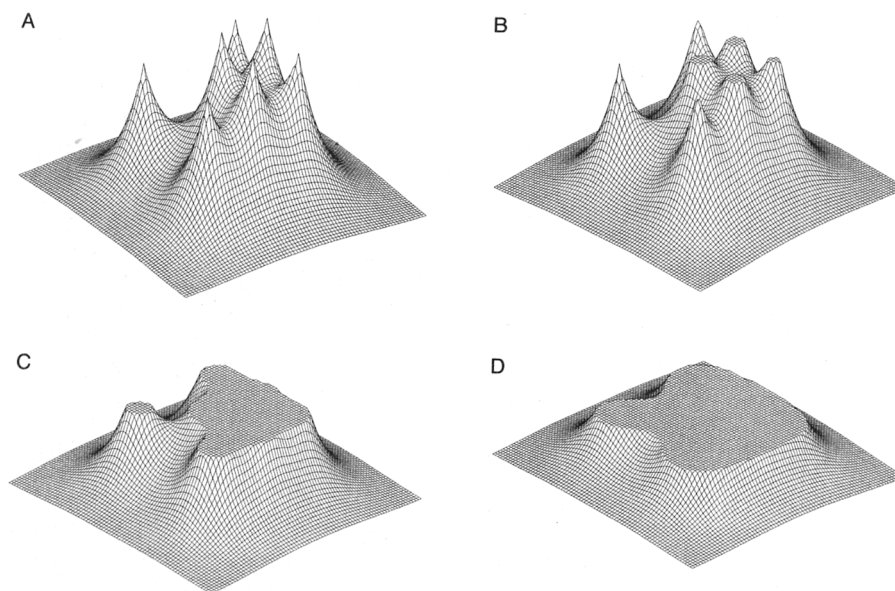


Figure 11. The CODE surface for Banks and Prinzmetal's (1976) displays (11A) with a high (11B), intermediate (11C), and low (11D) threshold applied to it (cf. Figure 9).

tions 1–5, respectively. Experiment 2 confirmed these results, producing mean reaction times of 583, 559, 596, 613, and 651 ms and mean probabilities of a correct response of .973, .987, .970, .942, and .915 for Conditions 1–5, respectively. The data confirmed Banks and Prinzmetal's (1976) expectations: Reaction time was faster for isolated targets than for good-figure targets, and reaction time was slower for camouflaged targets than for good-figure targets.

CODE. The CODE surface for the isolated- and camouflaged-target displays are illustrated in Figure 11A. There is no difference between targets and distractors in Figure 11, so the isolated target display has the same CODE surface as the three camouflaged target conditions. Figure 11B illustrates the application of a high threshold to the CODE surface, one that separates each of the items from the others in the display. Figure 11C illustrates an intermediate threshold that clumps the cluster of items into one group and isolates the singleton. Figure 11D illustrates a low threshold that groups all of the items together.

The CODE analysis provides some insight into the configuration of CODE and TVA that is required to fit the data. The low-threshold setting is an unlikely candidate for the CODE contribution because it ignores the spatial arrangement of the items. Banks and Prinzmetal (1976) found strong effects of spatial arrangement. The low-threshold setting could predict the display-size effect that Banks and Prinzmetal (1976) observed, but it could not predict the crucial difference between good figure and isolated target displays, in which the effects of grouping were pitted against display size and grouping won.

The intermediate- and high-threshold settings are reasonable candidates whose viability rests on the TVA analysis. The intermediate-threshold setting requires processing the clustered items in the camouflaged target conditions in parallel, and that may or may not be feasible depending on the signal to noise ratio in TVA.

Between-object effects. The high-threshold setting requires serial processing, and that requires a theory of between-object selection, which is outside the scope of CTVA. Nevertheless, some simple assumptions can be made that lead to testable predictions. I assumed that items are processed in an order that corresponds to their degree of isolation. Thus, the isolated item is processed first, then the two items with two neighbors, and finally, the item with three neighbors. Banks and Prinzmetal (1976) dismiss strategies like this as "far from optimal" (p. 362) because the target is more likely to occur in a nonisolated position in Conditions 1–4. However, the target is no more likely to occur in any other single position than in the isolated position, so there is no reason to prefer any other position to the isolated one. The strategies are no worse than random choice. Moreover, the strategies may be reasonable if they are consistent with habits or "natural tendencies" or if they interact with the recognition system in a way that benefits performance (see Logan, 1994).

This search strategy allows an estimate of the mean number of comparisons required to find the target (*search depth*) for the displays in Figure 10. If subjects examine only the four positions that targets can occur in, search depth will range from 1 to 4. The search strategy predicts a search depth of 1 for the isolated target condition and 2.5 for the good figure condition. In the camouflaged target condition, search depth should be 2.5 for the 2 two-neighbor positions and 4 for the three-neighbor position, averaging 3.

Within-object effects. CODE by itself also predicts some within-object effects in Banks and Prinzmetal's (1976) experiments. Figure 12 illustrates a 1-D CODE surface applied to a slice of Banks and Prinzmetal's (1976) displays. The top panel of Figure 12 shows three equally spaced items that correspond to the good figure condition. The middle and bottom panels of Figure 12 add a fourth item, placing it in between two of the

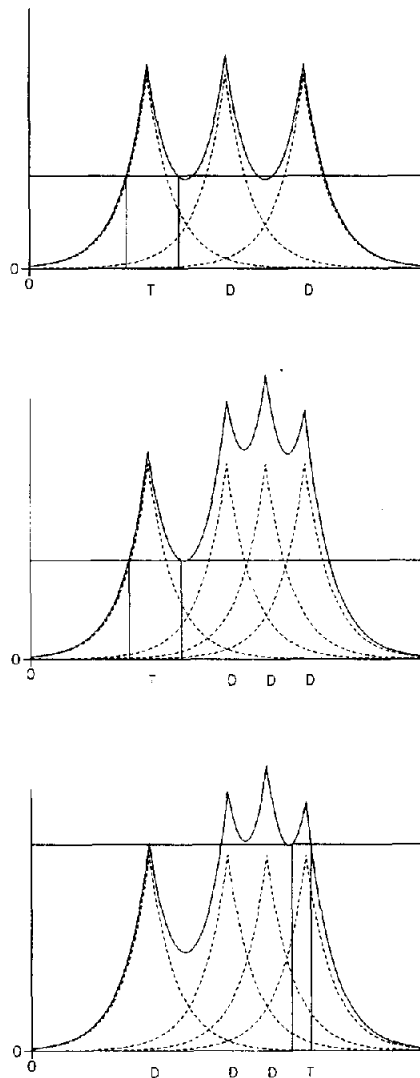


Figure 12. The CODE surface with a threshold applied to it for Banks and Prinzmetal's (1976) good figure condition (top panel), isolated target condition (middle panel), and camouflaged target condition (bottom panel) arrayed in one dimension. T = target; D = distractor.

three items from the top panel of Figure 12. The middle panel of Figure 12 represents the isolated target condition, and the bottom panel of Figure 12 represents the camouflaged target condition.

Thresholds have been applied to the different conditions and lines have been drawn to delimit the feature catch. The thresholds represent the high-threshold condition, in which CODE parses the display into objects that correspond to individual items. The thresholds were set at the lowest level that would allow the target item to be separated, which is a local minimum between items on the CODE surface. Thresholds higher than the local minimum will pick off individual items, but thresholds lower than the local minimum will group the target item with its neighbors (i.e., the local minimum represents the boundary between the high- and intermediate-threshold conditions).

The important points to be taken from Figure 12 concern the feature catches in the different conditions. The feature catches for the good figure condition and the isolated target condition are large and not contaminated much by their neighbors. The nearest neighbor is relatively far away, and the local minimum on the CODE surface between the target and the neighbor is relatively low. By contrast, the feature catch for the camouflaged-target condition is smaller and much more contaminated by its neighbor. The near neighbor raises the CODE surface in the region of the target and, consequently, raises the local minimum between itself and the target. This reduces the weight, c_x , on the target, relative to the good figure and isolated target conditions (by reducing the range of the limits of integration, excluding more of the tails of the target's feature distribution), and the reduction in the weight necessarily slows reaction time and decreases accuracy (see Equations 11–13). Adding insult to injury, the near neighbor intrudes more into the above-threshold region, giving it substantial weight in the feature catch. The extra item in the feature catch should slow reaction time and decrease accuracy further.

This analysis suggests that the number of near neighbors might be an important predictor of reaction times, because the threshold adjustment and noise effects are exacerbated by near neighbors. I counted the number of near neighbors for Banks and Prinzmetal's (1976) displays, portrayed in Figure 10, counting a horizontally, vertically, or diagonally adjacent item a near neighbor and not counting anything else. So, for example, the number of near neighbors in displays in the top row of Figure 10 is 1, 1, 2, 2, and 3 for Conditions 1–5, respectively. Averaged over all the displays, the number of near neighbors was .67, .67, 2.0, 2.0, and 2.67 for Conditions 1–5 respectively.

Regression analyses. I performed some regression analyses, predicting the 15 reaction times in Banks and Prinzmetal's (1976) Experiment 1 (see their Table 1) from the CODE-based measures of search depth and number of near neighbors and comparing the CODE predictions with those from Banks and Prinzmetal's (1976) grouping measure and display size. The correlations between the measures appear in Table 5. The simple and multiple regression equations appear in Table 6.

Individually, the CODE measures outperformed Banks and Prinzmetal's. The measure of search depth and the number of nearest neighbors were each more highly correlated with reaction time than Banks and Prinzmetal's grouping measure and display size. The CODE measures outperformed Banks and Prinzmetal's in multiple regression as well. Combining number of nearest neighbors with the measure of search depth resulted

Table 5
Correlations Between Reaction Times (RTs) in Banks and Prinzmetal's (1976) Experiment 1 and Measures of Grouping (–G), Display Size (D), Number of Near Neighbors (N), and Search Depth (S)

Predictor	RT	–G	D	N
–G	.567	—		
D	.311	.269	—	
N	.758	.523	.206	—
S	.826	.732	.000	.620

Table 6
Simple and Multiple Regression Equations Predicting Reaction Times (RTs) From Banks and Prinzmetal's (1976) Experiment 1 From Measures of Grouping ($-G$), Display Size (D), Number of Near Neighbors (N), and Search Depth (S)

R or r	Predictor	Equation
.5667	$-G$	$RT = 727 - 84.6G$
.3105	D	$RT = 531 + 17.9D$
.5900	$-G + D$	$RT = 664 - 79.1G + 9.9D$
.7580	N	$RT = 553 + 49.1N$
.8255	S	$RT = 488 + 57.4S$
.8831	$N + S$	$RT = 489 + 25.7N + 40.3S$

Note. R = multiple correlation from multiple regression; r = simple correlation from simple regression.

in a multiple correlation that was considerably higher than the multiple correlation from the grouping measure and display size. The multiple correlation including display size was only slightly higher than the simple correlation between the grouping measure and reaction time. Even without TVA, CODE provides a better account of Banks and Prinzmetal's (1976) data than their own analyses.

CTVA. The first step in applying TVA is setting the η values. In the Banks and Prinzmetal (1976) experiment, the distractors (T - F hybrids) are similar to the alternative targets (T and F). I used three levels of η : a high one for targets and distractors resembling themselves, an intermediate one for the mutual resemblance between targets and distractors, and a low one for the mutual resemblance between alternative targets. Thus, $\eta(x, T|x = T) = \eta(x, F|x = F) = \eta(x, D|x = D) = 1.0 > \eta(x, T|x = D) = \eta(x, D|x = T) = \eta(x, F|x = D) = \eta(x, D|x = F) = .02 > \eta(x, T|x = F) = \eta(x, F|x = T) = .01$. Bias (β) and attention weights (w_x) were set to 1.0 for T s, F s, and distractors.

Three between-item distances were included in the calculations: nearest neighbors, which were 1° of visual angle away (100 units); middle neighbors, which were 1.41° away (diagonally; 141 units); and far neighbors, which were 2° away (200 units). The standard deviation of the feature distributions was set to 50. In order to apply Equations 11 and 12 to the data, I treated the displays as if they were 1-D. I generated a CODE surface by adding together the feature distribution for the target and its nearest neighbor. This allowed me to define the local minimum surrounding the target and therefore set the threshold. I used the threshold set at the local minimum to compute the feature catch from each item in the display.

I fit parallel and serial models to the data. There were two versions of each type, one with the same threshold for each of the four critical display positions (set to the local minimum between the target and the nearest-possible neighbor, which is the added item in Figure 10, Conditions 3–5), and one with a different threshold for each critical position (set to the local minimum between the target and its nearest neighbor).

The parallel models focused on the four critical display positions (the four corner positions in each display). Each position contributed two "runners" to the race, one for each possible target (i.e., T vs. F), and the four positions raced against each

other. Reaction time and accuracy predictions were generated from Equation 12.

The serial models used the search strategy described above in the regression fits, focusing on the most isolated item first and proceeding through the display according to the degree of isolation. The race was run separately for each of the four positions in the display. There were three runners in the race at each position— T , F , and distractor. The distractor ran in the race because three out of four positions in each display contained distractors, and the appropriate action, if a distractor was present, was to proceed to the next display position. Reaction time and accuracy predictions were generated by first applying Equation 12 to generate processing times and accuracies for each display position and then integrating them with the serial search strategy. Reaction times for successive display positions were added, and the reaction time for each display was set to the mean of the different trajectories through the display. Accuracies for successive display positions were multiplied together, and the accuracy for each display was set to the mean of the different trajectories.

The models were fitted to the 15 reaction times in Banks and Prinzmetal's (1976) Experiment 1 (see their Table 1). The same parameters predicted accuracy, although Banks and Prinzmetal did not report it. They did report accuracy for their Experiment 2, which was a partial replication of their Experiment 1 with brief exposures, so we tried to fit those accuracy data. The serial model fits were better than the parallel model fits and the different-threshold fits were marginally better than the same-threshold fits ($r = .913$ for different- and .903 for same-threshold serial fits; $r = .808$ for different- and .807 for same-threshold parallel fits). The results of the different-threshold fits and the results of Banks and Prinzmetal's (1976) Experiment 1 are presented in Table 7.

The parallel models missed two essential features of Banks and Prinzmetal's (1976) results. First, they missed the display size effect, predicting longer reaction times for Pattern C displays than for Pattern B displays (see Figure 10). This occurred because the nearest neighbors were farther from the targets in Pattern B displays than in Pattern C displays. (Reaction times were slower in Pattern A displays than in Pattern B displays because the nearest neighbors were closer to the targets.) Second, and more important, the parallel models failed to predict the advantage of the isolated target condition over the good-figure condition. The near neighbors were the same distance away from the targets in the two conditions, so reaction times and accuracies were the same, contrary to what Banks and Prinzmetal (1976) observed.

The serial models captured the essential features of Banks and Prinzmetal's (1976) data. Pattern C was faster than Pattern B, and Pattern B was faster than Pattern A. More importantly, the isolated target displays were faster and more accurate than the good-figure displays in each pattern.

The feature catches (c_x) used in the fits for each pattern and condition are presented in Table 8. There is a feature catch for the target, for the near neighbor, the middle neighbor, and the far neighbor in each pattern. The correspondence between these values and Banks and Prinzmetal's (1976) displays can be gleaned from Figure 10A, Condition 5. *Target* corresponds to the position of the target; *near* corresponds to the position of the distractor immediately to the right or immediately below

Table 7

Observed and Predicted Reaction Times and Predicted Percent Correct Scores for Banks and Prinzmetal (1976)

Pattern	Condition				
	1	2	3	4	5
Banks and Prinzmetal's (1976) data					
A	598	558	609	640	759
B	573	551	638	667	735
C	557	552	649	703	683
Parallel processing					
A	585 99	585 99	671 99	671 99	675 99
B	536 99	536 99	670 99	670 99	674 99
C	586 99	586 99	676 99	676 99	676 99
Serial processing					
A	607 91	553 97	628 89	628 89	711 82
B	597 91	549 97	623 90	623 90	705 83
C	570 95	552 97	709 82	709 82	709 82

the target; *middle* corresponds to the position of the distractor along the diagonal; and *far* corresponds to the position of the distractor in the bottom left or top right positions. The distractor in the bottom right position was not considered in the fits. The entries in Table 8 correspond to target and distractor positions that were employed in the fits. Thus, for example, the near condition is blank in Pattern A, Conditions 1 and 2 because there were no very near neighbors in those displays (see Figure 10).

The feature catches for the same-threshold fits were .757, .114, .036, and .007 for all patterns and conditions.

Evaluation. CTVA did a reasonable job of accounting for Banks and Prinzmetal's (1976) data. As they anticipated, between-object effects were the most important factors in our account. However, in the CTVA analyses, serial processing models did better than parallel processing models. Contrary to Banks and Prinzmetal's suggestion, CTVA had to assume a serial search strategy that focused on the most isolated position first in order to capture the advantage of isolated target displays over good figure displays. These fits encourage further investigations.

Cohen and Ivry (1991): Density Effects in Conjunction Search

The difficulty of searching for targets that are conjunctions of separable features is another cornerstone prediction of Treisman's feature integration theory (Treisman & Gelade, 1980; Treisman & Sato, 1990). Reaction time increases as a linear function of display size with a steep slope. Since the original demonstration that conjunction search was harder than search

for the features that made up the conjunction, the result has been investigated vigorously. Often, the original result replicates, but some researchers have shown that conjunction search is sometimes easy, producing slopes near zero (e.g., Wolfe et al., 1989) and sometimes feature search is hard, producing slopes well above zero (e.g., Treisman & Gormican, 1988). Recently, Cohen and Ivry (1991) found that the conjunction search slope could be reduced considerably if the density of the displays was reduced by increasing the distance between adjacent items. This distance effect falls in the domain of CTVA, and CTVA accounts for it in a way that is similar to its account for distance effects on illusory conjunctions.

CODE. Conjunction search is difficult when distractors share features with the target (e.g., the distractors are green *X*'s and red *O*'s, and the target is a red *X*). CTVA interprets this as a similarity effect; the $\eta(x, i)$ values for distractors are high in conjunction search. In order to avoid target-present responses to distractors, the feature catch has to focus on individual items, either serially—one by one—or in parallel with separate spatial indices for each item. The threshold is set high so that the contribution of adjacent items to the feature catch is much smaller than the contribution of the item in the current focus of attention. Increasing distance between items has two effects in the theory: First, it decreases the overlap of distributions from adjacent items, and that decreases the probability of sampling features from adjacent items and reduces the likelihood of false target-present responses. Second, it lowers the local minima on the CODE surface, and that lets the system adopt a lower threshold (i.e., just above the local minimum). That increases the contribution of the target item to the feature catch and speeds processing (see Equations 12 and 10). Both of these factors would speed the search rate, as Cohen and Ivry (1991) observed.

CTVA. In order to apply TVA to conjunction search, the rules by which responses are chosen must be specified. Consider the case in which the target is a red *T* and the distractors are green *T*s and red *X*s. The person decides a target is present if he

Table 8

Feature Catches (c_x) for Parallel and Serial Fits to Banks and Prinzmetal (1976)

Pattern	Target	Near	Middle	Far
A				
Condition 1	.864		.067	.013
Condition 2	.864		.067	.013
Condition 3	.757	.114	.036	.007
Condition 4	.757	.114	.036	.007
Condition 5	.757	.114	.036	.007
B				
Condition 1	.941			.029
Condition 2	.941			.029
Condition 3	.757	.114		.007
Condition 4	.757	.114		.007
Condition 5	.757	.114		.007
C				
Condition 1	.864		.067	
Condition 2	.864		.067	
Condition 3	.757	.114		
Condition 4	.757	.114		
Condition 5	.757	.114		

or she determines that a perceptual object is both red and T . The person decides that a perceptual object is not a target if he or she determines that the object is not red or not T . Thus, the decision rules are as follows:

1. **IF** red **AND** T **THEN** terminate search and say "present."
2. **IF** not red **OR** not T **THEN** examine the next item.
3. **IF** there are no more items to be examined **THEN** say "absent."

In TVA, the time to decide that an object is red depends on the rate of processing, $v(x, \text{red})$ and the time to decide it is T depends on $v(x, T)$. The time to decide that an object is not red depends on the rate at which "not red" can be detected— $v(x, \text{notred})$, and the time to decide that an object is not T depends on $v(x, \text{not}T)$. In order to decide that a target is present, both red and T must be detected. Thus, the time to decide that a target is present is the maximum of the times required to decide that the object is red and it is a T : $\max(\text{red}, T)$. In order to decide that an object is not a target, it is sufficient to detect either not red or not T . Thus, the time to decide that an object is not a target is the minimum of the time to decide it is not red and the time to decide it is not T : $\min(\text{notred}, \text{not}T)$.

The decision about whether a given object is a target depends on a race between the process that decides an object is a target and the process that decides an object is not a target. The mathematics underlying the race are developed in Appendix D. When CODE and TVA are put together, the $v(x, i)$ values are modified by the feature catches, c_x , so reaction time and accuracy depend both on the factors that affect $v(x, i)$ (i.e., similarity between targets and distractors) and the factors that affect c_x (i.e., the density of the items in the display).

The finishing times and accuracies for individual comparisons must be combined over items to predict mean reaction time and accuracy for the whole display. The traditional way to do this in the conjunction search literature is to assume serial self-terminating processing, in which attention focuses on the items one by one (e.g., Cave & Wolfe, 1990; Treisman & Gelade, 1980; Treisman & Sato, 1990; Wolfe, 1994; Wolfe et al., 1989). In models like these, reaction time is the sum of the individual comparison times plus an additive constant, and accuracy is the product of the accuracies of the individual comparisons. It is also possible to combine individual finishing times and accuracies in various parallel models (e.g., Duncan & Humphreys, 1989; Pashler, 1987; Townsend & Ashby, 1983). Parallel models must find some way to keep individual items distinct from each other, or else illusory conjunctions would inflate the error rate. One way to keep items distinct is to spatially index them, and while spatial indexing is often thought of as a serial process (e.g., Ullman, 1984), Pylyshyn (1989) and colleagues (e.g., Pylyshyn & Storm, 1988; Trick & Pylyshyn, 1994) suggested that four or more spatial indices may be deployed in parallel. In principle, CTVA could provide a parallel-processing account of conjunction search if it assumed multiple spatial indices. For the present purposes, however, I configured CTVA as a serial self-terminating search process to make the exposition clearer. The key results depend on the finishing times and accuracies for individual items (see Appendix D). They should have the same effect on overall reaction time and accuracy no matter how they are combined.

In a serial, self-terminating version of CTVA, mean reaction time for target-present responses is

$$RT_P = FT_P + \left(\frac{D-1}{2}\right)FT_A + b, \quad (14)$$

where FT_P and FT_A are the finishing times for the processes that detect target presence and absence, respectively, D is display size, and b is an additive constant that represents residual time for perceptual encoding and motor processing. For target-absent responses, mean reaction time is

$$RT_A = (D)FT_A + b. \quad (15)$$

According to Equations 14 and 15, reaction time is a linear function of display size with a slope that depends on the finishing time of the process that detects target absence. The slope for target-present responses is approximately half the slope for target-absent responses, as is commonly found in conjunction search experiments (Treisman & Gelade, 1980).

The CTVA model also predicts accuracy, although Cohen and Ivry (1991) did not report it. The probability of a correct response for target-present displays is

$$P(C|P) = 1 - [1 - P(P)] \cdot P(A)^{D-1}, \quad (16)$$

where $P(P)$ and $P(A)$ are the probabilities that the processes that detect target presence and absence, respectively, function correctly. The probability of correctly detecting the target is one minus the probability of missing the target when it is present. The system will miss a target when it is present if it fails to detect the target, with probability $1 - P(P)$, and if it fails to false alarm to a distractor, with probability $P(A)^{D-1}$. These probabilities are independent, so they combine multiplicatively to produce the miss rate, which is subtracted from 1 in Equation 16 to produce the hit rate.

The probability of a correct response for target-absent displays is

$$P(C|A) = P(A)^D. \quad (17)$$

Averaged over experiments and target-present and target-absent conditions, the mean slope in Cohen and Ivry's (1991) clumped condition was 34 ms per item, and the mean slope in their spread condition was 14 ms per item. I estimated the slopes by calculating finishing times for individual comparisons using the equations in Appendix D. The finishing times can be inserted into Equations 14 and 15 to predict mean reaction times as a function of display size. I chose not to predict mean reaction times because there was considerable variation in slopes and in the ratio between target-present and target-absent slopes in Cohen and Ivry's (1991) experiments, which suggested that the serial, self-terminating model on which Equations 14 and 15 are based may not be appropriate for all of their data. Indeed, Cohen and Ivry (1991) argued that search was parallel in some of their conditions. Consequently, the CTVA analysis focused on finishing times, which would support predictions for parallel search models as well as serial, self-terminating ones.

The distances between items were set to 100 units in the clumped condition and 200 units in the spread condition. The standard deviation of the feature distributions was set to 50 in

each case. The threshold was set halfway between the local minimum between items and the peak of the feature distribution, in order to focus more sharply on the target item. With these parameters, the mean finishing time for the target-absent process was 3.12 units in the clumped condition and 1.67 units in the spread condition. Finishing times for target-absent processes determine the slopes in Equations 14 and 15; if one unit equals 10 ms, the predicted results are close to the average values in Cohen and Ivry's (1991) experiments.⁸

In the clumped condition, the feature catch included 39.4% of the target's feature distribution and 7.1% of each of the distractor's feature distributions. In the spread condition, the feature catch included 63.2% of the target's feature distribution and 2.2% of each of the distractor's feature distributions. The probability of correctly deciding target absence was .972 in the clumped condition and .992 in the spread condition. The mean finishing time and accuracy for the process that detected target presence was 3.04 units and .821 for the clumped condition and 2.51 units and .967 for the spread condition.

Evaluation. The CTVA model did a reasonable job of accounting for Cohen and Ivry's (1991) results. Cohen and Ivry (1991) proposed two processing mechanisms underlying their results: (a) a fine-grained localization process akin to Treisman's conception of conjunction search in the clumped condition and (b) a coarse-grained localization process different from Treisman's conception in the spread condition. By contrast, the CTVA analysis accounts for both conditions with the same processing mechanisms. The only difference between the conditions is the spacing of the item's feature distributions. While it remains possible that different mechanisms underlie performance in the different spacing conditions, the CTVA analysis suggests that further research with more incisive experiments will be necessary to rule out theories (such as CTVA) that propose a single mechanism.

Wolfe, Cave, and Franzel (1989): Double Versus Triple Conjunction Search

Most experiments in the conjunction search literature involve conjunctions of two features, or *double conjunctions*. Wolfe, Cave, and Franzel (1989) tested people in a *triple conjunction* task, in which targets were conjunctions of three features (e.g., large red Ts) and distractors contained only one of the target features (e.g., large green Xs, small red Xs, or small green Ts). They found that this triple conjunction search was much easier than double conjunction search. They interpreted their results in terms of preattentive processes rather than the attentive comparison process that operates on the selected items, arguing that triple conjunctions stood out more from the background of distractor items. To account for the triple conjunction results, they proposed *guided search* theory (Cave & Wolfe, 1990; Wolfe, 1994) as an extension of Treisman's feature integration theory. Whereas Treisman argued that preattentive and attentive processes were independent, Wolfe et al. (1989) argued that preattentive processes interacted with attentive processes, suggesting likely candidates for attention to focus on. Triple conjunction search was easy, they argued, because preattentive processes segregated triple conjunction targets from the

distractors more easily than they segregated double conjunction targets.

The field appears to have accepted the interpretation offered by Wolfe et al. (1989). Treisman and Sato (1990) revised feature integration theory to account for the triple conjunction results, proposing an inhibitory interaction between attentive and preattentive processes. They argued that inhibition was more effective when targets differed more from distractors, as with triple conjunctions that share only one feature with the target. Others in the field apparently agree with Wolfe and Treisman. Grossberg, Mingolla, and Ross (1994) modeled the Wolfe et al. (1989) data and attributed the ease of triple conjunction search to preattentive grouping processes. Surprisingly, no one appears to have tried to interpret the advantage of triple conjunctions in terms of the attentive comparison process.

In the course of modeling Cohen and Ivry's (1991; double) conjunction search data, it occurred to me that TVA can be extended easily to account for triple conjunction search. Moreover, TVA accounts for the advantage of triple conjunction search over double conjunction search in terms of the attentive comparison process and not in terms of preattentive grouping processes. Thus, TVA offers a new approach to the analysis of triple conjunction search that differs significantly from other current approaches (e.g., Grossberg et al., 1994; Treisman & Sato, 1990; Wolfe et al., 1989).

The TVA analysis can be extended to triple conjunction search by simply including a $v(x, i)$ value for each of the three features and their absence (i.e., not small, not red, and not T). As with double conjunctions, decisions about target presence are determined by the outcome of a race between the presence and absence of the target features:

$$Outcome_3 = \min[\max(r, T, s), \min(\bar{r}, \bar{T}, \bar{s})].$$

Notice that there are three runners in the race for target absence, which is more than the two that raced for target absence in standard conjunction search. This is important because the fastest of three runners will finish before the faster of two runners (see e.g., Logan, 1988, 1992), and this will reduce the slope of the function relating reaction time to display size because the slope is determined by the rate at which target absence is decided. Thus, the TVA analysis predicts shallower slopes in triple conjunction search than in standard, double conjunction search. A formal derivation of the finishing times and accuracies for target-present and target-absent responses is presented in Appendix D. The finishing times developed in Appendix D can be put into Equations 14 and 15 to predict slope values.

Finishing times for target-present and target-absent processes

⁸ Similar results are obtained if the threshold is set to the local minimum. The accuracies are a little lower and the finishing time difference between the clumped and spaced conditions is a little smaller. Using the same spacing parameters and η and λ values but setting the threshold to the local minimum, the finishing times (and accuracies) were 1.985 (.956) and 1.298 (.989) for clumped and spaced target-absent processes, respectively, and 3.037 (.821) and 1.958 (.954) for clumped and spaced target-present responses. The feature catches were .632 and .865 for clumped and spaced targets, respectively, and .159 and .066 for clumped and spaced distractors.

Table 9
Processes Detecting Target Presence and Target Absence
in Double and Triple Conjunction Search

Task	Finishing times		Accuracies	
	Target present	Target absent	Target present	Target absent
Double conjunction	4.742	3.122	.887	.972
Triple conjunction two features different	5.628	1.628	.763	.999
Triple conjunction one feature different	5.743	3.014	.849	.979

were modeled using the parameters from the clumped condition from the Cohen and Ivry (1991) fits. The target was 100 units from the neighboring distractors; the standard deviation of the feature distributions was 50; the η value representing the similarity between a feature and its absence was .01; and the threshold was set halfway between the local minimum and the peak of the target's feature distribution. Table 9 contains the predicted finishing times and accuracies.

Several effects in Table 9 are noteworthy. First, the time to detect target presence is greater for triple conjunctions than double conjunctions because the maximum of three values is larger than the maximum of two values (i.e., $\max(\text{red}, \text{large}, T) > \max(\text{red}, T)$). Second, the time to detect target absence, upon which the search slope depends (see Equations 14 and 15), is shorter for triple conjunctions than for double conjunctions. The difference is large when triple conjunction distractors differ from targets on two features, as Wolfe et al. (1989) observed. The difference is smaller when triple conjunction distractors differ from targets on only one feature, also as Wolfe et al. (1989) observed. Thus, TVA appears to account for the main trends in the data of Wolfe et al. (1989). Interestingly, it attributes the effects to the attentive comparison process rather than the preattentive processes.

Evaluation. TVA did a good job of accounting for the main differences between double and triple conjunction search. The credit goes entirely to TVA; Bundesen's (1990) theory can account for the differences without recourse to CODE. It is significant that TVA accounts for the differences entirely in terms of the attentive comparison process. This is important because the other approaches to the contrast between double and triple conjunctions do not attempt to model the comparison process. Some, such as Cave and Wolfe (1990) and Grossberg et al. (1994), provide formal accounts of preattentive processes but not the attentive comparison process. The TVA analyses suggest that other researchers may have misattributed the advantage of triple conjunction search to preattentive processes. The TVA analyses demonstrate that at least one construal of attentive processes—and one with considerable predictive power (Bundesen, 1990)—accounts for the advantage of triple conjunctions in terms of attention rather than preattention. In the context of that theory, it is possible that all of the advantage of triple conjunction is due to attentive processes or that part of the advantage is due to attention and part is due to preattention. It

is not possible, in the TVA theory, to account for the advantage entirely in terms of preattentive processes.

Target-Distractor Discriminability and the Size of the Spotlight

Several researchers have been concerned with factors that determine the size of the region that attention selects. Eriksen and St. James's (1986) zoom lens theory assumes that the spotlight can expand and contract according to task demands, but the resolving power diminishes as the size increases. Treisman and Gormican (1988) assumed that the size of the spotlight increased as the difficulty of discrimination decreased. Easy discriminations could be done in parallel all over the visual field, whereas difficult discriminations required sharply focused attention. Duncan and Humphreys (1989) made similar assumptions, arguing that the rate of processing increased as the similarity between targets and distractors decreased.⁹

The search literature provides strong empirical support for these speculations about the size of the spotlight, processing rate, and processing power. The rate of processing in visual search tasks, measured as the reciprocal of the slope of the linear function that relates reaction time to the number of items in the display, decreases as the difficulty of discrimination increases (Treisman & Gormican, 1988), and it decreases as the similarity between targets and distractors increases (Duncan & Humphreys, 1989).

CTVA provides a straightforward account of the relation between discrimination difficulty and search rate. According to Equation 7, accuracy depends on the ratio of the correct $v(x, i)$ value to the sum of the other $v(x, j)$ values. Extending Equation 7 to sum over all of the items in the display and rearranging the terms yields

$$P(x \in i \text{ first}) = \frac{v(x, i)}{\sum_{z \in S} \sum_{j \in R} v(z, j)} = \frac{v(x, i)}{v(x, i) + \sum_{\substack{z \in S/j \in R \\ x, i \neq z, j}} v(z, j)} \quad (18)$$

Equation 18 makes it clear that accuracy depends on the sum of the rates of processing of the other categorizations of the other items in the display, that is, $\sum \sum v(z, j)$. If that sum is large, accuracy is low. If that sum is small, accuracy is high. The magnitude of the sum depends on the similarities of the other categorizations to the correct categorization and on the number of other (similar) items in the display.

⁹ Duncan and Humphreys' (1989) theory is not easily characterized as a spotlight theory because they were not specific about the mechanism of selection. A recent extension of their theory by Humphreys and Müller (1993) seems contrary to the spotlight view, because it assumes that processing is always parallel (i.e., distributed over the whole visual field rather than focused in a single beam).

From Equation 12, the rate of processing for any given item is

$$v(x, i) = c_x \eta(x, i) \beta_i \frac{w_x}{\sum_{z \in S} w_z}$$

The processing rate, $v(x, i)$, depends on the magnitude of the feature catch, c_x , and the similarity between the item and the target (or nontarget) categorization, $\eta(x, i)$. Processing rate can be held constant as one or the other of these factors increases if the other decreases by a corresponding amount. Thus, c_x and $\eta(x, i)$ trade off against each other. A low c_x and a high $\eta(x, i)$ can produce the same processing rate as a high c_x and a low $\eta(x, i)$, and that provides the account of the relation between processing rate and discriminability. When the discriminability between targets and distractors is high, $\eta(x, i)$ for the distractors is low, so c_x can be high. A high c_x occurs when the threshold is set low and several items are processed at once, in parallel. Thus, high discriminability allows parallel processing. By contrast, when discriminability between targets and distractors is low, $\eta(x, i)$ for the distractors is high, so c_x must be set low to prevent false categorizations of distractors. The model lowers c_x by raising the threshold so that items are segregated perceptually from each other, and that segregation promotes serial processing. Thus, low discriminability encourages serial processing. High discriminability produces big spotlight beams and low discriminability produces small ones, consistent with the data and with theorists' speculations.

Eriksen and Eriksen (1974): Distance Effects With Distracting Flankers

Eriksen and Eriksen (1974) published an important article on which much of the debate over space-based and object-based attention was grounded. They showed that people were influenced by distractor items that flanked the target even when there was no uncertainty about target location. Flankers that were associated with the same response as the target facilitated reaction time and accuracy, whereas flankers that were associated with the opposite response from the target impaired reaction time and accuracy. These effects are modulated by distance between targets and flankers and by factors that place the target and flankers in the same or different perceptual groups, so the task is an important test case for CTVA.

Method and results. Eriksen and Eriksen (1974) presented their subjects with displays like those in Figure 13. The task was to determine the identity of the central letter and move a lever to the left or right, depending on the letter. Two letters were mapped onto each response. *H* and *K* were mapped onto one response, and *S* and *C* were mapped onto the other. The central letter always appeared in the same position, .5° above the fixation point. The displays were exposed for 1 s, so reaction time was the most important dependent variable.

The most important independent variable was the compatibility of the flankers and the target. In *response compatible* displays (Conditions 1 and 2), the target and the flankers both called for the same response. In *response incompatible* displays (Condition 3), the target and flankers called for opposite re-

1. Noise Same as Target

```

      HHHHHHH
    HHH H HHH
  HHH   H   HHH

```

2. Noise Response Compatible

```

      KKKHKKK
    KKK H KKK
  KKK   H   KKK

```

3. Noise Response Incompatible

```

      SSSHSSS
    SSS H SSS
  SSS   H   SSS

```

4. Noise Heterogeneous - Similar

```

      NWZHNWZ
    NWZ H NWZ
  NWZ   H   NWZ

```

5. Noise Heterogeneous - Dissimilar

```

      GJQH GJQ
    GJQ H GJQ
  GJQ   H   GJQ

```

6. Target Alone

```

      H

```

Figure 13. Examples of displays from Eriksen and Eriksen's (1974) experiments, showing noise same as target (top panel), noise response compatible (second panel), noise response incompatible (third panel), noise heterogeneous and similar (fourth panel), noise heterogeneous and dissimilar (fifth panel), and target alone (bottom panel), and showing the distance manipulation (Conditions 4 and 5 are neutral).

sponses. The other conditions were controls that can be used to assess facilitatory and inhibitory components of the *response compatibility effect* (i.e., the difference between compatible and incompatible displays). Distance between the target and flankers (.06°, .5°, and 1.0°) was the other important independent variable.

The results, displayed in Figure 14, showed a strong response compatibility effect. The difference in reaction time was large when the flankers were close to the target and diminished as distance increased. This effect is very robust, having been replicated many times with many variations on the procedure. The number of flanking letters does not seem to be a crucial factor; similar effects can be obtained with one (Andersen, 1990; Flowers & Wilcox, 1982; Kramer & Jacobson, 1991) and two (Eriksen & Schultz, 1979; Coles, Gratton, Bashore, Eriksen, & Donchin, 1985) on each side.

The distance between the target and the flankers is important, regardless of the number of flankers. The response compatibil-

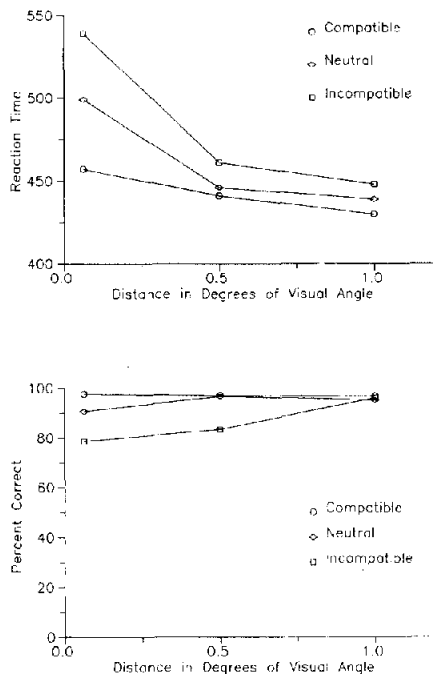


Figure 14. Mean reaction times (top panel) and accuracy (bottom panel) as a function of distance between the target and the flanking distractors in Eriksen and Eriksen's (1974) experiment.

ity effect decreases as distance increases in many experiments (Andersen, 1990; Eriksen & Hoffman, 1972; Flowers & Wilcox, 1982; Kramer & Jacobson, 1991). The distance effect is interpreted as strong evidence for space-based attention. The decline in facilitation and interference with distance is interpreted in terms of the width of the beam of the attentional spotlight. Facilitation and interference occur to the extent that the flankers fall within the beam (but see van der Heijden, 1992).

The response compatibility effect is also influenced by Gestalt grouping principles that determine whether the target and distractors are seen as part of the same or different perceptual groups (or objects). Facilitation and interference are stronger when the target and distractors are part of the same group than when they are part of different groups (Baylis & Driver, 1992; Driver & Baylis, 1989; Harms & Bundesen, 1983; Kramer & Jacobson, 1991; also see Kahneman & Henik, 1981). This is interpreted as strong evidence for object-based attention: Object-based theorists argue that attention selects all of the properties of the selected object, relevant and irrelevant. Distractors are processed when they are selected together with the target—when they fall in the same group—but not when they fall in different groups.

Object-based research on the Eriksen and Eriksen (1974) paradigm has ignored space in general and proximity in particular as an organizing principle, conceding proximity to the space-based opposition. This is surprising because the spacing effects in the original Eriksen and Eriksen (1974) experiment could have been due to grouping rather than distance itself. In their displays, depicted in Figure 13, the distance between the flankers was held constant as the distance between targets and

flankers increased. This manipulation would cause the distractors to be grouped together by proximity and separated from the target.

The confounding of grouping and distance raises an important question: How do subjects know which item is the target? The original design of the Eriksen and Eriksen (1974) task was intended to remove the requirement of locating the target. The display appeared in the same position from trial to trial and the target always appeared in the same position relative to the other items in the display and relative to the fixation point. Thus, target location was highly predictable. Nevertheless, the predictability of target location does not mean that subjects did not have to engage in some processing to find it. Even if the target is always the middle item, subjects must need to compute *middle* in order to find it (Logan, 1995). It may have been easier to find the target when distractors were separated.

Grouping effects based on principles other than proximity are outside the scope of CTVA. The between-item effects involved in finding the middle item are also outside the immediate scope of CTVA. While there may well be between-item effects in the Eriksen and Eriksen (1974) task, there are certainly within-item effects, and those within-item effects are the focus of the CTVA analysis.

CODE. The CODE analysis defines the feature catch in the Eriksen and Eriksen (1974) task. For simplicity, I focused on a three-item version of the paradigm, with one target and two identical flanking distractors, rather than the seven-item version Eriksen and Eriksen (1974) initially studied. Results are similar across three- and seven-item versions. CODE analysis would suggest that the outside flankers in the seven-item version are too far from the target to have much of an impact on it.

Feature distributions and the CODE surface for the Eriksen and Eriksen (1974) task are presented in Figure 15. The top panel illustrates a narrow spacing condition, and the bottom panel illustrates a wide spacing condition. Also illustrated in Figure 15 are thresholds for each condition, set just above the local minimum between the target and the flankers, which define the feature catch.

Two effects of the distance between flankers and targets are apparent in Figure 15. First, when the flankers are close, a much greater area of their distributions falls within the feature catch than when the flankers are far. Thus, flankers should have a greater impact on the feature catch when the flankers are close than when they are far from the target. Second, when the flankers are close, the local minimum between target and distractors is higher, so the threshold is higher. The area of the target's feature distribution that falls within the feature catch is smaller in close-spaced displays than in far-spaced displays, so overall reaction time should be slower.

Both of these effects are observed in the literature: Many investigators report a diminution in the flanker effect as distance increases (Andersen, 1990; Eriksen & Hoffman, 1972, 1973; Flowers & Wilcox, 1982; Kramer & Jacobson, 1991). Those same studies found faster reaction times with greater target-flanker distances, though most investigators did not comment on that effect.

CTVA. The Eriksen and Eriksen (1974) paradigm presents four different kinds of stimuli that need to be represented in

CTVA: Two alternative targets (e.g., H and S) and two alternative distractors (response compatible and response incompatible). Setting the η values is straightforward: η for the target H , given that an H is presented, should be 1.0. η for the distractor H , given that the target is an H , should be 1.0. η for the target S , given that the target is an H , should be between 1.0 and 0.0. So should η for the distractor S , given the target H . β should be the same—1.0—for the two alternative targets, and w_x should be the same—1.0—as well.

This parameterization of TVA, which is much like the one for the preceding paradigms, fails to produce the basic Eriksen and Eriksen (1974) results. It predicts a higher error rate on response incompatible trials than on response compatible trials, but it predicts equivalent reaction times. TVA gets the ordering of difficulty right—response incompatible displays are harder than response compatible ones—but it predicts that the effects will appear in error rate rather than reaction time, and the results are nearly always the opposite. The major effects are on reaction time; the effects on accuracy are weak or nonexistent (see e.g., Eriksen & Eriksen, 1974).

The faulty predictions result from construing TVA as a simple race model. The probability of a correct response depends on the ratio,

$$P(\text{correct}) = \frac{v(T, H) + v(D_1, H) + v(D_2, H)}{v(T, H) + v(T, S) + v(D_1, H) + v(D_1, S) + v(D_2, H) + v(D_2, S)} \quad (19)$$

On response compatible trials, $v(D_1, H)$ and $v(D_2, H)$ will be large, because the flankers, like the target, are H s. On response incompatible trials, however, $v(D_1, H)$ and $v(D_2, H)$ will be small because the flankers are S s rather than H s. Consequently, the probability of a correct response will be higher on compatible trials than on incompatible trials, by an amount that depends on the magnitude of $v(D_1, H)$ and $v(D_2, H)$. So TVA predicts more errors on incompatible trials, and the difference in error rate may be quite large.

Mean reaction time depends on the denominator of Equation 19 (following the logic expressed in Equations 6 and 12 and the derivation in Appendix A). On compatible trials, $v(D_1, H)$ and $v(D_2, H)$ will be large and $v(D_1, S)$ and $v(D_2, S)$ will be small, because the flankers are H s and not S s. The situation is reversed on incompatible trials. The values of $v(D_1, S)$ and $v(D_2, S)$ will be large and $v(D_1, H)$ and $v(D_2, H)$ will be small, because the flankers are S s and not H s. The important point is that the magnitude of the denominator will be the same in both cases. What is lost in $v(D, H)$, in going from compatible to incompatible trials, is gained in $v(D, S)$. What is lost in $v(D, S)$, in going from incompatible to compatible trials, is gained in $v(D, H)$. Consequently, mean reaction time will remain the same: TVA cannot account for the ubiquitous compatibility effect on reaction time.¹⁰

In order to fit the Eriksen and Eriksen (1974) results, I configured TVA a counter model (Townsend & Ashby, 1983), letting the race run until several "runners" had finished. There were two counters, with criteria $K_H = K_S = 3$, and the counting process finished as soon as one counter accumulated its criterion number of counts. The probability of responding correctly and mean reaction time for correct responses were computed from Equations 9 and 10. The $v(x, i)$ values were determined by setting η equal to 1.0 for H , given H and S , given S , .01 for H given S and S given H , and .5 for H or S given a neutral distractor. β and w_x were set to 1.0. There were two distractors, located 50, 100, and 150 units on either side of the target. The standard deviation of target and distractor feature distributions was 50. The results of the fits are plotted in Figure 16.

¹⁰ I think this prediction is generally true of race models. The instance theory of automaticity (Logan, 1988), for example, cannot account for the Stroop (1935) effect. Reaction time should be just as fast on incompatible trials as on compatible trials because the word should retrieve the same number of traces in both cases. Accuracy should be much lower on incompatible trials because there should be more word traces than color traces in the race, so the word should be more likely to win. The subject should produce an error whenever the word wins on an incompatible trial, so error rate should be very high. This problem can be solved by allowing the retrieval process to retrieve more than one trace before terminating. The retrieval process in the instance theory could drive a counter model, as I have done here with CTVA, or it could drive a random walk model, as in Nosofsky and Palmeri's (in press) exemplar-based model of speeded classification. These models are straightforward generalizations of the simple race model, and the statistics underlying the retrieval process remain the same.

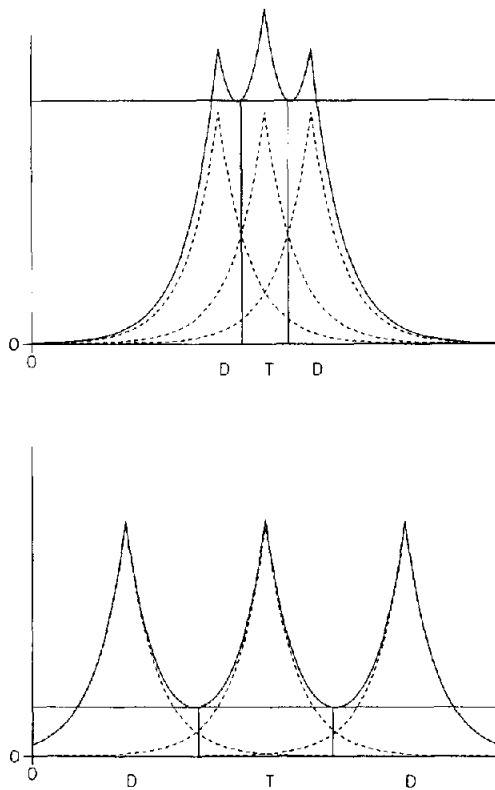


Figure 15. The CODE surface for a three-item version of Eriksen and Eriksen's (1974) experiment with a threshold applied just above the local minimum between the central target (T) and flanking distractors (D). Top panel = narrow spacing; bottom panel = wide spacing.

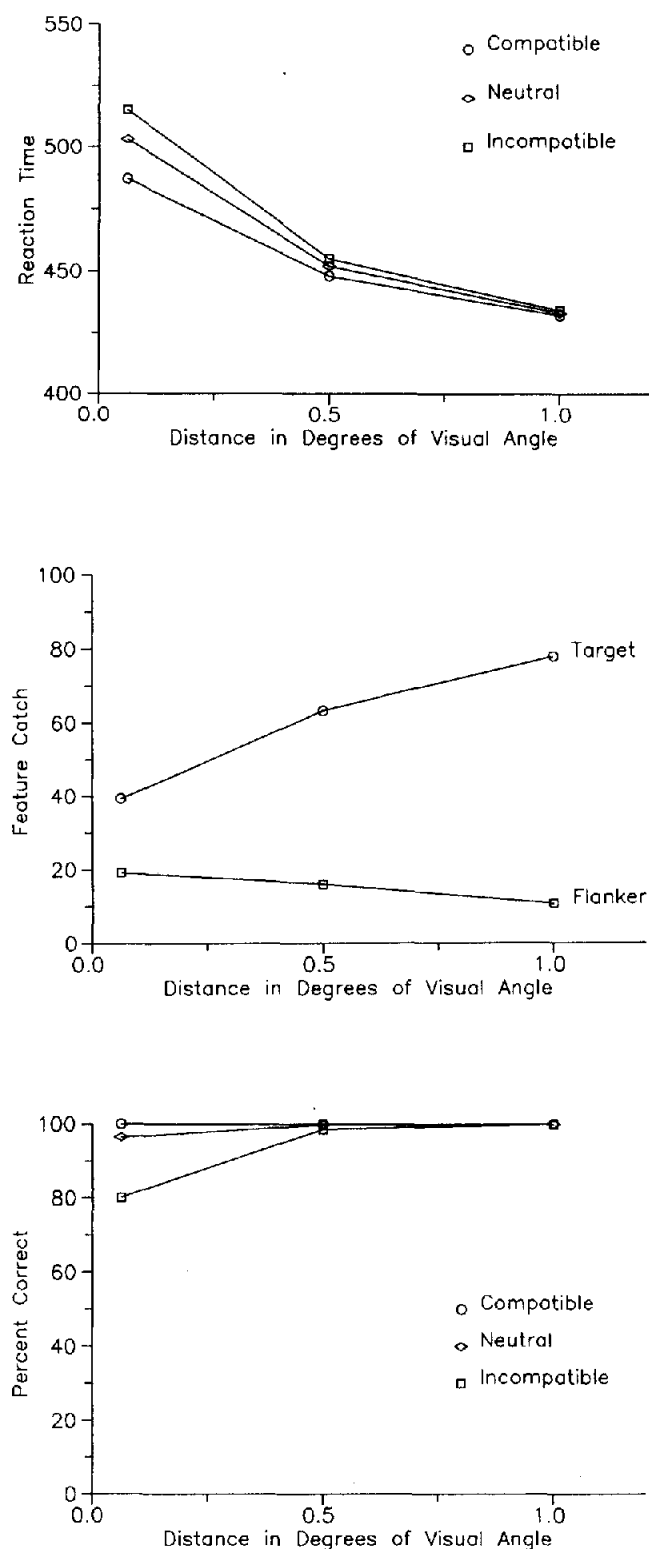


Figure 16. Mean reaction times for compatible, neutral, and incompatible conditions (top panel), feature catches for target and noise items (middle panel), and accuracy (expressed as percent correct) for compatible, neutral, and incompatible conditions (bottom panel) predicted by the CODE theory of visual attention for the Eriksen and Eriksen (1974) experiment.

The mean reaction times, in the top panel of Figure 16, correlate strongly with the observed data in Figure 14: $r = .901$. The predicted reaction times show the two patterns characteristic of distance effects in the Eriksen and Eriksen (1974) paradigm. First, the compatibility effect decreases as distance between the target and the distractors increases. There is a strong compatibility effect when the distance is small. Compatible responses are faster than neutral responses, which in turn, are faster than incompatible responses. The ordering of conditions remains the same as distance increases, but the magnitude of the differences decreases. Thus, CTVA captures the effect reported many times in the literature (Andersen, 1990; Eriksen & Hoffman, 1972, 1973; Flowers & Wilcox, 1982; Kramer & Jacobson, 1991). The predicted compatibility effects in Figure 16 are smaller than the observed ones in Figure 14, perhaps because between-object effects, which were not modeled, contributed to the observed effects.

Second, mean reaction time in all conditions decreases as the distance between targets and distractors increases. Averaged over compatibility conditions, the predicted effects were close to the observed ones (502, 452, and 433 ms predicted vs. 498, 449, and 439 ms observed). The distance effect is found in all experiments in which distance is manipulated, although the investigators typically do not comment on it (Andersen, 1990; Eriksen & Hoffman, 1972, 1973; Flowers & Wilcox, 1982; Kramer & Jacobson, 1991). The reduction in reaction time with distance is predicted by CTVA. The threshold is set at the local minimum in the CODE surface between the target and the distractors. As distance increases, the local minimum moves farther away from the target and this has two effects, both of which speed processing. It increases the feature catch, c_x , from the target and it decreases the feature catch from the distractors. These effects can be seen in the middle panel of Figure 16, which plots the feature catch for the target and one of the distractors as a function of distance.

Predicted response accuracy is plotted in the bottom panel of Figure 16. Accuracy is at ceiling for response compatible displays but varies as a function of distance for incompatible and neutral displays. At the shortest distance, accuracy is well below ceiling for response incompatible displays, increasing rapidly as distance increases. Accuracy for neutral displays is intermediate between compatible and incompatible displays, increasing as distance increases. These results capture the pattern observed by Eriksen and Eriksen (1974) presented in Figure 14.

Evaluation. The CTVA analysis accounted for the major effects in the Eriksen and Eriksen (1974) paradigm. Response compatible displays were more difficult than neutral displays, which in turn, were more difficult than response incompatible displays. Construing CTVA as a counter model (rather than a simple race model) allowed it to account for the difficulty in terms of the appropriate dependent variable, showing strong effects on reaction time and substantial effects on accuracy, as Eriksen and Eriksen (1974) found. The CTVA analysis predicted an overall reduction in reaction time as distance increased, which is commonly observed in the Eriksen and Eriksen (1974) paradigm but has never before been accounted for. Zoom lens models, such as Eriksen and St. James's (1986), would predict the opposite result, because they argue that resources are spread more thinly as the spotlight expands, and spreading resources more thinly slows responding.

The CTVA analysis does not account for between-object effects, which have been shown to be important in the Eriksen and Eriksen (1974) paradigm (e.g., Baylis & Driver, 1992; Driver & Baylis, 1989; Harms & Bundesen, 1983; Kramer & Jacobson, 1991). Inspection of Eriksen and Eriksen's (1974) displays in the present Figure 13 suggests that between-object grouping by proximity effects may contribute something to the distance effect. These effects are beyond the scope of the current version of CTVA but remain an important topic for future research.

General Discussion

Answers to the Five Questions

This article began with five questions that challenge any theory of visual spatial attention. How does CTVA address them? The answers have been implicit in the exposition of the theory throughout the paper. Now it is time to make them explicit.

How is space represented? The theory assumes that space is represented in two ways. From a bottom-up perspective, space is represented as a CODE surface. Each object is distributed in space and the CODE surface is the sum of the distributions. The CODE surface is determined completely by bottom-up processes. The theory assumes it is constructed by obligatory parallel processes that operate simultaneously over the whole visual field. From a top-down perspective, space is represented in terms of perceptual groups defined by the intersection of the CODE surface and a threshold. The threshold is set by top-down processes, and top-down processes can operate on the groups produced by the threshold setting. Processes that apprehend spatial relations, for example, may operate on the groups that CODE provides (Logan & Sadler, 1996).

What is an object? In CODE, an object is a perceptual group. Thus, an object is whatever falls within an above-threshold region of the CODE surface. CODE defines a hierarchy of objects in a principled fashion, by moving the threshold up and down the CODE surface. Low thresholds produce a small number of multi-element objects; high thresholds produce a larger number of single-element objects.

Theoretical integration. At this point, the theoretical integration should be clear: *The above-threshold region of CODE surface is BOTH an object and a spotlight.* CTVA selects objects and regions of space in the same act of attention. The difference between object-based and space-based attention is a matter of perspective. In CTVA, the two views are complementary rather than adversarial.

What determines the shape of the spotlight? The spotlight in CTVA is the above-threshold region of the CODE surface. The shape of the above-threshold region is determined jointly by the shape of the CODE surface and the threshold. The spotlight can have different shapes, depending on the threshold setting, but the shapes are constrained by the shape of the CODE surface, which depends deterministically on the proximity of the items in the display. CTVA does not banish omnipotent homunculi entirely because top-down processes determine the threshold setting, but it eliminates much of the work the homunculus had to do in space-based theories by constraining the shape of the spotlight to match the topography of the CODE surface.

How does selection occur within the focus of attention? Selection within the focus of attention occurs according to the principles of Bundesen's (1990) TVA model of selection. The person controls a bias parameter that makes a particular categorization more likely and a priority parameter that makes relevant objects more likely to be selected. In the counter-model version of the theory, the person also controls the response criteria that determine the number of counts required to categorize an object.

How does selection between objects occur? Selection between perceptual objects depends on top-down processes that apply conceptual representations of spatial relations to above-threshold regions of the CODE surface. The top-down processes include spatial indexing and reference frame alignment. The top-down processes are addressable by language, so that one person's utterances can control another person's attention (Logan, 1995). Selection between objects is the least well-specified part of the theory. Logan and Sadler (1996) sketched the computational requirements of the apprehension of spatial relations between objects, but they did not implement them at the same level of specificity as the other components of CTVA.

Benefits of CTVA

The marriage of CODE and TVA is beneficial in several respects. First and foremost, it provides quantitative accounts of seven important phenomena that have shaped the current literature on visual spatial attention. These accounts are unique because the accounts of competing theories are primarily qualitative. Moreover, CTVA provides some new insights into the phenomena that were not apparent in the qualitative accounts. The CTVA analysis of Prinzmetal's (1981) experiments on grouping effects on illusory conjunctions suggested that subjects grouped the displays only occasionally and most often treated the display items as separate objects. The analysis of Cohen and Ivry's (1989, 1991) experiments on distance effects in illusory conjunctions and conjunction search showed that a single mechanism could account for what appeared to be qualitatively different effects. The CTVA analysis of Banks and Prinzmetal's (1976) experiments showed that a serial search strategy, which they explicitly discounted, turned out to be necessary to account for the advantage of perceptually isolating the target. And the CTVA analysis of double and triple conjunction search suggested that the advantage of triple conjunctions may stem, in part at least, from attentive processes that compare display items with a description of the target, rather than the preattentive processes proposed by other theorists (Grossberg et al., 1994; Treisman & Sato, 1990; Wolfe et al., 1989).

The CTVA analyses were beneficial because they were among the first to provide a formal representation of space in the attention literature. Theories of visual attention agree that space is important and location information is special, but few say anything explicit about the representation of space and the properties of the representation (but see Ashby et al., 1996; Maddox et al., 1994). The CODE theory is important because it brings grouping by proximity back into the repertoire of object-based approaches to attention and provides a reasonable account of that grouping principle (Compton & Logan, 1993; van Oeffelen & Vos, 1982, 1983).

The CTVA analyses were also beneficial in that they showed the power of Bundesen's (1990) TVA model. Bundesen and colleagues applied the model primarily to partial and whole report tasks (Bundesen, 1987; Bundesen, Pedersen, & Larsen, 1984; Bundesen, Shibuya, & Larsen, 1985; Shibuya & Bundesen, 1988). Bundesen (1990) extended it to deal with other phenomena, and the current analysis extends it even further. The idea that selective attention can be based on a race between alternative candidates is exceptionally powerful and promising (e.g., Bundesen, 1993). It is especially interesting because the theory is tractable mathematically and assumes independence of processes. By contrast, many current formal models of attention, typically based on connectionist architectures, assume highly interactive processes and, consequently, must be analysed by simulation rather than simple mathematics (e.g., Cohen et al., 1990; Grossberg et al., 1994; Humphreys & Müller, 1993; Mozer, 1991; Phaf et al., 1990).

Limitations of CTVA

The CTVA model is limited in several respects. Some of the limitations point out important directions for future research, but some are stumbling blocks from which CTVA may never recover. Some of the limitations stem from the fact that CTVA is abstract. It says nothing about the nature of the features that comprise the feature distributions, and it says nothing about how similarity between perceptual objects and category templates is computed. It does not deal with other grouping principles, such as grouping by similarity, and it does not deal with motion. These limitations can be overcome by future research, and I will suggest possible solutions to some of these problems later.

More serious limitations stem from CODE's assumption that objects can be idealized as points in space (i.e., if the threshold is high enough). This assumption prevents CODE from dealing with objects that extend in space, with structured objects, and with interconnected or overlapping objects. This is an important limitation because many objects in the world have these properties. Many objects, such as the page you are reading, extend in space and cannot be easily idealized as points. Many objects are structured—things are built from interconnected parts (Biederman, 1987; Marr & Nishihara, 1978)—and the representations of structured objects cannot be idealized as simple points. Moreover, objects often overlap and occlude each other, and that is not easily captured in the pointilistic CODE representation.

In principle, it may be reasonable to idealize the locations of objects as points. That strategy is a common one in the linguistic and psycholinguistic literature on the apprehension of spatial relations (Herskovits, 1986; Jackendoff & Landau, 1991; Talmy, 1983). Even in that literature, however, some objects are idealized as lines, regions, and volumes, and that is hard to reconcile with the CODE idealization. Moreover, idealization of objects as points may be a more difficult problem for object recognition (identification) than for localization, for reasons described above.

It may be possible to deal with these problems by relaxing the assumption that objects are idealized as points, allowing objects

to occupy 1-D, 2-D, and 3-D regions in space. It may be possible to account for distance and grouping effects by assuming that the boundaries of objects vary in a manner similar to the variation in the pointilistic objects in CODE. The position of a line, for example, might vary according to a Laplace distribution in a direction orthogonal to its main axis. However, much of the elegance of CODE may be lost in the translation.

The difficulty with extended, structured, and overlapping objects is mitigated somewhat in the experimental paradigms that CTVA and the other theories of visual spatial attention address. Most experiments on visual search, partial report, and so on, present subjects with separate objects with a simple structure, and the CODE representation may be well suited for those displays (but see Wolfe, 1996). It may not be unreasonable to idealize a display of randomly positioned letters as a set of unconnected points. Thus, CTVA is a reasonable model of current research in visual spatial attention.

Another difficulty with CTVA is that it defines objects only in terms of location. Proximity is the only grouping principle that determines what an object is. While many researchers would agree that location is an important defining characteristic of an object, most would argue that it is not the only one. Grouping by similarity (Baylis & Driver, 1992), common fate (Driver & Baylis, 1989), and connectedness (Kramer & Jacobson, 1991) have been shown to produce strong object-based effects independent of proximity. It may be possible to incorporate the effects of grouping by similarity and common fate into CTVA (see below), but connectedness may be difficult because it implies a hierarchical structure that is not captured in CODE's idealization of objects as points (see Palmer & Rock, 1994).

Capacity Limitations and the Locus of Selection

The locus of selection and the nature of capacity limitations are longstanding issues in the attention literature, occupying psychologists since the time of Broadbent (1958) if not earlier. The typical theory of attention includes early preattentive processes that are unlimited in capacity followed by attentive processes that are limited in capacity. Controversy surrounds the locus of the boundary between preattentive and attentive processes and the involvement of capacity limitations in each stage of processing. The CODE theory of visual attention takes a position on these issues, departing somewhat from the typical view.

Locus of selection. The locus of selection issue has been articulated in at least two ways in the literature: One concerns *attended items*, addressing the kind of information on which attentional selection is based. Advocates of early selection argue that items are selected on the basis of physical features, like location and color (van der Heijden, 1992), while advocates of late selection argue that attentional selection is based on identity, meaning, or category membership (e.g., Shiffrin & Schneider, 1977). The second way of articulating the issue concerns *unattended items*, addressing the level of processing attained by stimuli that attention does not select. Advocates of early selection argue that unattended stimuli receive only cursory analysis of physical features (Broadbent, 1958) and advocates of late selection argue that unattended stimuli are fully processed, to

the level of identification (Deutsch & Deutsch, 1963). The two approaches appear similar, and many consider them equivalent. However, they really address different issues. As van der Heijden (1992) points out, all stimuli could be processed fully, to the level of identification, but selection could be based on location nevertheless. Late selection in one sense could be paired with early selection in the other sense.

CTVA is both an early selection and a late selection theory with respect to attended items. It is an early selection theory from the perspective of CODE, because between-object selection is based on location, and selection by location is traditionally associated with early selection. However, it is a late selection theory from the perspective of TVA, because TVA selects items by categorizing them (Bundesen, 1990). Items race to be categorized, and the first one (or the first K) to finish is (are) selected.

CTVA is an early selection theory with respect to unattended items, because unattended items are not categorized. Categorization occurs when an item wins the race (or when K items finish). Unselected items lose the race and therefore are not categorized. They receive only cursory processing.

Note that CTVA does not accept the common assumption of a chain of increasingly abstract processes going from stimulus to response, beginning with low-level representations and proceeding to identity, categorization, and meaning. Like TVA, CTVA assumes only two levels of representation, precategorical and categorical. The precategorical representation consists of the feature distributions and the CODE surface; the categorical representation consists of categorizations of display items. In principle, the same kinds of information exist in both representations. The precategorical representation contains the perceptual information that supports categorization, and the categorical representation contains categorizations of perceptual information. Abstract categories, like *mammal*, are defined in terms of perceptual features in the precategorical representation, just as concrete categories, such as *red*, are (for further discussion, see Logan, 1995).

Note as well that CTVA does not assume that all possible categorizations of the display can be processed in parallel over the whole display. Some categorizations, such as deciding whether a display instantiates a categorical spatial relation like *above* or *beside* require more than the TVA part of CTVA. Logan (1994, 1995) argued that apprehension of spatial relations requires integrating information from several attentional fixations, whereas TVA describes what happens in a single fixation. Apprehension of spatial relations requires the underspecified late location part of CTVA depicted in Figures 1 and 6. Other categorizations that require more than one fixation of attention likely cannot be done by the TVA part of CTVA. It is not immediately clear what kinds of categorization can and cannot be done by TVA. Future research and further specification of the TVA and late location parts of CTVA will be required before an answer emerges.

Capacity limitations. Theories of attention assume that the capacity for processing information is unlimited, limited, or fixed. According to Townsend and Ashby (1983), capacity is unlimited if the rate at which one item is processed does not depend on the number of items being processed simultaneously. Capacity is limited if the rate at which an item is processed

depends on the number of other items being processed. Capacity is fixed if it is limited, and the limit is constant across displays, tasks, and situations. CTVA assumes that capacity is limited.

According to Bundesen (1990), the processing capacity, C , of TVA and CTVA can be defined as the sum of all of the $v(x, i)$ values across all perceptual categorizations of all elements in the visual field, that is

$$C = \sum_{x \in S} \sum_{i \in R} v(x, i).$$

According to this definition, capacity is unlimited if the $v(x, i)$ values do not change when a new item is added to the display; that is, C increases by $\sum_{i \in R} v(x, i)$ when a new item is added; capacity is limited if the $v(x, i)$ values decrease when a new item is added to the display; C increases by an amount less than $\sum_{i \in R} v(x, i)$ when a new item is added; and capacity is limited and fixed if the $v(x, i)$ values decrease so that C stays constant.

According to this definition, CTVA and TVA are limited-capacity models. This follows from the definition of $v(x, i)$ in Equations 5 and 12. The value of $v(x, i)$ is the product of $\eta(x, i)$, β_i , and the normalized attentional weight, $w_x / \sum w_z$. As new items are added to the display, the attentional weight on item x decreases (see Equation 6) and, consequently, $v(x, i)$ decreases. Bundesen (1990) argued that if the items in the display were homogeneous, that is, if $\sum_{i \in R} v(x, i)\beta_i$ was constant for all items in the display, that capacity, C , would be fixed as well as limited.¹¹ In many applications of CTVA, the homogeneity assumption will be violated because $v(x, i)$ depends on the feature catch, c_x , (see Equation 12) and the feature catch will be different for different items in the display (i.e., whenever items are unevenly spaced). Thus, CTVA assumes limited capacity but usually not fixed capacity.

In some applications, CTVA does not use attention weights to select targets to process in the same way that TVA does. In the fits to the Eriksen and Eriksen (1974) data, the attention weights, w_x , were set to 1 and the central target item was selected by the late location system outside of CODE and TVA (i.e., using Logan's 1995 theory). In those applications, the $v(x, i)$ values are not affected by adding other items to the display, so CTVA assumes unlimited processing capacity.

Note that processing capacity is not the same theoretical construct as processing resources. Processing capacity plays a role in resource theories, but it is only one of several constructs at work in those theories. Most resource theories make the strong assumption that processing capacity is both limited and fixed across displays, tasks, and situations (i.e., C is constant in all contexts), and neither TVA nor CTVA make that assumption.

¹¹ The fixed-capacity version of TVA does not assume that capacity is fixed at the same value for all displays, tasks, and situations. The same experimental procedure can be complicated in a way that violates the homogeneity assumption (e.g., by crowding so many items in a display that lateral masking is produced), and the factors that limit capacity in one situation may not be the ones that limit it in another (e.g., capacity may be limited by display contrast in one situation and by item similarity in another). Thus, the TVA idea of fixed capacity is quite different from the resource-theory idea of fixed capacity.

Resource theories go beyond the idea of processing capacity, making additional assumptions about how capacity can be allocated. They argue that resources can be allocated in parallel rather than in series and the allocation is graded rather than all or none. Moreover, they assume that performance changes in a continuously graded fashion as resource allocation varies (see e.g., Kahneman, 1973; Navon & Gopher, 1979; Norman & Bobrow, 1975). None of these ideas is entailed by the concept of processing capacity. TVA and CTVA are largely mute on the issue of resources and therefore immune to the criticisms of resource theory (e.g., Allport, 1980; Duncan, 1980; Logan, in press; Navon, 1984; Neisser, 1976).

Future Directions

A theory as broad as CTVA is a fertile ground for future research. By incorporating Bundesen's (1990) TVA, CTVA inherits the phenomena that TVA accounts for, and TVA is already a far-reaching theory. One important direction for future research is to look inward and test the assumptions underlying both TVA and CTVA. The distributional assumptions are important to test because they were made largely for mathematical convenience. Other distributions may do as well as or better than the exponential and the Laplace (see e.g., Ashby et al., 1996; Compton & Logan, 1993; Maddox et al., 1994). Other directions for future research are more outward-looking, trying to extend the theory to new domains. In the remainder of the article, I will describe three that are high on my agenda.

Proximity and grouping effects in partial report. Several investigators have found that performance in partial report tasks is influenced by perceptual grouping and by the presence of nearby distractors. Fryklund (1975) showed that subjects do better if the items they are supposed to report are adjacent to each other in coherent groups. Merikle (1980) found something similar, showing that partial report performance was better when the to-be-reported subset was compatible with the Gestalt grouping of the display than when it was incompatible. It should be possible to account for these results with CTVA, using versions of TVA that Bundesen and colleagues developed for partial report tasks (e.g., Bundesen, 1987; Bundesen et al., 1984; Bundesen et al., 1985; Shibuya & Bundesen, 1988).

The key to fitting these data may lie in a proximity effect reported by Snyder (1972) and Mewhort, Campbell, Marchetti, and Campbell (1981). In partial report tasks that probe for a single item rather than a set of items, errors are often correct reports of the letters adjacent to the target item. The CODE theory of visual attention would explain this result in terms of the feature catch. Items adjacent to the target are likely to intrude in the feature catch for the target because significant parts of their feature distributions are likely to fall in the above-threshold region centered on the target. Adjacent items are more likely than nonadjacent items to intrude in the target's feature catch because the feature distribution falls off exponentially as distance increases. Thus, in principle, CTVA can account for the Snyder (1972) and Mewhort et al. (1981) results. The question is whether it can account for them quantitatively, using reasonable parameter values.

The same idea can be extended to account for the grouping

effects reported by Fryklund (1975) and Merikle (1980): Items close to each other or in the same perceptual group are likely to intrude in each other's feature catch. If the task requires identification of adjacent items or items in the same group, these intrusions might be beneficial, perhaps priming responses appropriate for other to-be-reported items. However, if the task requires identification of nonadjacent items or items in different perceptual groups, then intrusions from adjacent items and items in the same perceptual group might be harmful, priming inappropriate responses that compete with the required responses to to-be-reported items. To test this idea, CTVA would have to be extended to include the TVA account of multi-item partial report performance (e.g., Bundesen, 1987; Bundesen et al., 1984; Bundesen et al., 1985; Shibuya & Bundesen, 1988) and, possibly, to include priming of not-yet-reported items.¹²

Grouping by similarity. CODE and CTVA deal only with grouping by proximity, yet many other factors affect perceptual grouping and grouping by those factors affects performance in attention tasks. An important direction for future research is to extend CODE and CTVA to deal with other grouping principles. Grouping by similarity is a good candidate for the first step in that direction because it is well studied perceptually (e.g., Beck, Prazdny, & Rosenfeld, 1983; Bergen, 1991) and it has powerful effects on attention (e.g., Baylis & Driver, 1992; Duncan & Humphreys, 1989; Harms & Bundesen, 1983; Humphreys & Müller, 1993; Ivry & Prinzmetal, 1991; Wolfe, 1994).

The mechanisms for dealing with similarity effects may already be present in CTVA. The similarity parameters in TVA may interact with CODE to limit access to the attentional system to items that share common characteristics. Manipulating β_i increases access for items similar to category i , that is, with high $\eta(x, i)$ values, and decreases access for items dissimilar to category i , (that is, with low $\eta(x, i)$ values). If the items in the display are dissimilar—if the distribution of $\eta(x, i)$ values is distinctly bimodal with some very high and some very low values—then manipulating β should “parse” the display into two groups—one with high $\eta(x, i)$ values and one with low $\eta(x, i)$ values. However, if the items in the display are similar—if the distribution of $\eta(x, i)$ values is unimodal and compact—then manipulating β should not separate the items.

The effects of manipulating β and η on the feature distributions can be seen in Figure 17. Figure 17 represents feature distributions and the CODE surface for displays like OXXOXO. Similarity between the Xs and Os decreases going from the top left to the bottom right, and the effective feature distributions for the Os decrease in area as similarity decreases. The feature distributions for the Os were multiplied by the product of β_i and $\eta(x, i)$, which reduces the area under each feature distribution. Feature distributions with low values of $\eta(x, i)$ are suppressed, whereas feature distributions with high values of $\eta(x, i)$ maintain their salience. The CODE surface, however, is built before β has its effect in the current version of CTVA, so the CODE surface remains the same as similarity varies.

The proposed modification of CTVA is presented in Figure 18. In the modification, manipulations of β feed back to the

¹² Since this article went to press, Logan and Bundesen (1996) applied CTVA to these partial report tasks with considerable success.

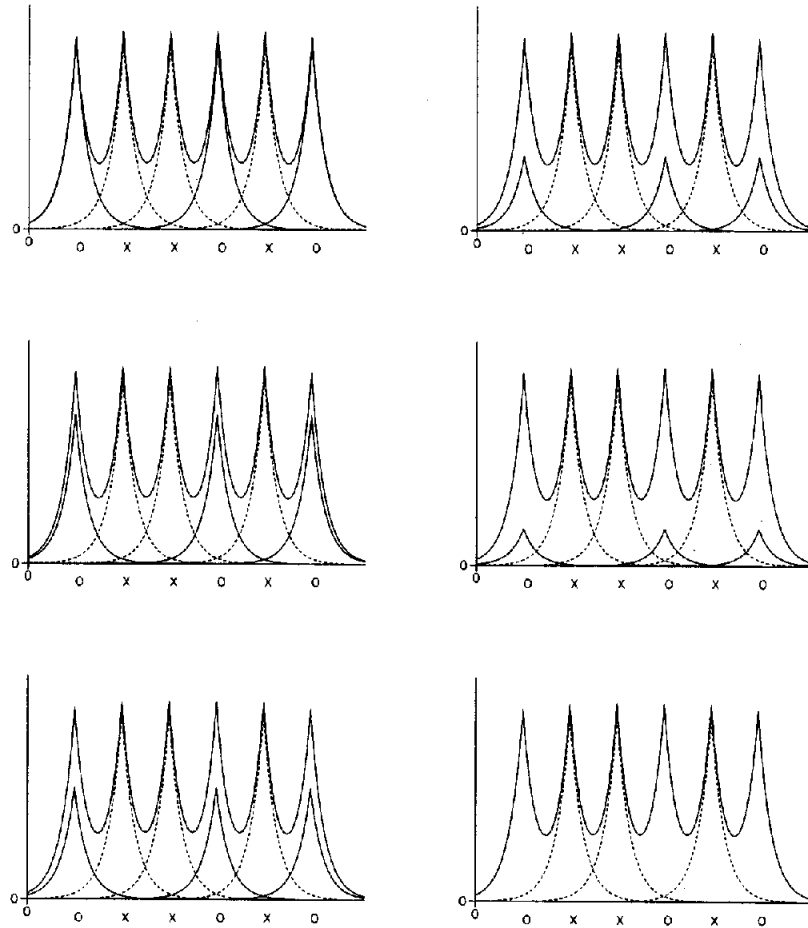


Figure 17. Feature distributions and original CODE surfaces for items that differ in similarity (X and O), with $\eta(x, i)\beta_i = 1.0$ for the Xs but varying between 1.0 and 0.0 for the Os (1.0, 0.8, and 0.6 from top to bottom in the left-hand panels; 0.4, 0.2, and 0.0 from top to bottom in the right hand panels).

CODE surface so that the CODE surface changes as similarity between the Xs and the Os decreases. Figure 18 plots *effective CODE surfaces* that were produced by multiplying all of the feature distributions in the display by the product of β_i and $\eta(x, i)$ and then summing the feature distributions. The multiplication changes the shape of the CODE surface so that items with low values of $\eta(x, i)$ are suppressed, whereas feature distributions with high values of $\eta(x, i)$ remain prominent.

It remains to be seen whether this modification of CTVA can account for similarity effects in grouping and attention experiments. The idea can be tested quite stringently by requiring the model to account for both grouping judgments and effects on performance in attention experiments with the same parameter values. That test, however, is beyond the scope of this article.

One limitation of the proposed approach is that it depends on top-down specification of β_i , which requires foreknowledge of the categorical difference between the groups to be segregated. Textbook demonstrations of grouping by similarity do not (seem to) require foreknowledge of the dimension that distinguishes the groups. Moreover, visual search for singleton targets—items that differ in some unforeseen property from the

distractors—is almost as easy as search for predesignated targets (Müller, Heller, & Zeigler, 1995; Treisman, 1988). It is possible that there is some interaction between bottom-up and top-down processes that allow the system to set the appropriate β values to achieve segregation, but there is not much time for those interactions to take place because similarity grouping effects are apparent very quickly (Beck et al., 1983) and singleton popout is very fast (i.e., cost of not knowing the target dimension is small; Müller et al., 1995; Treisman, 1988). Perhaps CTVA will have to be supplemented by some other mechanism that segregates dissimilar items and isolates dissimilar targets (cf. Cave & Wolfe, 1990; Humphreys & Müller, 1993).

Attention and automaticity. Theories of visual spatial attention are intended to interface with theories of other aspects of cognition, such as memory retrieval, but they rarely do. Consequently, theories of visual spatial attention are largely ahistorical, capturing a moment in a person's life without describing how the knowledge that is necessary to support current performance was acquired. Similarly, theories of other aspects of cognition rarely say anything about visual spatial attention and the perceptual processes that allow them to interface with the ex-

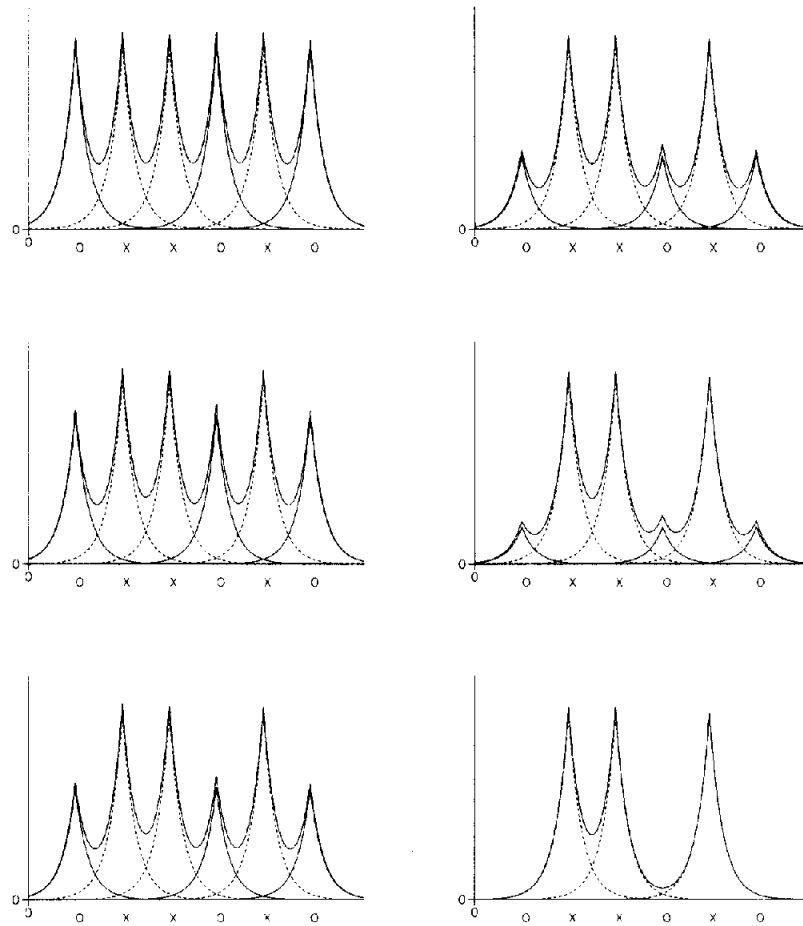


Figure 18. Feature distributions and modified CODE surfaces for items that differ in similarity (X and O), with $\eta(x, i)\beta_i = 1.0$ for the Xs but varying between 1.0 and 0.0 for the Os (1.0, 0.8, and 0.6 from top to bottom in the left-hand panels; 0.4, 0.2, and 0.0 from top to bottom in the right-hand panels).

ternal world. An important goal for future research is to integrate CTVA with other theories of cognition, especially theories that describe learning.

I am particularly interested in interfacing CTVA with the *instance theory of automaticity* (Logan, 1988, 1992) and a recent generalization of the theory by Nosofsky and Palmeri (in press) called the *exemplar-based random walk* (EBRW) model. The instance theory and EBRW describe the acquisition and expression of automaticity in a manner that relates it to theories of memory (Hintzman, 1988; Jacoby & Brooks, 1984), concept learning (Hintzman, 1986; Medin & Schaffer, 1978; Nosofsky, 1988), problem solving (Ross, 1984, 1987), judgment (Kahneman & Miller, 1986), and social categorization (Smith & Zaraté, 1992). So interfacing CTVA with instance theory should go a long way toward a general account of cognition.

The instance theory is an excellent candidate for interfacing with CTVA because they are both race models. The instance theory describes automaticity as performance based on retrieval of past solutions from memory, and during retrieval, the different traces of past solutions in memory (the instances) race against each other, with the first trace to finish determining performance

(Logan, 1988, 1992). Until now, the instance theory has assumed a binary similarity gradient, with traces either identical to each other or completely different, and it has assumed that the retrieval time distribution was the same for each trace. These assumptions were made largely for mathematical convenience, in order to support proofs that mean reaction time and the entire distribution of reaction times would decrease as a power function of practice (Logan, 1992). The EBRW model is an improvement over the instance theory because it assumes that similarity is graded continuously and retrieval time varies as a function of similarity. Moreover, EBRW generalizes the idea of a simple race, in which the first instance retrieved is the winner, to a relay race, in which several instances are retrieved before the process terminates. The idea is similar to the counter-model generalization of CTVA in Equations 9 and 10.

The integration of CTVA and the instance-EBRW theory would interpret each instance as an $\eta(x, i)$ parameter, with a retrieval time that depends on $v(x, i)$. As in the original instance theory, different traces of the same stimulus would have distinct but identical $\eta(x, i)$ (and $v(x, i)$) values, so that retrieval time would depend on the number of instances in mem-

ory as well as their similarity to the current object of attention. The novel contribution (from EBRW) is to allow nonidentical traces to enter the race with retrieval times that are functions of their similarity to the current object of attention. Moreover, the attentional mechanisms in CTVA would allow a principled account of the effects of attention in the acquisition and expression of automaticity, which is a central topic in recent investigations of the instance theory (Logan & Etherton, 1994; Logan, Taylor, & Etherton, 1996). Of course, the proof will be in the pudding. It remains to be seen whether these speculations can provide reasonable accounts of the attention and learning phenomena associated with automaticity.

Conclusions

The combination of CODE and TVA accounted for many phenomena in the literature on visual spatial attention. The CTVA model provided coherent answers to the five questions that challenge current theories of attention. It integrated object-based and space-based approaches to attention, arguing that the output of CODE, which TVA selects, is both an object and a region of space. The major contribution of CTVA was to provide coherent accounts of seven major empirical phenomena that shaped the current literature on visual spatial attention. This was an important contribution because the CTVA accounts were quantitative, whereas previous accounts were only qualitative.

The strengths of CTVA derive equally from the representational assumptions of the CODE theory and the processing assumptions of the TVA theory. By itself, CODE addresses only the phenomenology of grouping by proximity; combined with TVA, it addresses attention. By itself, TVA underestimates the importance of space and cannot account for the effects of distance and grouping by proximity; combined with CODE it provides a more complete and more balanced account of attentional phenomena. The CTVA model is strong primarily because it was built from strong components; CODE and especially TVA were impressive theories to begin with. Perhaps the most important contribution of CTVA is to show that strong theories can be made even stronger by combining them with other theories and that, ultimately, psychology can progress by developing theories cumulatively (Posner, 1982).

References

- Allport, D. A. (1980). Attention and performance. In G. Claxton (Ed.), *Cognitive psychology* (pp. 112-153). London: Routledge & Kegan Paul.
- Andersen, G. J. (1990). Focused attention in three-dimensional space. *Perception and Psychophysics*, 47, 112-120.
- 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.
- Attneave, F. (1960). In defense of homunculi. In W. A. Rosenblith (Ed.), *Sensory communication* (pp. 777-782). Cambridge, MA: MIT Press.
- Banks, W. P., & Prinzmetal, W. (1976). Configurational effects in visual information processing. *Perception and Psychophysics*, 19, 361-367.
- Baylis, G. C., & Driver, J. (1992). Visual parsing and response competition: The effects of grouping. *Perception and Psychophysics*, 51, 145-162.
- Baylis, G. C., & Driver, J. (1993). Visual attention and objects: Evidence for hierarchical coding of location. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 451-470.
- Beck, J., Prazdny, K., & Rosenfeld, A. (1983). A theory of textural segmentation. In J. Beck, B. Hope, & A. Rosenfeld (Eds.), *Human and machine vision*. New York: Academic Press.
- Bergen, J. R. (1991). Theories of visual texture perception. In D. M. Regan (Ed.), *Spatial vision* (pp. 114-134). New York: Macmillan.
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94, 65-96.
- Broadbent, D. E. (1958). *Perception and communication*. London: Pergamon Press.
- Broadbent, D. E. (1971). *Decision and stress*. London: Academic Press.
- 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., Pedersen, L. F., & Larsen, A. (1984). Measuring efficiency of selection from briefly exposed visual 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.
- Cave, K. R., & Wolfe, J. M. (1990). Modeling the role of parallel processing in visual search. *Cognitive Psychology*, 22, 225-271.
- Chastain, G. (1982). Feature mislocalizations and misjudgments of intercharacter distance. *Psychological Research*, 44, 51-66.
- Clark, H. H. (1973). Space, time, semantics, and the child. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 27-63). New York: Academic Press.
- Cohen, A., & Ivry, R. (1989). Illusory conjunctions inside and outside the focus of attention. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 650-663.
- Cohen, A., & Ivry, R. (1991). Density effects in conjunction search: Evidence for a coarse location mechanism of feature integration. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 891-901.
- 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.
- Coles, M. G. H., Gratton, G., Bashore, T. R., Eriksen, C. W., & Donchin, E. (1985). A psychophysiological investigation of the continuous flow model of human information processing. *Journal of Experimental Psychology: Human Perception and Performance*, 11, 529-553.
- Compton, B. J., & Logan, G. D. (1993). Evaluating a computational model of perceptual grouping by proximity. *Perception and Psychophysics*, 53, 403-421.
- Compton, B. J., & Logan, G. D. (1996). *Judgments of perceptual groups: Reliability and sensitivity to stimulus transformation*. Unpublished manuscript.
- Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, 70, 80-90.
- Driver, J., & Baylis, G. C. (1989). Movement and visual attention: The spotlight metaphor breaks down. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 448-456.

- Duncan, J. (1980). The demonstration of capacity limitation. *Cognitive Psychology*, 12, 75-96.
- Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, 113, 501-517.
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, 96, 433-458.
- Egley, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General*, 123, 161-177.
- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception and Psychophysics*, 16, 143-149.
- Eriksen, C. W., & Hoffman, J. E. (1972). Temporal and spatial characteristics of selective encoding from visual displays. *Perception and Psychophysics*, 12, 201-204.
- Eriksen, C. W., & Hoffman, J. E. (1973). The extent of processing of noise elements during selective encoding from visual displays. *Perception and Psychophysics*, 14, 155-160.
- Eriksen, C. W., Pan, K., & Botella, J. (1994). Attentional distribution in visual space. *Psychological Research*, 56, 5-13.
- Eriksen, C. W., & Schultz, D. W. (1979). Information processing in visual search: A continuous flow conception and experimental results. *Perception and Psychophysics*, 25, 249-263.
- Eriksen, C. W., & St. James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception and Psychophysics*, 40, 225-240.
- Eriksen, C. W., & Yeh, Y. Y. (1985). Allocation of attention in the visual field. *Journal of Experimental Psychology: Human Perception and Performance*, 11, 583-597.
- Flowers, J. H., & Wilcox, N. (1982). The effect of flanking context on visual classification: The joint contribution of interactions at different processing levels. *Perception and Psychophysics*, 32, 581-591.
- Fryklund, I. (1975). Effects of cued-set spatial arrangement and target-background similarity in the partial-report paradigm. *Perception and Psychophysics*, 17, 375-386.
- Grossberg, S., Mingolla, E., & Ross, W. D. (1994). A neural theory of attentive visual search: Interactions of boundary, surface, spatial, and object representations. *Psychological Review*, 101, 470-489.
- Harms, L., & Bundesen, C. (1983). Color segregation and selective attention in a nonsearch task. *Perception and Psychophysics*, 33, 11-19.
- Herskovits, A. (1986). *Language and spatial cognition: An interdisciplinary study of the prepositions in English*. Cambridge, England: Cambridge University Press.
- Hillstrom, A. P., & Logan, G. D. (in press). Process dissociation, cognitive architecture, and response time: Comments on Lindsay and Jacoby (1994). *Journal of Experimental Psychology: Human Perception and Performance*.
- 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 memory in a multiple-trace memory model. *Psychological Review*, 95, 528-551.
- Humphreys, G. W., & Müller, H. J. (1993). SEArch via Recursive Rejection (SERR): A connectionist model of visual search. *Cognitive Psychology*, 25, 43-110.
- Ivry, R., & Prinzmetal, W. (1991). Effect of feature similarity on illusory conjunctions. *Perception and Psychophysics*, 49, 105-116.
- Jackendoff, R., & Landau, B. (1991). Spatial language and spatial cognition. In D. J. Napoli & J. A. Kegl (Eds.), *Bridges between psychology and linguistics: A Swarthmore festschrift for Lila Gleitman*. Hillsdale, NJ: Erlbaum.
- 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* (pp. 1-47). San Diego, CA: Academic Press.
- Juola, J. F., Bouwhuis, D. G., Cooper, E. E., & Warner, C. B. (1991). Control of attention around the fovea. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 125-141.
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.
- Kahneman, D., & Henik, A. (1977). Effects of visual grouping on immediate recall and selective attention. In S. Dornic (Ed.), *Attention and Performance VI* (pp. 307-332). Hillsdale, NJ: Erlbaum.
- Kahneman, D., & Henik, A. (1981). Perceptual organization and attention. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 181-211). Hillsdale, NJ: Erlbaum.
- Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, 93, 136-153.
- 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.
- Kahneman, D., Treisman, A., & Gibbs, B. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, 24, 175-219.
- Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, 4, 219-227.
- Kramer, A. F., & Jacobson, A. (1991). Perceptual organization and focused attention: The role of objects and proximity in visual processing. *Perception and Psychophysics*, 50, 267-284.
- LaBerge, D., & Brown, V. (1989). Theory of attentional operations in shape identification. *Psychological Review*, 96, 101-124.
- Lasaga, M. I., & Hecht, H. (1991). Integration of local features as a function of global goodness and spacing. *Perception and Psychophysics*, 49, 201-211.
- Logan, G. D. (1980). Attention and automaticity in Stroop and priming tasks: Theory and data. *Cognitive Psychology*, 12, 523-553.
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, 95, 492-527.
- 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. (1994). Spatial attention and the apprehension of spatial relations. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 1015-1036.
- Logan, G. D. (1995). Linguistic and conceptual control of visual spatial attention. *Cognitive Psychology*, 28, 103-174.
- Logan, G. D. (in press). The automaticity of academic life: Unconscious applications of an implicit theory. In R. S. Wyer (Ed.), *Advances in social cognition* (Vol. 10). Mahwah, NJ: Erlbaum.
- 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). San Diego, CA: Academic Press.
- Logan, G. D., & Etherton, J. L. (1994). What is learned during automatization? The role of attention in constructing an instance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 1022-1050.
- Logan, G. D., & Sadler, D. (1996). A computational analysis of the apprehension of spatial relations. In P. Bloom, M. Peterson, L. Nadel, & M. Garrett (Eds.), *Language and space* (pp. 493-529). Cambridge, MA: MIT Press.
- 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.
- Luce, R. D. (1959). *Individual choice behavior*. New York: Wiley.
- Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, 109, 163-203.
- Maddox, W. T., Prinzmetal, W., Ivry, R., & Ashby, F. G. (1994). A probabilistic multidimensional model of location information. *Psychological Research*, 56, 66-77.
- Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London*, 200, 269-294.
- Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, 85, 207-238.
- Merkle, P. M. (1980). Selection from visible persistence by perceptual groups and category membership. *Journal of Experimental Psychology: General*, 109, 279-295.
- Mewhort, D. J. K., Campbell, A. J., Marchetti, F. M., & Campbell, J. I. D. (1981). Identification, localization and "iconic memory": An evaluation of the bar probe task. *Memory and Cognition*, 9, 50-67.
- Milner, P. M. (1974). A model for visual shape recognition. *Psychological Review*, 81, 521-535.
- Mozer, M. C. (1991). *The perception of multiple objects: A connectionist approach*. Cambridge, MA: MIT Press.
- Müller, H. J., Heller, D., & Ziegler, J. (1995). Visual search for singleton targets within and across feature dimensions. *Perception and Psychophysics*, 57, 1-17.
- Navon, D. (1977). Forest before trees? The precedence of global features in visual perception. *Cognitive Psychology*, 9, 353-383.
- Navon, D. (1984). Resources—A theoretical soup stone? *Psychological Review*, 91, 216-234.
- Navon, D., & Gopher, D. (1979). On the economy of the human processing system. *Psychological Review*, 86, 214-255.
- Neisser, U. (1976). *Cognition and reality*. San Francisco: Freeman.
- Norman, D. A., & Bobrow, D. (1975). On data-limited and resource-limited processes. *Cognitive Psychology*, 7, 44-64.
- Nosofsky, R. M. (1988). Exemplar-based accounts of realtions between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 700-708.
- Nosofsky, R. M., & Palmeri, T. J. (in press). *An exemplar-based random walk model of speeded classification*. *Psychological Review*.
- Palmer, S. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology*, 9, 441-474.
- Palmer, S., & Rock, I. (1994). Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic Bulletin and Review*, 1, 29-55.
- Pashler, H. (1987). Detecting conjunctions of color and form: Reassessing the serial search hypothesis. *Perception and Psychophysics*, 41, 191-201.
- Phaf, R. H., van der Heijden, A. H. C., & Hudson, P. T. W. (1990). SLAM: A connectionist model for attention in visual selection tasks. *Cognitive Psychology*, 22, 273-341.
- Pinker, S. (1984). Visual cognition: An introduction. *Cognition*, 18, 1-63.
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32, 3-25.
- Posner, M. I. (1982). Cumulative development of attentional theory. *American Psychologist*, 37, 168-179.
- Posner, M. I., & Boies, S. J. (1971). Components of attention. *Psychological Review*, 78, 391-408.
- 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.
- Prinzmetal, W. (1981). Principles of feature integration in visual perception. *Perception and Psychophysics*, 30, 330-340.
- Prinzmetal, W., & Keysar, B. (1989). Functional theory of illusory conjunctions and neon colors. *Journal of Experimental Psychology: General*, 118, 165-190.
- Prinzmetal, W., & Mills-Wright, M. (1984). Cognitive and linguistic factors affect visual feature integration. *Cognitive Psychology*, 16, 305-340.
- Prinzmetal, W., Treiman, R., & Rho, S. H. (1986). How to see a reading unit. *Journal of Memory and Language*, 25, 461-475.
- Polyshyn, Z. (1984). *Computation and cognition*. Cambridge, MA: MIT Press.
- Polyshyn, Z. (1989). The role of location indices in spatial perception: A sketch of the FINST spatial-index model. *Cognition*, 32, 65-97.
- Polyshyn, Z., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, 3, 179-197.
- Ratcliff, R. (1981). Theory of order relations in perceptual matching. *Psychological Review*, 88, 552-572.
- Ross, B. H. (1984). Reminders and their effects in learning a cognitive skill. *Cognitive Psychology*, 16, 371-416.
- Ross, B. H. (1987). This is like that: The use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 629-639.
- Shibuya, H., & Bundesen, C. (1988). Visual selection from multielement displays: Measuring and modeling effects of exposure duration. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 591-600.
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, 84, 127-190.
- Shulman, G. L., Remington, R., & McLean, J. P. (1979). Moving attention through space. *Journal of Experimental Psychology: Human Perception and Performance*, 5, 522-526.
- Smith, E. R., & Zarate, M. A. (1992). Exemplar-based model of social judgment. *Psychological Review*, 99, 3-21.
- Snyder, C. R. R. (1972). Selection, inspection, and naming in visual search. *Journal of Experimental Psychology*, 92, 428-431.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643-662.
- Talmy, L. (1983). How language structures space. In H. L. Pick & L. P. Acredolo (Eds.), *Spatial orientation: Theory, research, and application* (pp. 225-282). New York: Plenum.
- Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes*. Cambridge, England: 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. (1990). Variations on the theme of feature integration: Reply to Navon (1990). *Psychological Review*, 97, 460-463.
- 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., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 459-478.
- Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, 14, 107-141.
- Trick, L. M., & Polyshyn, Z. W. (1994). Why are small and large num-

- bers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review*, 101, 80–102.
- Tsal, Y. (1983). Movements of attention across the visual field. *Journal of Experimental Psychology: Human Perception and Performance*, 9, 523–530.
- Ullman, S. (1984). Visual routines. *Cognition*, 18, 97–159.
- Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. W. J. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT Press.
- Vandalloise, C. (1991). *Spatial prepositions: A case study from French*. Chicago: University of Chicago Press.
- 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 and 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.
- Vecera, S. P., & Farah, M. J. (1994). Does visual attention select objects or locations? *Journal of Experimental Psychology: General*, 123, 146–160.
- Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin and Review*, 1, 202–238.
- Wolfe, J. M. (1996). Extending guided search: Why guided search needs a preattentive “item map.” In A. F. Kramer, M. G. H. Coles, & G. D. Logan (Eds.), *Converging operations in the study of visual selective attention* (pp. 247–270). Washington DC: American Psychological Association.
- 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.
- Wolford, G. (1975). Perturbation model for letter identification. *Psychological Review*, 82, 184–199.
- Wolford, G., & Shum, K. H. (1980). Evidence for feature perturbations. *Perception and Psychophysics*, 27, 409–420.
- Yantis, S. (1992). Multielement visual tracking: Attention and perceptual organization. *Cognitive Psychology*, 24, 295–340.

Appendix A

The Mathematical Basis of TVA

Bundesen's (1990) TVA depends on the exponential distribution to predict reaction time and accuracy. The purpose of this appendix is to provide the derivations and explain the reasoning underlying them.

Exponential Distribution and $v(x, i)$

Bundesen (1990) interprets the $v(x, i)$ values as rate parameters in exponential distributions. The exponential distribution's density function is

$$f(t) = \lambda \exp[-\lambda t]. \quad (\text{A1})$$

Its cumulative distribution function is

$$F(t) = \int_0^t f(t) dt = 1 - \exp[-\lambda t]. \quad (\text{A2})$$

The relation between $v(x, i)$ and the exponential distribution depends on the hazard function, $h(t)$. Bundesen (1990) assumed that the $v(x, i)$ values were hazard functions for exponential distributions. The general expression for the hazard function is

$$h(t) = \frac{f(t)}{1 - F(t)}. \quad (\text{A3})$$

The hazard function for an exponential distribution is 0 for time < 0 and constant over time > 0 , as can be verified by inserting Equations A1 and A2 into A3. The value of the constant is λ , the rate parameter for the exponential distribution.

The hazard function is useful for many reasons (see Luce, 1986; Townsend & Ashby, 1983). For our purposes, the hazard function is useful because it leads directly to the distribution function:

$$F(t) = 1 - \exp \left[- \int_{-\infty}^t h(x) dx \right]. \quad (\text{A4})$$

Substituting the constant hazard function for the exponential distribution, $h(x) = \lambda$, into Equation A4 yields:

$$F(t) = 1 - \exp[-\lambda t], \quad (\text{A5})$$

which is the same as Equation A2. Thus, in Bundesen's (1990) theory, the $v(x, i)$ values are directly interpretable as rate parameters— λ values—for exponential distributions. The exponential distributions are important because they yield estimates of reaction time immediately: The mean and standard deviation of the exponential and, hence, reaction time, are both $1/\lambda$.

Reaction Time

Exponential distributions behave nicely when they race against each other. The density function, $f_{\min}(t)$, for the minima of two distributions, $f_1(t)$ and $f_2(t)$, is

$$f_{\min}(t) = f_1(t)[1 - F_2(t)] + f_2(t)[1 - F_1(t)]. \quad (\text{A6})$$

If the distributions are exponential, then A6 becomes

$$\begin{aligned} f_{\min}(t) &= \lambda_1 \exp[-\lambda_1 t] \exp[-\lambda_2 t] + \lambda_2 \exp[-\lambda_2 t] \exp[-\lambda_1 t] \\ &= (\lambda_1 + \lambda_2) \exp[-(\lambda_1 + \lambda_2)t]. \end{aligned} \quad (\text{A7})$$

Thus, the distribution of minima sampled from two exponential distributions is itself an exponential distribution with a rate parameter equal to the sum of the rate parameters from the parent distributions from which the samples were drawn. This result can be generalized, using Equations A6 and A7 recursively, to prove that the distribution of minima sampled from n exponential distributions is itself an exponential distribution with a rate parameter equal to the sum of the n rate parameters. This generalization is important because it allows Bundesen (and me) to predict the mean and the standard deviations of the finishing times of the race; they are simply the reciprocal of the rate parameter of the exponential distribution that describes the race.

Accuracy

Response probabilities can be derived from Equations A6 and A7. The probability that $f_1(t)$ wins the race can be obtained by integrating the first term on the right-hand side of A6, that is, $f_1(t)[1 - F_2(t)]$, and the probability that $f_2(t)$ wins the race can be obtained by integrating the second term on the right-hand side of A6, that is, $f_2(t)[1 - F_1(t)]$. The results for exponential distributions can be obtained by integrating the two terms on the right-hand side of the top line of Equation A7. The probability that $f_1(t)$ finishes first is

$$\begin{aligned} P(1 \text{ first}) &= \int_0^\infty \lambda_1 \exp[-\lambda_1 t] \exp[-\lambda_2 t] dt \\ &= \frac{\lambda_1}{\lambda_1 + \lambda_2}, \end{aligned} \quad (\text{A8})$$

and the probability that $f_2(t)$ finishes first is

$$\begin{aligned} P(2 \text{ first}) &= \int_0^\infty \lambda_2 \exp[-\lambda_2 t] \exp[-\lambda_1 t] dt \\ &= \frac{\lambda_2}{\lambda_1 + \lambda_2}. \end{aligned} \quad (\text{A9})$$

These results can be generalized by recursion to samples from n different exponential distributions. In general, the probability that the sample from one distribution finishes first is simply the ratio of the rate parameter for that distribution to the sum of the rate parameters for all of the distributions in the race. Substituting $v(x, i)$ for λ_i yields the following general equations: The probability that item x finishes first (i.e., that x is the first item categorized as i) is

$$P(x \text{ first}) = \frac{v(x, i)}{\sum_{z \in S} v(z, i)}, \quad (\text{A10})$$

and the probability that categorization i finishes first (i.e., that i is the first categorization made of x) is

$$P(i \text{ first}) = \frac{v(x, i)}{\sum_{j \in R} v(x, j)}. \quad (\text{A11})$$

Equations A10 and A11 can be used to generate predictions about accuracy, if x is the correct item to select and the other items in S are incorrect (Equation A10) or if i is the correct categorization and the other categorizations in R are incorrect.

Appendix B

Pigeonholing and Filtering in TVA

There are two selection mechanisms in TVA, pigeonholing and filtering. Following Broadbent (1971), pigeonholing involves selecting a categorization for a display item whereas filtering involves selecting a display item to be processed. Pigeonholing and filtering are separable selection mechanisms in that increasing the likelihood of a particular categorization (pigeonholing) should not affect the likelihood that a particular item is the first object of the categorization (filtering), and increasing the likelihood that a particular item is selected (filtering) should not affect the likelihood that the item is categorized in a particular way (pigeonholing).

In TVA, pigeonholing is accomplished by manipulating β_i and filtering is accomplished by manipulating w_x . The manipulations seem similar because β_i has is computed by multiplying the $\eta(x, i)$ value (see Equation 5), and w_x has is computed by multiplying the $\eta(x, i)$ value by a pertinence parameter, π_i (see Equation 6). Many people (including me) have difficulty understanding how pigeonholing and filtering could be separable given the similarity in the way their manipulations are effected. The purpose of this appendix is to make clear the reasons why the mechanisms are separable.

β and Pigeonholing

Bundesen (1990) argued that manipulating the bias to categorize an item as i , β_i , affected the probability of categorizing every item in the visual field as i without affecting the probability that any particular item would be the first one categorized as i . This effect, known as pigeonholing (Broadbent, 1971), can be seen by expanding the $v(x, i)$ terms in Equation A11 so that they represent the product of $\eta(x, i)$ and β_i .

The probability that i is the first categorization made of x is:

$$\begin{aligned} P(i \text{ first}) &= \frac{v(x, i)}{\sum_{j \in R} v(x, j)} \\ &= \frac{\eta(x, i)\beta_i}{\sum_{j \in R} \eta(x, j)\beta_j}. \end{aligned} \quad (\text{B1})$$

Increasing β_i will increase the numerator of Equation B1 and therefore increase the probability that i is the first categorization to finish for x .

The probability that x is the first item categorized as i is given by Equation A10:

$$\begin{aligned} P(x \text{ first}) &= \frac{v(x, i)}{\sum_{z \in S} v(z, i)} \\ &= \frac{\eta(x, i)w_x}{\sum_{z \in S} \eta(z, i)w_z}. \end{aligned} \quad (\text{B2})$$

This probability is independent of β_i . In other words, increasing β_i will not increase the probability that any particular item, x , will finish first (i.e., be the first to be categorized as i) because β_i increases the probability that every item will finish first by the same amount. If $\beta_i = 0.9$, then every $\eta(x, i)$ value is multiplied by 0.9, whether η is large or small for that particular x . Thus, increasing β_i has the effect of shrinking the time scale of the race between the different items. However, shrinking the time scale does not affect the order in which the items finish. Consequently, manipulating β_i affects the probability of categorizing an item as i without affecting which item will be the first one to be categorized as i . This is the biasing effect of pigeonholing, as envisioned by Broadbent (1971).

(Appendix continues on next page)

Put differently, Equations B1 and B2 show that the effect of increasing β_i is spread over every item in the display—it increases the probability that every item will be categorized as i . Again, this is the biasing effect of pigeonholing, envisioned by Broadbent (1971) and incorporated into TVA by Bundesen (1990).

Filtering and w_x

Bundesen (1990) argued that manipulating the attentional weight on an item x , w_x , affected the probability that item x would be selected without affecting the probability that x would be categorized in any particular way. Increasing w_x will increase the numerator of Equation B2 and therefore increase the probability that item x is selected first. Increasing w_x will not increase the likelihood that item x will be categorized in any particular way, because w_x affects all categorizations of x to the same extent. This follows from Equation B1. The attentional weight, w_x , drops out of the equation and therefore has no effect on the probability that x will be categorized as i . Thus, if $w_x/\sum w_x$ was 0.9, the η values for each categorization of x would be multiplied by 0.9, shrinking

the time scale with which the categorizations finished but not affecting which categorization finished first.

The logic seems clear when w_x is the focus of the argument. However, the person does not manipulate w_x directly, but instead, manipulates π_i , which determines w_x . The effect is shown in Equation 6, which is reproduced here:

$$w_x = \sum_{i \in R} \eta(x, i) \pi_i.$$

The potential for confusion stems from the fact that π_i and β_i both have their effects by multiplying the $\eta(x, i)$ values. How can π and β have separate effects when they both multiply η ?

The answer lies in the scope of the effects. Changes in β affect all items equally (see Equation 5). Changes in π cause changes in attentional weights (see Equation 6), and a change in the attentional weight of an item affects all categorizations of the item equally (see Equation 5). Thus, the effects of β_i are spread over all the items in the display and consequently change the likelihood that every item is categorized as i , whereas the effects of π_i are spread over all categorizations of item x and consequently change the likelihood of all possible categorizations of item x .

Appendix C

Details of the CTVA Fits

This section is intended to describe the fits of the CTVA model in enough detail to allow interested readers to replicate them for themselves. In order to make analytically tractable fits, I made a number of simplifying assumptions. The most important one was to fit all of the data with 1-D CODE surfaces, for which the boundaries of the above-threshold regions were defined as points that could be found by computing local minima on the CODE surface. The boundaries of above-threshold regions on more realistic 2-D CODE surfaces were lines that would not be easy to compute analytically. Fortunately, many of the data sets I fitted displayed items in linear arrays, so the 1-D CODE surfaces were appropriate.

The fits were calculated deterministically using the equations in the text of the article. One of the virtues of CTVA is that predictions can be derived analytically without stochastic simulation. In principle, the CTVA predictions can be calculated with pencil and paper using the equations in the text and the procedures described here. In practice, I used Pascal programs to generate predictions, to make it easier to explore the effects of varying parameters and to find parameter values that produced good fits to the data.

Prinzmetal (1981)

The first step in fitting Prinzmetal's (1981) data was to define the feature catch. I used two 1-D CODE surfaces to define the feature catch, one for objects within groups and one for objects between groups. The within-group surface was constructed, essentially, by drawing a horizontal line through the rows of circles in Figure 3 and positioning the centers of the feature distributions in the centers of the circles. Objects within groups were 125 units apart, and objects between groups were 250 units apart. I chose a high and low threshold along that CODE surface, based on the local minima between the centers of the circles. For the high threshold, I set the limits of integration one unit more than the local minimum to create separate above-threshold regions for each circle with two units of distance separating them. For the low threshold, I set the limits of integration one unit less than the local minimum so that the above-threshold regions centered on each circle would overlap

and form one large region. The within-group feature catch was calculated by integrating a 1-D Laplace distribution (see Equations 1 and 3) between the limits of the above-threshold regions.

When the distractor was in the same group (left panels of Figure 7) and the threshold was low, the feature catches for the target and distractor features were equal, since they all fell within the same above-threshold region. When the threshold was set high, the feature catch for the target included the area under its distribution within the above-threshold region centered on the target and the area under the distractor distribution that fell within the above-threshold region centered on the target. (Recall that the center of the distractor distribution was 250 units away from the center of the target distribution.) The high-threshold feature catch for the distractor included the area of the distractor distribution that fell within the above-threshold region centered on the distractor and the area of the target distribution that fell within the above-threshold region centered on the distractor.

The between-group CODE surface was constructed in a similar manner to the within-group surface, by drawing a vertical line between vertically aligned circles in Figure 7 and positioning the centers of the feature distributions on the centers of the circles. I used the limits of integration defined for the within-group high and low thresholds to compute the between-group feature catch. When the distractor was in a different group and the threshold was low, the feature catch for the target included the area of the target distribution that fell within the above-threshold region centered on the target plus the area of the distractor distribution that fell within the above-threshold region centered on the target. The feature catch for the distractor was defined similarly, computing the areas of the target and distractor distributions that fell within the above-threshold region centered on the distractor. When the threshold was high, the calculation was essentially the same except that the above-threshold regions were slightly smaller.

The second step in fitting CTVA to the data was to calculate response probabilities. The feature catches defined in the first step were plugged into Equation 12 to compute $v(x, i)$ values, and the $v(x, i)$ values were plugged into Equation 7 to compute response probabilities. In order to report target presence, the two target features—a horizontal and a ver-

tical line—had to finish the race. Evidence for each feature raced against evidence for its absence. Thus, *horizontal present* raced against *horizontal not present*. I set the η values for absent features equal to 1 minus the η values for present features. The probability of correctly detecting a target was set equal to the product of the probability of correctly detecting the horizontal feature and the probability of correctly detecting the vertical feature. Attention weights (based on π values) were fixed at 1. The fits in Table 1 were obtained by manipulating three parameters—the standard deviation of the feature distributions, the biases (β values) for target presence and absence (which were constrained to sum to 1.0), and the η values for horizontal and vertical features (which were constrained to be equal). I did not try to optimize the fit formally (i.e., with a curve-fitting program that searches for parameter values that minimize least squares, etc.), but I did try to find parameters that approximated the observed data. The fits in Table 1 are based on a feature-distribution standard deviation of 50, β of 0.9 and 0.1 for feature presence and absence, respectively, and η of 0.99 and 0.01 for feature presence and absence, respectively.

Cohen and Ivry (1989)

Experiments 1 and 2

The fits to Cohen and Ivry's (1989) data involved computing the feature catches and then the response probabilities. To compute the feature catch, a 1-D CODE surface was created with the centers of the feature distributions 50 units apart in the near condition and 250 units apart in the far condition. The threshold was set just above the local minimum between the two objects (i.e., almost midway between the objects). The feature catch for the target was computed by integrating a 1-D Laplace distribution within the limits of the above-threshold region centered on the target. The feature catch for the target also included the area of the distractor distribution that fell within the above-threshold region centered on the target.

Once the feature catches were computed, response probabilities were computed using Equations 12 and 7. The attention weights and biases were set equal to 1.0 for each color and letter categorization. Eta values were set for each of the four colors and for each of the two letters. Eta values for colors that were present in the display ranged between 0 and 1; η values for colors that were not present in the display were set to 1 minus the eta values for colors that were present. Similarly, the η value for the target letter ranged between 0 and 1 and the η value for the unrepresented letter was set to 1 minus the value for the target. So, for example, if the target was a pink *F* and the distractor was a green *O*, and the η values for pink and green were set to 0.9 and the η values for yellow and blue would be set to 0.1; if the η value for *F* were set to 0.9 and the η value for *X* would be set to 0.1. The different colors raced against each other, following Equation 7, as did the letters. The probabilities of the various combinations of outcomes listed in Table 2 were computed by multiplying and adding the probabilities computed from Equation 7.

The fits in Table 2 depended on two free parameters: the standard deviation of the feature distributions, which was set at 50, and the η values for color and letter presence, which were set equal to each other at 0.9. The η values for absent colors and letters were constrained to equal 1 minus the η values for present colors and letters.

Experiments 3 and 4

The fits to Experiments 3 and 4 involved computing 1-D CODE surfaces for each of the 12 conditions listed in Table 3, calculating the fea-

ture catches, feeding the feature catches into Equations 12 and 7, and combining the different outcomes to produce the predicted response probabilities. The 12 conditions differed in terms of the placement of the two letters for the conjunction task and in terms of the placement of the two digits for the primary task. The closest spacing was between letters and digits (e.g., between *w* and *X* in Condition Small CD in Table 3). It was set equal to 25 units. The closest spacing between letters was twice as large (e.g., between *X* and *Y* in Condition Small CD in Table 3). It was set equal to 50 units. All other distances were multiples of 25 or 50.

The thresholds were set just above the local minimum on the CODE surface between each of the letters from its nearest neighbor. Thus, in Condition Small CD in Table 3, thresholds were set at the local minima between *w* and *X* and between *Y* and *z*. In Condition Far CD, thresholds were set at the local minimum between *X* and *Y*. The feature catches were computed by integrating the area of the distribution for the target letter and the distractor letter that fell within the above-threshold region surrounding the target letter, by integrating the area of the distribution for target and distractor letters that fell within the above-threshold region surrounding the distractor letter, and then averaging the two values. I did this because Cohen and Ivry (1989) did not report data separately for targets and distractors in the alternative positions (i.e., targets could appear in the positions occupied by the *X*'s or the *Y*'s in Table 3; Cohen and Ivry averaged over positions in each of the 12 conditions, so I did the same).

The feature catches were plugged into Equations 12 and 7 in the same manner as in the analysis of Experiments 1 and 2. I calculated the probability of detecting each color and the probability of detecting each letter, and I combined them by multiplying and adding to create the six categories listed in Table 2. I computed illusory conjunction rates in the same way Cohen and Ivry (1989) did, by subtracting half of the probability of a color feature error from the probability of a color conjunction error (correct target letter; correct distractor color). Those values appear in Table 3 and Figure 9.

The fits in Table 3 and Figure 9 depend on two free parameters: the standard deviation of the feature distributions (set at 100) and the η values for color and letter presence (set at 0.825), which were constrained to be equal. The η values for color and letter absence were set at 1 minus the η values for color and letter presence. Altogether, there were 72 data points to be predicted in each experiment—six response categories in 12 conditions. The correlation between the CTVA predictions and the data was 0.955 in Experiment 3 and 0.942 in Experiment 4. The data from the two experiments correlated 0.976 with each other, so CTVA captured a large proportion of the reliable variance.

Banks and Prinzmetal (1976)

The fits to Banks and Prinzmetal's (1976) data were difficult because the displays were 2-D rather than linear arrays of characters. Nevertheless, I approximated the spatial distribution of items with 1-D CODE surfaces. I constructed three 1-D CODE surfaces to represent near, middle, and far neighbors. Each surface had two feature distributions, one representing a potential target and another representing the distractor. The distances between the centers of the distributions were 100, 141, and 200 units for near, middle, and far distractors, respectively. The standard deviation of the feature distributions was set to 50.

The CTVA fits set the threshold in two different ways. The *same threshold* fits assumed that there was only one threshold applied to the whole display. It was set just above the local minimum between the target and the near distractor 100 units away. This threshold overestimated the minimum threshold in the isolated target condition and un-

(Appendixes continue on next page)

derestimated it (slightly) in the camouflaged target condition, where other near neighbors would raise the local minimum on the 2-D CODE surface and move it toward the target. I calculated feature catches for near, middle, and far distractors using this threshold by integrating the feature distributions for the distractors within the limits of the above-threshold region surrounding the potential target.

The *different threshold* fits set the threshold just above the local minimum between a target and its nearest neighbor, so the threshold varied between conditions. In the isolated target conditions, for example, the nearest neighbor was 141 (Conditions A and C) and 200 (Condition B) units away; in the camouflaged target condition, the nearest neighbor was 100 units away. These thresholds underestimated the local minimum in each display because they ignored the contribution of the other feature distributions to the 2-D CODE surface. Nevertheless, they were a reasonable approximation that could be computed analytically.

The feature catches computed in these two ways were used to compute $v(x, i)$ values, and the $v(x, i)$ values were used to compute response probabilities and processing times for the parallel and serial models described in the text of the article. Those computations should be sufficiently clear, so I will not describe them further here.

Cohen and Ivry (1991)

The fits to Cohen and Ivry's (1991) conjunction search experiments were relatively straightforward. I used 1-D CODE surfaces to compute the feature catches and considered only two feature distributions in each CODE surface, one representing the target and one representing the distractor in target-present displays and one representing each of two distractors in target-absent displays. I set the feature distributions 100 units apart in the clumped condition and 200 units apart in the spread

condition. In order to increase accuracy, I set the threshold halfway between the local minimum between the feature distributions and the peak of one of the feature distributions. The standard deviation of the feature distributions was set to 50. These feature catches were used to generate $v(x, i)$ values, and the $v(x, i)$ values were used to generate accuracies and processing times for individual comparisons, using the equations developed in Appendix D. I interpreted these accuracies and processing times in terms of a serial search model, following common practice in the search literature (e.g., Cave & Wolfe, 1990; Treisman & Gelade, 1980; Treisman & Sato, 1990; Wolfe, 1994), but they could be interpreted in terms of a parallel processing model in which several items were processed in parallel (cf. Pashler, 1987; Pylyshyn, 1989).

Eriksen and Eriksen (1974)

The fits to Eriksen and Eriksen's (1974) data were straightforward. I generated a 1-D CODE surface from three items—a central target and two flanking distractors. There were three distances between the target and the distractors, which I set to 50, 100, and 150 units. I set the standard deviation of the feature distributions to 50 units. The thresholds were set just above the local minimum between the target and the distractors, and the feature catch for targets and distractors was computed by integrating their respective distributions within the above-threshold region.

The feature catches were used to modify the $v(x, i)$ values, and the $v(x, i)$ values were used in Equations 9 and 10 to predict accuracy and reaction time. There were three free parameters. The η value for H given H and S given S was fixed at 1.0, and the η values for S given H and H given S and those for S given a neutral distractor and H given a neutral distractor were allowed to vary. The third parameter was the counter criteria, K_H and K_S , which were constrained to be equal to each other.

Appendix D

TVA and Conjunction Search

Standard Conjunction Search

Conjunction search requires discriminating perceptual objects that contain all of the target features from objects that do not contain all of them. The standard conjunction search task involves two features. For example, if the target is a red T , the discrimination is between red T s on the one hand, and not-red T s, red not- T s, and not-red not- T s on the other (see e.g., Treisman & Gelade, 1980). The TVA analysis assumes there is a $v(x, i)$ value for each of these alternatives, where x is the perceptual object and i is red, T , not-red, or not- T . The TVA analysis assumes further that the categorizations race against each other and decisions about whether an object is a target depend on the outcome of the race. If the item is not a target, the race is straightforward: red, T , not-red, and not- T race against each other and the object is not a target if not-red or not- T finish before red and T . If the item is a target, the race is more complicated. Both red and T must finish before not-red or not- T . This complicates the formal analysis of the race. In essence, the slower of red and T race against the faster of not-red and not- T . Thus,

$$\text{Outcome} = \min[\max(r, T), \min(\bar{r}, \bar{T})].$$

The probability density function for $\max(\text{red}, T)$ is

$$f_{\max}(x) = v_r \exp[-v_r x] (1 - \exp[-v_T x]) + v_T \exp[-v_T x] (1 - \exp[-v_r x]),$$

and its distribution function is

$$F_{\max}(x) = (1 - \exp[-v_r x]) + (1 - \exp[-v_T x]) - (1 - \exp[-(v_r + v_T)x]).$$

The probability density function for $\min(\text{notred}, \text{not}T)$ is

$$f_{\min}(x) = (v_{\bar{r}} + v_{\bar{T}}) \exp[-(v_{\bar{r}} + v_{\bar{T}})x],$$

and its distribution function is

$$F_{\min}(x) = 1 - \exp[-(v_{\bar{r}} + v_{\bar{T}})x].$$

The probability density function for the minimum of these random variables is

$$f(x) = f_{\min}(x)(1 - F_{\max}(x)) + f_{\max}(x)(1 - F_{\min}(x)).$$

Substituting the density and distribution functions into this expression yields

$$\begin{aligned} f(x) = & (v_r \exp[-v_r x])(1 - \exp[-v_T x]) \\ & + (v_T \exp[-v_T x])(1 - \exp[-v_r x])(\exp[-(v_{\bar{r}} + v_{\bar{T}})x]) \\ & + [(v_{\bar{r}} + v_{\bar{T}}) \exp[-(v_{\bar{r}} + v_{\bar{T}})x]] \{ \exp[-v_r x] \\ & + \exp[-v_T x] - \exp[-(v_r + v_T)x] \}, \end{aligned}$$

which is the distribution of finishing times for the race. The mean of this distribution is

$$FT = \frac{1}{v_r + v_{\bar{r}} + v_{\bar{T}}} + \frac{1}{v_T + v_{\bar{r}} + v_{\bar{T}}} - \frac{1}{v_r + v_T + v_{\bar{r}} + v_{\bar{T}}}.$$

This mean finishing time is not conditionalized on the outcome of the race. It includes cases in which $\max(\text{red}, T)$ wins as well as cases in which $\min(\text{notred}, \text{not}T)$ wins. In order to model conjunction search, we need the mean finishing times conditional on each runner winning the race and we need the probabilities that each runner will win. The probabilities can be derived from two additive terms on the right-hand side of the unconditional probability density function. Thus, the probability $P(P)$ that $\max(\text{red}, T)$ will win the race (i.e., the probability that a target is judged to be present) is

$$\begin{aligned} P(P) &= \int_0^\infty [(v_r \exp[-v_r x])(1 - \exp[-v_T x]) \\ &\quad + (v_r \exp[-v_T x])(1 - \exp[-v_{\bar{r}} x])](\exp[-(v_{\bar{r}} + v_{\bar{T}})x]) dx \\ &= \frac{v_r}{v_r + v_{\bar{r}} + v_{\bar{T}}} + \frac{v_T}{v_T + v_{\bar{r}} + v_{\bar{T}}} - \frac{v_r + v_T}{v_r + v_T + v_{\bar{r}} + v_{\bar{T}}}, \end{aligned}$$

and the mean conditional finishing time for $\max(\text{red}, T)$ is

$$\begin{aligned} FT_P &= \frac{1}{P(P)} \left(\frac{v_r}{(v_r + v_{\bar{r}} + v_{\bar{T}})^2} \right. \\ &\quad \left. + \frac{v_T}{(v_T + v_{\bar{r}} + v_{\bar{T}})^2} - \frac{v_r + v_T}{(v_r + v_T + v_{\bar{r}} + v_{\bar{T}})^2} \right). \end{aligned}$$

The probability $P(A)$ that $\min(\text{notred}, \text{not}T)$ will win the race (i.e., the probability that a target is not present) is

$$\begin{aligned} P(A) &= \int_0^\infty ((v_{\bar{r}} + v_{\bar{T}}) \exp[-(v_{\bar{r}} + v_{\bar{T}})x]) \\ &\quad (\exp[-v_r x] + \exp[-v_T x] - \exp[-(v_{\bar{r}} + v_{\bar{T}})x]) dx \\ &= \frac{v_{\bar{r}} + v_{\bar{T}}}{v_r + v_{\bar{r}} + v_{\bar{T}}} + \frac{v_{\bar{r}} + v_{\bar{T}}}{v_T + v_{\bar{r}} + v_{\bar{T}}} - \frac{v_{\bar{r}} + v_{\bar{T}}}{v_r + v_T + v_{\bar{r}} + v_{\bar{T}}}, \end{aligned}$$

and the mean conditional finishing time for $\min(\text{notred}, \text{not}T)$ is

$$FTA = \frac{1}{P(A)} \left(\frac{v_{\bar{r}} + v_{\bar{T}}}{(v_r + v_{\bar{r}} + v_{\bar{T}})^2} + \frac{v_{\bar{r}} + v_{\bar{T}}}{(v_T + v_{\bar{r}} + v_{\bar{T}})^2} - \frac{v_{\bar{r}} + v_{\bar{T}}}{(v_r + v_T + v_{\bar{r}} + v_{\bar{T}})^2} \right).$$

The TVA analysis assumes that people search through the perceptual objects in the display in a self-terminating fashion. Under that assumption, the mean conditional finishing times determine reaction time and the slope of the function relating reaction time to the number of items in the display. The expressions for mean reaction time and accuracy as a function of display size are given in the main body of the article.

Triple Conjunction Search

Wolfe, Cave, and Franzel (1989) tested people in a *triple conjunction* task, in which targets were conjunctions of three features (e.g., small red Ts) and distractors contained only one of the target features (e.g., small green Xs, large red Xs, or large green Ts). The TVA analysis can be extended to triple conjunction search as well by including a $v(x, i)$ value for each of the three features and their absence (i.e., not small, not red, and not T). As with double conjunctions, decisions about target presence are determined by the outcome of a race between the presence and absence of the target features:

$$\text{Outcome}_3 = \min[\max(r, T, s), \min(\bar{r}, \bar{T}, \bar{s})].$$

Notice that there are three runners in the race for target absence.

which is more than the two that raced for target absence in standard conjunction search. This is important because the fastest of three runners will finish before the faster of two runners (see, e.g., Logan, 1988, 1992), and this will reduce the slope of the function relating reaction time to display size because the slope is determined by the rate at which target absence is decided. Thus, the TVA analysis predicts shallower slopes in triple conjunction search than in standard, double conjunction search. This is an important conclusion because the difference is attributed to the comparison process, whereas Wolfe et al. (1989) and others (e.g., Grossberg, Mingolla, & Ross, 1994; Treisman & Sato, 1990) attributed it to preattentive processes that precede the comparison process.

The mean finishing times and response probabilities for triple conjunction search can be determined in the same way as for double conjunction search. For target-present decisions, the probability of a correct decision is

$$\begin{aligned} P(P) &= \frac{v_r}{v_r + v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}}} + \frac{v_T}{v_T + v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}}} + \frac{v_s}{v_s + v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}}} \\ &\quad - \frac{v_r + v_T}{v_r + v_T + v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}}} - \frac{v_r + v_s}{v_r + v_s + v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}}} \\ &\quad - \frac{v_T + v_s}{v_T + v_s + v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}}} + \frac{v_r + v_T + v_s}{v_r + v_T + v_s + v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}}}, \end{aligned}$$

and mean finishing time is

$$\begin{aligned} FT_P &= \frac{1}{P(P)} \left[\frac{v_r}{(v_r + v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}})^2} + \frac{v_T}{(v_T + v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}})^2} \right. \\ &\quad + \frac{v_s}{(v_s + v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}})^2} - \frac{v_r + v_T}{(v_r + v_T + v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}})^2} \\ &\quad - \frac{v_r + v_s}{(v_r + v_s + v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}})^2} - \frac{v_T + v_s}{(v_T + v_s + v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}})^2} \\ &\quad \left. + \frac{v_r + v_T + v_s}{(v_r + v_T + v_s + v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}})^2} \right]. \end{aligned}$$

For target-absent decisions, the probability of a correct decision is

$$\begin{aligned} P(A) &= \frac{v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}}}{v_r + v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}}} + \frac{v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}}}{v_T + v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}}} + \frac{v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}}}{v_s + v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}}} \\ &\quad - \frac{v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}}}{v_r + v_T + v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}}} - \frac{v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}}}{v_r + v_s + v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}}} \\ &\quad - \frac{v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}}}{v_T + v_s + v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}}} + \frac{v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}}}{v_r + v_T + v_s + v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}}}, \end{aligned}$$

and mean finishing time is

$$\begin{aligned} FTA &= \frac{1}{P(A)} \left[\frac{v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}}}{(v_r + v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}})^2} + \frac{v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}}}{(v_T + v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}})^2} \right. \\ &\quad + \frac{v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}}}{(v_s + v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}})^2} - \frac{v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}}}{(v_r + v_T + v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}})^2} \\ &\quad - \frac{v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}}}{(v_r + v_s + v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}})^2} - \frac{v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}}}{(v_T + v_s + v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}})^2} \\ &\quad \left. + \frac{v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}}}{(v_r + v_T + v_s + v_{\bar{r}} + v_{\bar{T}} + v_{\bar{s}})^2} \right]. \end{aligned}$$

Received July 1, 1995

Revision received December 20, 1995

Accepted January 10, 1996 ■