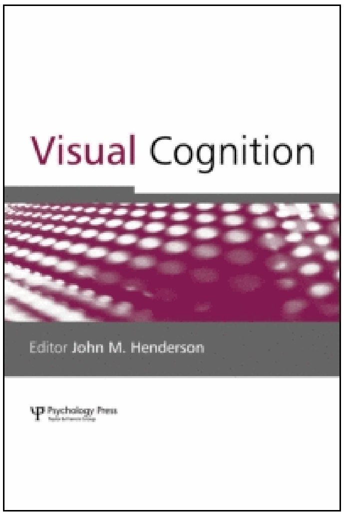

Visual Cognition
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713683696

Visual Selective Attention: Outlines of a Choice Model, a Race Model and a Computational Theory

Claus Bundesen

To cite this Article Bundesen, Claus(1998) 'Visual Selective Attention: Outlines of a Choice Model, a Race Model and a Computational Theory', Visual Cognition, 5: 1, 287 — 309
To link to this Article: DOI: 10.1080/713756774
URL: http://dx.doi.org/10.1080/713756774

VISUAL COGNITION, 1998, *5* (1/2), 287–309

Visual Selective Attention: Outlines of a Choice Model, a Race Model and a Computational Theory

Claus Bundesen
Centre for Visual Cognition, University of Copenhagen, Copenhagen, Denmark

A computational theory of visual selective attention is presented. The theory developed out of a choice and a race model for visual selection from multi-element displays. The choice model (Bundesen, Pedersen, & Larsen, 1984) provides a rule for computation of selection probabilities, which accounts for effects of the selection criterion and the numbers of targets and distractors in the stimulus. The race model (Shibuya & Bundesen, 1988) provides a process interpretation of the choice model and accounts for effects of the exposure duration of the stimulus. The computational theory (TVA; Bundesen, 1990) was constructed by integrating the race model with a biased-choice model for single-stimulus recognition (Luce, 1963). TVA describes two mechanisms (filtering and pigeonholing) by which selection is assumed to be carried out, and it organizes a large body of empirical data on human performance in visual recognition and attention tasks. A recent theoretical development (CTVA; Logan, 1996; Logan & Bundesen, 1996) combines TVA with a theory of perceptual grouping by proximity. CTVA explains effects of spatial separation between items in multi-element displays. The neural localization of the operations described in TVA is considered in the final section.

INTRODUCTION

This paper describes the development of a computational theory of visual selective attention. The theory was originally published under the title "A theory of visual attention" (TVA; Bundesen, 1990). TVA was developed from earlier models of visual selection in particular experimental tasks. The first of these models was a "choice model" for visual selection from multi-element displays (Bundesen, Pedersen, & Larsen, 1984). The choice model provides a

Requests for reprints should be addressed to Claus Bundesen, Centre for Visual Cognition, Psychological Laboratory, University of Copenhagen, Njalsgade 90, DK-2300 Copenhagen S, Denmark. E-mail: bundesen@axp.psl.ku.dk

This research was supported by grants from the Danish Ministry of Education and Research and the International Human Frontier Science Program Organization. Thanks are due to Werner Schneider and anonymous reviewers for helpful comments on an earlier draft of the manuscript.

rule for computation of selection probabilities, and it accounts for effects of both the selection criterion and the numbers of targets and distractors in the stimulus. The choice model is simple, and it is described in the first section of this paper.

The second section describes a "race model" for selection from multi-element displays (Bundesen, 1987; Shibuya & Bundesen, 1988). The choice model can be derived from the race model, so the race model explains all that the choice model explains. The race model also provides a picture of the temporal course of processing. The choice model is only a descriptive model, but the race model is a process model.

The third section outlines the more general theory, TVA. The race model can be derived from TVA, so TVA explains all that the race model explains. TVA also provides a picture of the mechanisms by which selection is assumed to be carried out. The race model is only a process model, but TVA is a computational theory of selective attention in vision.

The third section describes further a recent development of TVA, the "CODE theory of visual attention" (CTVA; Logan, 1996; Logan & Bundesen, 1996). TVA organizes a large body of empirical data on human performance in visual recognition and attention tasks, but TVA has been criticized for neglecting spatial effects in visual attention (e.g. van der Heijden, 1993). CTVA combines TVA with a theory of perceptual grouping by proximity (van Oeffelen & Vos, 1982). The resulting theory explains a wide range of spatial effects in visual attention.

The final section considers the neural localization of the operations described in TVA. Pattern recognition and categorization are assumed to occur in a ventral visual pathway, but the ventral stream of processing is modified by attentional weighting. It is suggested that the weighting is done by a system in the posterior parietal cortex that takes input from the ventral pathway and feeds back to the ventral pathway.

CHOICE MODEL

The choice model for visual selection was developed to describe the results of a series of "partial-report" experiments (cf. Sperling, 1960) performed in our laboratory. In a partial-report experiment, the subjects are presented with a briefly exposed visual display showing a number of different elements. The task is to report as many of those elements as possible that satisfy a particular selection criterion (targets) and ignore any other elements (distractors). For example, the display may show a mixture of red and black letters, and the task may be to name as many of the red letters as possible and ignore the black ones (selection by colour). Or the display may show a mixture of letters and digits, and the task may be to report the digits and ignore the letters (selection by alphanumeric class). Usually, the exposure duration is shorter than the reaction time for eye movements so that information from the display must be extracted from a single fixation.

In our early partial-report experiments (Bundesen et al., 1984; Bundesen, Shibuya, & Larsen, 1985), the exposure duration was kept constant at a value of approximately 100 msec. Before and after the presentation of a stimulus display, the field was dark. The selection criterion varied between experiments. We investigated selection by brightness, selection by colour, selection by shape and selection by alphanumeric class. In each experiment with a given selection criterion, we systematically varied both the number of targets and the number of distractors in the displays.

The results showed that the number of correctly reported targets depended on the selection criterion, the number of targets and the number of distractors in the display. The joint effects of the three factors could be described by a simple model, the choice model.

In the choice model for visual selection, subjects can report a given display element if, and only if, they succeed in encoding the element into a short-term memory store with very limited capacity.[1] The capacity is limited to K elements, and K is one of two basic parameters in the model. In the experimental conditions I have described, subjects are assumed to have enough time to encode K elements from the stimulus display; that is, enough time to fill up the short-term store with elements from the display. In these conditions, the number of correctly reported targets depends on the ability of subjects to encode targets rather than distractors into the short-term store.

Read-in to the short-term store is conceived as selective sampling of elements from the stimulus display. Each element in the display has a certain attentional weight. Until the short-term store has been filled up with elements, the probability that any not-yet-encoded element is the next one to be encoded equals the weight of that element divided by the sum of the weights of all those elements that have not yet been encoded (cf. Luce, 1959).

In our experiments, the elements were alphanumeric characters, and it was plausible to assume that targets had higher weights than distractors, but any two targets within the same display had the same weight, and any two distractors within the same display had the same weight. Without loss of generality, then, the weight of a target could be set to 1 and the weight of a distractor to α. Parameter α is the second of the two basic parameters in the model.

Parameter α is a measure of the efficiency of selection. If α equals 0 (i.e. if the weight of a distractor is 0), then the probability that a distractor is encoded into the short-term store also equals 0. Thus, selection is perfect; only targets are encoded. If α equals 1 (i.e. if the weight of a distractor equals the weight of

[1] In some applications of the model, this assumption has been relaxed. Bundesen et al. (1985) assumed that each target that entered the short-term store was correctly reported with probability θ. Our estimate for θ was .92

a target), then the probability that an element is encoded into the short-term store is independent of whether the element is a target or a distractor. Thus, encoding is non-selective.

Figure 1 shows a maximum likelihood fit of the model to results from an experiment on partial report by colour (blue *vs* green; Bundesen et al., 1985). The mean number of correctly reported targets is plotted as a function of the number of distractors (D) in the display with the number of targets in the display as the parameter. The observed data points are marked by squares, circles and a diamond. The curves show predictions from the choice model with short-term storage capacity K at a value of 3.5 elements and parameter α at .06. (Non-integral values of K were treated as probability mixtures. For example, a value of 3.53 for K was treated as a mixture of the values 3 and 4 such that, on any trial, $K = 4$ with a probability of .53).

Figure 2 shows a maximum likelihood fit of the model to results from a similar experiment on partial report by alphanumeric class (letters *vs* digits; Bundesen et al., 1985). The fit was obtained with parameter K at 3.5 elements and parameter α at .42. It is reassuring that the estimate for K was the same for selection by alphanumeric class and for selection by colour; the capacity of the

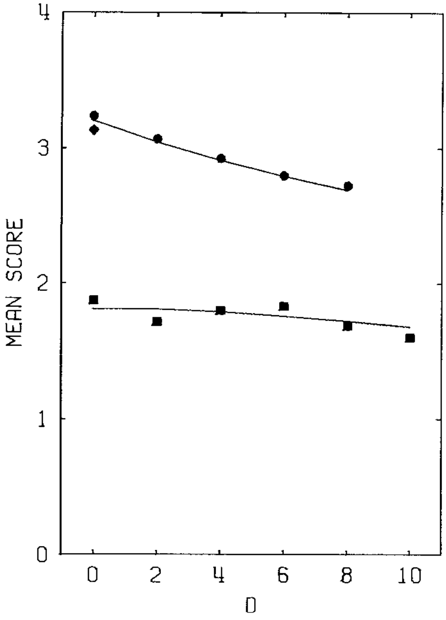

FIG. 1. Number of correctly reported targets (mean score) for subject HS as a function of the number of distractors (D) with the number of targets as a parameter for selection by colour. The number of targets was 2 (squares), 4 (circles) or 12 (diamond). Unmarked points connected by straight lines represent a theoretical fit to the data by the choice model. Reproduced with permission from Bundesen et al. (1985). ©1985 by The International Association for the Study of Attention and Performance.

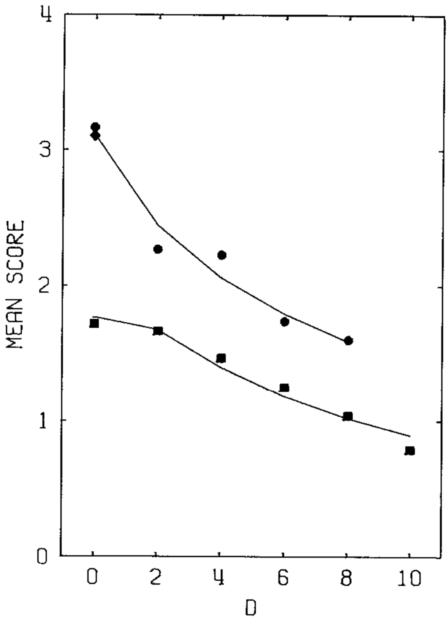

FIG. 2. Number of correctly reported targets (mean score) for subject HS as a function of the number of distractors (*D*) with the number of targets as a parameter for selection by alphanumeric class. The number of targets was 2 (squares), 4 (circles) of 12 (diamond). Unmarked points connected by straight lines represent a theoretical fit to the data by the choice model. Reproduced with permission from Bundesen et al. (1985). © Copyright 1985 by The International Association for the Study of Attention and Performance.

short-term store should not depend on the selection criterion. The effect of the selection criterion was completely accounted for by variation in parameter α; α was much higher for selection by alphanumeric class than for selection by colour.

RACE MODEL

The choice model is easily understood and readily applied, and it has helped to provide a good description of many experimental results. But the choice model is lacking in depth. It provides a rule for calculating selection probabilities, but it provides no explanation for the rule.

Bundesen et al. (1985) noted that the choice model can be derived from a simple race model of the underlying processes. The race model describes the selection process as a race between the elements in the visual field: The presentation of a stimulus display releases an encoding process for each of the elements in the display, and the elements that are selected are those elements that get encoded into the short-term store before the stimulus presentation terminates and before the short-term store has been filled up. If the encoding times for different elements are stochastically independent and exponentially distributed, the choice model must hold. That is, the choice model can be derived mathematically from the general race model if one assumes that the encoding times are mutually independent, exponentially distributed random variables (Bundesen et al., 1985; also see Bundesen, 1993b).

Shibuya and Bundesen (1988) developed a particular race model called FIRM, which stands for "Fixed-capacity independent race model" (see also Shibuya, 1991). FIRM describes the processing of a stimulus display as a two-stage process. During the first stage of processing, a weight is computed for each element in the display. The weight is a measure of the strength of the sensory evidence that the element is a target.

During the second stage of processing, a fixed amount of processing capacity (C elements per second) is distributed among the elements in the display, and the race between the elements takes place. The amount of processing capacity that is allocated to an element determines how fast the element can be encoded into the short-term store. The total processing capacity of C elements per second is distributed across the elements in proportion to their weights. Thus, every element x is allocated a processing capacity equal to C times the weight of x divided by the sum of weights across all elements.

The time taken to encode an element x is assumed to be exponentially distributed, with probability density function $\mu_x \exp(-\mu_x t)$, where the rate parameter μ_x equals the amount of processing capacity that is allocated to the element. Encoding times for different elements are stochastically independent. The elements actually selected are those elements whose encoding processes are completed before the stimulus presentation terminates and before the short-term store has been filled up.

Because the choice model for visual selection can be derived from FIRM, FIRM explains all that the choice model explains. In addition, FIRM predicts effects of variations in the exposure duration of a stimulus display. Shibuya and Bundesen (1988) tested such predictions in a comprehensive partial-report experiment. The stimulus displays showed mixtures of letters and digits, and the task was to report as many of the digits as possible while ignoring the letters. Exposure durations ranged from 10 to 200 msec, and each display was terminated by a pattern mask.

Figure 3 shows the probability distribution of the number of correctly reported targets as a function of the exposure duration, the number of targets (T) and the number of distractors (D) in the display for one representative subject. For example, the panel in the lower right-hand corner shows the results for displays with six targets and four distractors. The top curve represents the probability that the subject reported at least one element correctly as a function of exposure duration; the curve below shows the probability that the subject reported at least two elements correctly, and so on. The observed data points

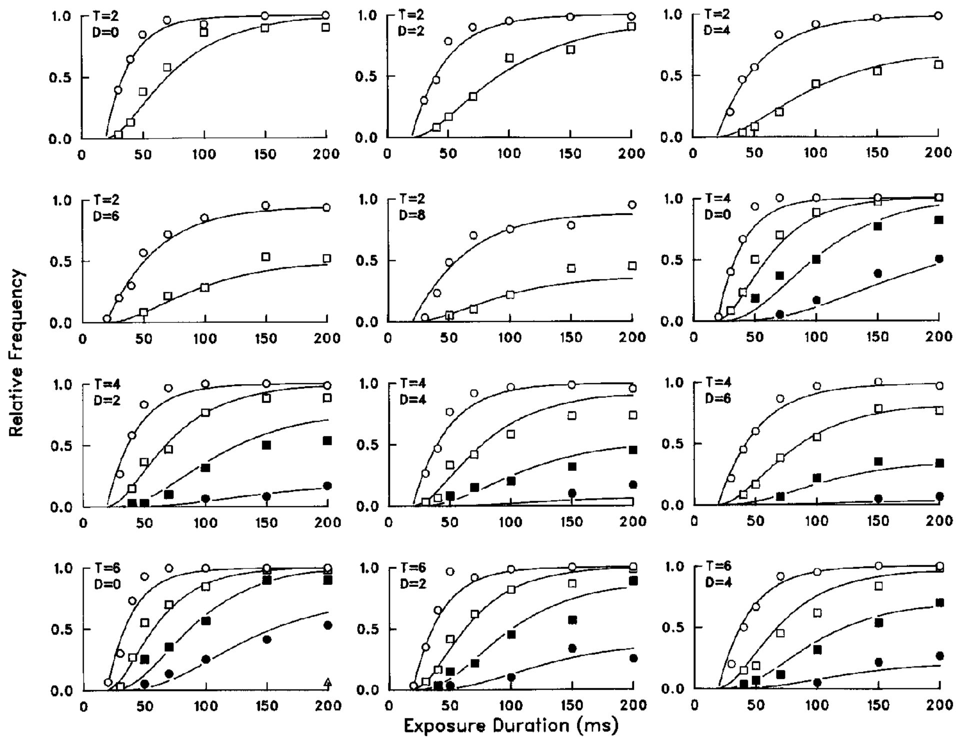

FIG. 3. Relative frequency of scores of j or more correctly reported targets as a function of exposure duration, with j, number of targets (T), and number of distractors (D) as parameters in the experiment of Shibuya and Bundesen (1988). Data are shown for subject MP. Parameter j varies within panels; j is 1 (open circles), 2 (open squares), 3 (solid squares), 4 (solid circles) or 5 (triangle). T and D vary among panels. Smooth curves represent a theoretical fit to the data by the race model. For clarity, observed frequencies less than 0.02 were omitted from the figure. Reproduced with permission from Shibuya and Bundesen (1988). © 1988 by the American Psychological Association.

are marked by circles and squares. The smooth curves show predictions from FIRM with short-term storage capacity K at a value of 3.7 elements, total processing capacity C at 49 elements per second, and parameter α (the weight ratio of a distractor to a target) at .40. As can be seen, the results agreed well with the predictions.

COMPUTATIONAL THEORY

TVA

TVA is a generalization of FIRM. It integrates findings on single-stimulus recognition, whole report, partial report, detection, and search in a race model framework. In TVA, both visual recognition and selection of elements in the visual field consist in making perceptual categorizations. A perceptual categorization has the form "x belongs to i", where x is an element in the visual field and i is a perceptual category. Examples of perceptual categories are the class of red elements (a colour category), the class of letters of type A (a shape category) and the class of elements in the right visual field (a location category).

That a perceptual categorization is made means that the categorization is encoded into a limited-capacity short-term memory store. If and when one makes the perceptual categorization that x belongs to i (i.e. if and when the perceptual categorization is encoded into the short-term store), element x is said both to be selected and to be recognized as a member of category i. Thus, an element is said to be selected if, and only if, it is recognized as a member of one or other category. Similarly, an element is said to be represented in the short-term store if, and only if, some categorization of the element is represented in the store.

At the moment a perceptual categorization of an element completes processing, the categorization enters the short-term store, provided that memory space for the categorization is available in the store. The capacity of the store is limited to K different elements, and space is available for a new categorization of element x if element x is already represented in the store (with another categorization) or if less than K elements are represented in the store. There is no room for a categorization of element x if the short-term store has been filled up with other elements.

Consider the event that a particular perceptual categorization, "x belongs to i", completes processing at time t. The hazard function of this event has a certain value which I call the v value of the perceptual categorization.[2] Thus the v value is a measure of speed of processing, a measure of the speed at which

[2]The hazard function of the event is the conditional probability that the event occurs at time t, given that the event has not occurred before time t.

the perceptual categorization is processed. In TVA, the v value is determined by two basic equations. By equation 1,

$$v(x, i) = \eta(x, i)\,\beta_i \frac{w_x}{\sum_{z \in S} w_z} \tag{1}$$

where $\eta(x, i)$ is the instantaneous strength of the sensory evidence that element x belongs to category i; β_i is a perceptual bias associated with category i; S is the set of all elements in the visual field; and w_x and w_z are attentional weights of elements x and z, respectively.

The attentional weights are derived from pertinence values. Every perceptual category is assumed to have a certain pertinence value. The pertinence of a category is a measure of the current importance of attending to elements that belong to the category. The weight of an element x in the visual field is given by equation 2:

$$w_x = \sum_{j \in R} \eta(x, j)\,\pi_j \tag{2}$$

where R is the set of all perceptual categories; $\eta(x, j)$ is the instantaneous strength of the sensory evidence that element x belongs to category j; and π_j is the pertinence value of category j. By equation 2, the attentional weight of an element is a weighted sum of pertinence values. The pertinence of a given category enters the sum with a weight equal to the strength of the sensory evidence that the element belongs to the category.

By equations 1 and 2, v values can be expressed as functions of eta, beta and pi values. When eta, beta and pi values are given, processing times for different perceptual categorizations are assumed to be stochastically independent.

In most applications of the theory to the analysis of experimental data, eta, beta and pi values are assumed to be constant during the presentation of a stimulus display. When eta, beta and pi values are constant, v values are also constant. The v values were defined as hazard functions, and when v values are kept constant, processing times become exponentially distributed. The v value of the perceptual categorization that element x belongs to category i becomes the exponential rate parameter for the processing time of this perceptual categorization.

Mechanisms of selection

The theory embraces two mechanisms of selection (cf. Broadbent, 1970, 1971): a mechanism for selection of elements (filtering) and a mechanism for selection of categories (pigeonholing). The filtering mechanism is represented by attentional weights, which are derived from pertinence values. As an example, if selection of red elements is required the pertinence of "red" should be high. Equation 2 implies that when "red" has a high pertinence, red elements get high attentional weights. Accordingly, by equation 1, processing of red elements is fast, so red elements are likely to win the processing race and be encoded into the short-term memory store.

The pigeonholing mechanism is represented by perceptual bias parameters. Pertinence values determine which elements are selected, but perceptual bias parameters determine how the elements are categorized. If particular types of categorizations are desired, bias parameters of the relevant categories should be high. By equation 1, then, the desired types of categorizations are likely to be made.

Applications

TVA has been applied to the results of a broad range of paradigms concerned with single-stimulus recognition and selection from multi-element displays (see Bundesen, 1990, 1996). For single-stimulus recognition, the theory provides a simple derivation of a successful model of effects of visual discriminability and bias—the biased-choice model of Luce (1963; see Bundesen, 1990, p. 527).

Parallel Processing. For selection from multi-element displays, the fixed-capacity independent race model, FIRM, can be derived as a special case of TVA. FIRM describes the selection process as a race between the elements in the visual field. TVA describes the selection process as a race between perceptual categorizations of elements in the visual field. The correspondence between FIRM and TVA is established by identifying the encoding of an element in the short-term store with the encoding of a perceptual categorization of the element. The K parameter of FIRM corresponds to the K parameter of TVA, and the attentional weights of FIRM correspond to the attentional weights of TVA, the w parameters.

Consider the C parameter of FIRM and the associated assumption that the total processing capacity is constant. In TVA, the total processing capacity C can be defined as the sum of v values across all perceptual categorizations of all elements in the visual field:

$$C = \sum_{x \in S} \sum_{i \in R} v(x, i) \tag{3}$$

By equation 1 of TVA, C can often be treated as a constant. For example, suppose we are dealing with stimuli that are equal in discriminability, etc., so that there is a constant k such that

$$\sum_{i \in R} \eta(x, i) \beta_i = k$$

for every element x. Equations 1 and 3 then imply that

$$C = \sum_{x \in S} \sum_{i \in R} \eta(x, i) \beta_i w_x / \sum_{z \in S} w_z = k$$

To put this into words, for the stimulus material in question, the total processing capacity C is a constant.

Serial Processing. Serial processing is assumed to occur when the time cost of shifting attention is outweighed by gain in speed of processing once attention has been shifted. For example, suppose targets and distractors are equal in attentional weights, and consider search through a set of two display items. If the items are processed in parallel with a total capacity of C, the mean time to the first completion of an item is $1/C$, but the mean time from the first to the second completion is twice as high, $2/C$. If processing is serial with a capacity of C, the mean time to the first completion is again $1/C$. The mean time from the first to the second completion equals the time taken to shift attention between items plus $1/C$. Therefore, to optimize performance, processing should be serial if, and only if, the time taken to shift attention is less than $1/C$. For example, if $1/C$ is 100 msec per item, processing should be serial if a shift of attention takes less than 100 msec.

Bundesen (1990) extended this line of reasoning and considered the time cost of shifting attention to a group of n items and then processing the n items in parallel. The time cost per member of the group is found to be shortest if group size n is about the same as the product of the processing capacity (C) and the time taken to shift attention. Numerical calculations suggest that typical data on slow feature (Treisman & Gormican, 1988) and conjunction (Treisman & Gelade, 1980) search can be explained by attention shifting among groups of display items, so that processing is parallel within groups but serial between groups (see Bundesen, 1990, p. 537).

CTVA

Logan has recently proposed a theory—the "CODE theory of visual attention" (CTVA)—that integrates space-based and object-based approaches to visual attention (Logan, 1996; see also Logan & Bundesen, 1996). The theory combines TVA with van Oeffelen and Vos' (1982, 1983) "COntour DEtector" (CODE) theory of perceptual grouping by proximity. In the theory of van Oeffelen and Vos, locations of items in visual displays are represented as distributions in one- or two-dimensional space. Van Oeffelen and Vos originally used normal distributions, but Compton and Logan (1993) found that Laplace distributions worked just as well, and they are easier to work with.

Grouping by proximity is modelled as follows (see Fig. 4). First, each stimulus item is represented by a distribution centred on the position that the object occupies in one- or two-dimensional space. In the one-dimensional case

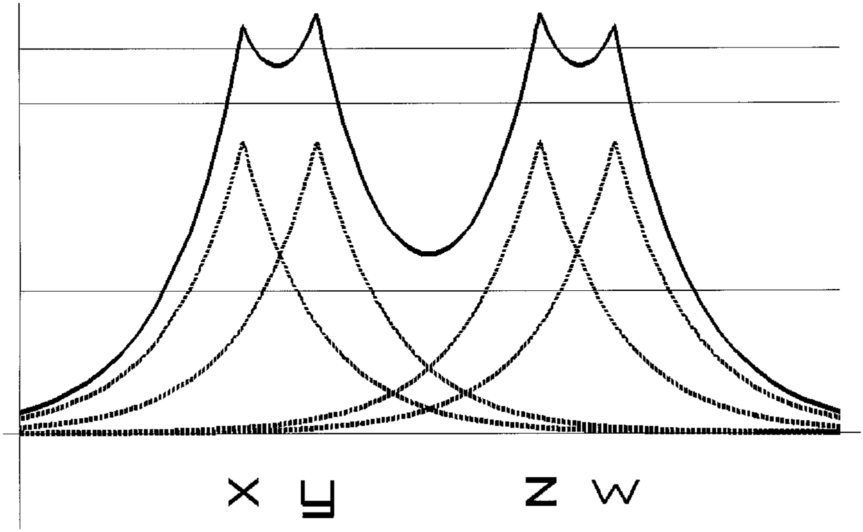

FIG. 4. Perceptual grouping by proximity. Laplace distributions (broken curves) and a CODE surface (solid curve) are shown for four items (*x*, *y*, *z* and *w*) arrayed in one dimension. Thresholds applied to the CODE surface are shown by horizontal lines. The low threshold includes all four items in one group. The middle threshold generates two groups with two items each. The high threshold separates all four items.

(e.g. a linear array of items positioned along a *u*-axis), item *y* may be represented by the Laplace distribution:

$$f_y(u) = \tfrac{1}{2}\lambda_y \exp\left(-\lambda_y \mid u - \theta_y \mid\right) \tag{4}$$

with scale parameter λ_y and position parameter θ_y. Second, a CODE *surface* is constructed by summing the distributions for different items over space. Third, a threshold is applied to the CODE surface, cutting off one or more "above-threshold regions." A perceptual group is defined as an above-threshold region of space; that is, a region for which the code surface is above the threshold.

Groups of different sizes are formed by raising and lowering the threshold. A low threshold produces a small number of groups with many items in each group; a high threshold produces a large number of groups with few items in each. Grouping is hierarchical, so that the smaller groups are nested within the larger groups.

To link CODE to TVA, Logan (1996) assumed that the distribution for an item is a distribution of information about the features of the item. Thus, in equation 4, $f_y(u)$ is the density of information about features of *y* at spatial position *u*. Logan further assumed that TVA samples information from one or more above-threshold regions of the CODE surface. An above-threshold region of the CODE surface is the same as an element in the visual field. At any point in time, the set of regions (elements) from which information is being sampled by TVA is called the "field of spatial attention, A" (Logan & Bundesen, 1996). The field of spatial attention can be a single element such as a particular letter in a multi-letter display. The field of spatial attention can also be a set of elements that make up a single higher-level element. One example is the set of all individual letters in a multi-element display. Another example is a set of letters that form a smaller perceptual group within a multi-letter display.

The amount of information in a given above-threshold region of the CODE surface about a feature from a particular stimulus item is called the "feature catch" from that item in the given above-threshold region (see Fig. 5). It equals the area or volume of the distribution for the item that falls within the limits of the above-threshold region. The amount of information is greatest for items that fall within the above-threshold region, in the sense that the points that represent the centres of their distributions fall within the region. The amount of information is positive for all items in the display, but it decreases as the spatial distance of the item from the given region is increased. Accordingly, in CTVA, "location is special" (cf. Bundesen, 1991, 1993a; Nissen, 1985; van der Heijden, 1993; van der Velde & van der Heijden, 1993).

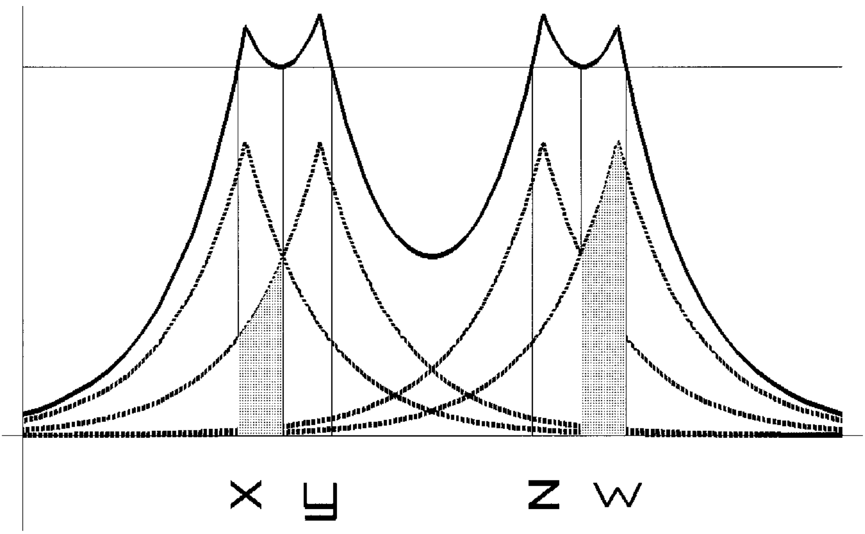

FIG. 5. Feature catch. Laplace distributions (broken curves) and a CODE surface (solid curve) are shown for items *x*, *y*, *z* and *w*. A threshold (horizontal line) applied to the CODE surface generates four above-threshold regions (separated by vertical lines). The feature catch from item *y* in the above-threshold region formed by item *x* [i.e. $c(x, y)$] equals the shaded area to the left. The feature catch from item *w* in the above-threshold region formed by item *w* [i.e. $c(w, w)$] equals the shaded area to the right.

Suppose a threshold is applied to the CODE surface for a multi-element display so that each item in the display forms a separate above-threshold region.[3] Let x and y be items in the display; that is, above-threshold regions of the CODE surface. The catch in the x region of features extracted from the y region, $c(x, y)$, is a measure of the likelihood of sampling features stemming from item y in the processing of item x. In the one-dimensional case,

$$c(x, y) = \int_{\text{region } x} f_y(u)\, du \tag{5}$$

where $f_y(u)$ is given by equation 4, and the integration is done over the above-threshold region formed by item x (cf. Fig. 5).

The feature catch modulates the information input to TVA. Formally, this is represented by replacing eta values $\eta(x, i)$ by "effective" eta values $\eta_e(x, i)$ in equations 1 and 2 of TVA. The effective eta value for the categorization that item x is a member of category i (i.e. x has feature i) is given by:

$$\eta_e(x, i) = \sum_{y \in S} c(x, y)\eta(y, i) \tag{6}$$

where S is the set of all items in the display. By equation 6, the effective evidence that item x has feature i depends upon the evidence that item y has feature i to the extent that features stemming from item y are caught in the above-threshold region formed by item x.

Substituting $\eta_e(x, i)$ for $\eta(x, i)$ in equation 1 of TVA yields the CTVA equation:

$$v(x, i) = \eta_e(x, i)\, \beta_i \frac{w_x}{\Sigma_{z \in S} w_z} \tag{7}$$

Substituting $\eta_e(x, i)$ for $\eta(x, i)$ in equation 2 of TVA, and incorporating the assumption that the attentional weight of an item is 0 if the item is outside the field of spatial attention A, yields the CTVA equation

$$w_x = \begin{cases} \Sigma_{j \in R}\, \eta_e(x, j)\pi_j & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases} \tag{8}$$

CTVA becomes identical to TVA when (a) $A = S$ (i.e. the field of spatial attention coincides with the set of items in the display) and (b) $c(x, x) = 1$ for every item x, but $c(x, y) = 0$ when x is different from y. For example, in many partial-report experiments, it is plausible that (a) the field of spatial attention encompasses all items in the display, and (b) inter-item distances are so long

[3]The condition that each item in the display forms a separate above-threshold region allows feature catch factors to be written as functions of ordered pairs of items, $c(x, y)$. This makes the present formalization more simple than the formalizations in Logan (1996) and Logan and Bundesen (1996)

that feature catches from adjacent items can be neglected. In such cases, an analysis based on CTVA reduces to an analysis based on TVA (e.g. an analysis in terms of the fixed-capacity independent race model of Shibuya & Bundesen, 1988). Thus, CTVA can be viewed as a generalization of TVA, and TVA can be viewed as a special case of CTVA.

Parallel Processing and Feature Integration. Implicit in the account of CTVA by Logan and Bundesen (1996) is a distinction between two types of parallel processing of a multi-element display. The first type of parallel processing is used when a low threshold is applied to the CODE surface, so that the whole stimulus display is processed as a single unit. In this case, feature catch factors are close to 1, so processing is fast, but all features that are extracted are bound to the display as a whole, rather than being bound to the individual elements in the display. Being bound to the display as a whole, the features may be thought of as "free-floating" within the display (c.f. Treisman & Gelade, 1980; Treisman & Schmidt, 1982).

The second type of parallel processing requires that a high threshold is applied to the CODE surface, so that each element in the display is represented by a separate above-threshold region of the CODE surface. In this case, feature catch factors are well below 1, so processing is not so fast, but features are bound to individual elements and not to the display as a whole.

To illustrate the distinction between the two types of parallel processing, consider some recent results reported by Cave (1994; see also Kim & Cave, 1995). In one experiment, subjects searched for the letter *F* in a display of eight letters. The search display was occasionally followed by a probe, which was a small square at one of the eight letter locations. Responses to probes at the target's location were faster than responses to probes at distractor locations, and the size of this difference was greater when the distractor set included letters that were highly confusible with the target (e.g. *E* and *P*). Cave concluded that effects of selective attention were stronger when the distractors were physically more similar to the target.

A possible explanation of the findings of Cave (1994) is as follows. Performance reflected a mixture of the two types of parallel processing. When target–distractor similarity was low, parallel processing with a low threshold predominated. In this case, search was undertaken by testing for the presence of a particular feature in the stimulus display (c.f. the account of "feature search" in the theory of Treisman & Gelade, 1980). The whole display was processed as a single unit and the subject tested whether the unit (the display) had a particular (target-specific) feature. Because the display was processed as a single unit, all features that were extracted were bound to the display as a whole, rather than being bound to units at the level of individual letters. With only one unit in the visual field, that unit received all of the available processing capacity. There was no attenuation of processing due to spatial attention, so processing was fast. When this mode of processing was used, reaction times to probe dots were the same for dots presented at target locations as for dots presented at distractor locations.

When target–distractor similarity was high, parallel processing with a high threshold predominated. Performance could not be based on whether a particular visual feature was present or absent in the display, and search was undertaken by processing target and distractors in parallel as individual units. Thus, spatial attention was directed to the individual letters, so that the attended part of the visual field was restricted to a set of spatial regions, one for each letter; rates of processing were then attenuated, but features were bound to individual letters and not to the display as a whole. In this mode of processing, the target and distractors competed for processing capacity, but the target had a higher attentional weight than the distractors, so the target attracted more processing capacity than the distractors (cf. equation 7). By spatial and temporal proximity, the probe dot tended to be included in the same perceptual unit as the (target or distractor) letter it replaced, so the probe dot received more processing capacity when it replaced a target than a distractor. Hence, when this mode of processing was used, processing of the probe dot was faster for dots presented at target locations than for dots presented at distractor locations.

Other Applications. Logan (1996) applied CTVA to the results of a number of studies on the effects of perceptual grouping and spatial distance between items on reaction times and error rates in visual attention tasks. The studies included effects of grouping (Prinzmetal, 1981) and effects of distance between items (Cohen & Ivry, 1989) on occurrence of illusory conjunctions, effects of grouping (Banks & Prinzmetal, 1976) and effects of distance between items (Cohen & Ivry, 1991) in visual search, and effects of distance between target and distractors in the flankers task (Eriksen & Eriksen, 1974). Logan and Bundesen (1996) applied CTVA to a variety of spatial effects in partial report (e.g. Mewhort, Campbell, Marchetti, & Campbell, 1981). Providing quantitative accounts for such phenomena seems an important step forward.

NEURAL LOCALIZATION

Figure 6 is a flowchart of the algebraic operations of equations 1 and 2 of TVA.[4] Figure 6 can also be viewed as a flowchart of the corresponding operations in CTVA (equations 7 and 8). The latter view presupposes that (a) the eta values

[4]In Fig. 6, equation 1 of TVA is implemented by first multiplying $\eta(x, i)$ by β_i and then multiplying the product of $\eta(x, i)$ and β_i by $w_x/\sum_{z\in S} w_z$. Alternatively, the equation could be implemented by first multiplying $\eta(x, i)$ by $w_x/\sum_{z\in S} w_z$. and then multiplying by β_i.

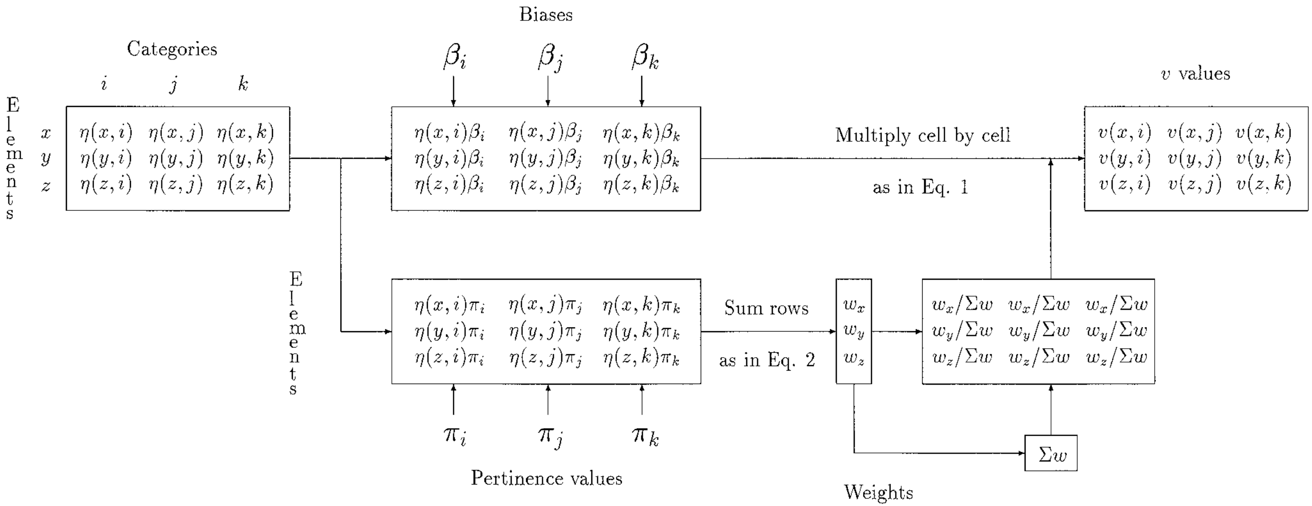

FIG. 6. Flowchart of the algebraic operations of equations 1 and 2 of TVA.

in the figure are interpreted as effective eta values [e.g. $\eta(x, i)$ is interpreted as $\eta_e(x, i)$], and (b) the set of items shown in the figure is the set of items in the field of spatial attention (i.e. $A = \{x, y, z\}$).

Figure 6 suggests a distinction between two streams of processing—a main stream of processing leading from sensory input to pattern recognition and categorization (the upper route from left to right), and an attentional loop (the lower part of the figure), in which attentional weights are computed and fed back to the main stream. Pattern recognition and categorization are often assumed to be subserved by a ventral visual pathway running from primary visual cortex (area V1) via visual areas V2 and V4 to infero-temporal (IT) cortex (see Goodale & Milner, 1992; Mishkin, Ungerleider, & Macko, 1983; Ungerleider & Mishkin, 1982). On this hypothesis, the main stream of processing in Fig. 6 should be located in the ventral pathway. The loop for attentional weighting should take input from the ventral pathway and feed back to the ventral pathway, but intermediate computations in the loop may be undertaken in dorsal rather than ventral areas of the brain. Indeed, as argued below, extant evidence suggests that the attentional weights are represented in the posterior parietal (PP) cortex (for related suggestions, see Schneider, 1995).

Electrophysiological Evidence

Results suggesting that a map of attentional weights is located in the posterior parietal cortex come from electrophysiological studies of visual responses of parietal neurons in awake, behaving monkeys. Bushnell, Goldberg and Robinson (1981) trained the same rhesus monkeys on several visual tasks. In one task (the fixation task), the monkeys were required to fixate a spot of light turned on by pressing a lever. After an unpredictable interval (1–4 sec), the fixation point dimmed and remained dim for about 500 msec. If the monkey released the lever while the light was dim, he was rewarded by receiving a drop of water. A second spot of light presented in the periphery of the visual field shortly after the onset of the fixation point was behaviourally irrelevant to the monkey. Another task (the peripheral attention task) was like the fixation task, except that on 50% of the trials the peripheral stimulus dimmed but not the fixation point. In this task, the monkey should release the lever in response to dimming of either the fixation point or the peripheral stimulus.

Responses to the peripheral stimulus were recorded from single neurons in area 7 of posterior parietal cortex. About half of the studied cells showed a significantly stronger response (i.e. greater increase in the rate of firing) to the peripheral stimulus in the peripheral attention task than in the fixation task. A similar enhancement of the visual response to the peripheral stimulus (relative to the response to the peripheral stimulus in the fixation task) was found in tasks in which the monkeys were required to make a saccadic eye movement to the peripheral stimulus or to reach out and touch the stimulus. The fact that the enhancement occurred when the monkeys behaved differently when attending to the stimulus suggested that the enhancement was involved in a general process of visual selective attention. In particular, the results seem consistent with the hypothesis that the rate of firing in a neuron with the peripheral stimulus in its receptive field represented the attentional weight of the peripheral stimulus.

Bushnell et al. (1981) noted that area 7 receives rich inputs from the limbic system (substantia innominata, claustrum and cingulate gyrus). They suggested that, whereas the activity in the limbic neurons depends on the way in which the stimulus is pertinent to the animal (the quality of the motivation involved), "area 7 may serve to integrate motivation-specific information into a general attentional mechanism" (p. 769). More specifically, the summation described in equation 2 of TVA may be done in area 7.

Neuropsychological Evidence

Posner, Walker, Friedrich and Rafal (1984) investigated effects of parietal injury on allocation of attention. Patients with unilateral damage to the parietal lobe were tested using the cost–benefit cueing paradigm developed by Posner and his associates (e.g. Posner, Nissen, & Ogden, 1978). The patients were instructed to fixate the centre of a screen where the outline of a small box was presented. Each trial began with a cue event, which was a brief brightening of one of two peripheral boxes located at 8° to the left and 8° to the right of the fixation box, respectively. At intervals varying from 0 to 1000 msec after cue onset, a bright asterisk (target) was presented in one of the peripheral boxes. On 80% of the trials, the target appeared in the cued box, so that the cue was valid. On the remaining 20% of the trials, the target appeared in the uncued peripheral box, so that the cue was invalid. In either case, the subject's task was to press a single key as quickly as possible whenever the target appeared.

The main findings of Posner et al. (1984) are illustrated in Table 1. The data in the table are mean reaction times for the group of six patients with right-hemisphere lesions on trials with a cue–target interval of 1000 msec. When the cue was valid, there was only a modest advantage of targets on the side ipsilateral to the lesion. When the cue was invalid, reaction times were longer, and there was a greater advantage of targets on the side ipsilateral to the lesion. Thus, very long reaction times were noted when the cue indicated that the target would appear on the same side as the lesion, but the target in fact appeared on the opposite side. Posner et al. (1984) explained this "extinction-like" reaction time pattern by assuming that the parietal patients had problems in "disengaging" attention from stimuli on the side ipsilateral to their lesion.

The parietal patients of Posner et al. (1984) showed no general deficit in disengaging attention. Posner et al. proposed that the patients showed a deficit in disengaging attention from particular spatial locations (viz., locations on the

TABLE 1
Observed and Predicted Mean Reaction Times for Right Parietal Patients in the Main Experiment of Posner et al. (1984)

	Side Relative to Lesion			
	Contralateral		*Ipsilateral*	
Cue Validity	*Observed*	*Predicted*	*Observed*	*Predicted*
Invalid	906	910	629	625
Valid	564	548	461	477

Note: The observed data are mean reaction times (msec) for the group of six patients with right parietal lesions on trials with a cue-target interval of 1000 msec. The means are calculated from the data presented in Table 3 of Posner et al. (1984). The predictions assume that attentional weights were reduced by 36% in the contralesional visual field

side ipsilateral to the lesion). Later work with parietal patients by Posner, Walker, Friedrich and Rafal (1987) indicated that, when both the cue and the target appeared on the side ipsilateral to the lesion, but the target was displaced from the cued location, reactions were slower when the direction of the displacement was contralesional rather than ipsilesional. Posner et al. (1987) tentatively concluded that the patients showed a deficit in disengaging attention from a spatial location when the disengagement operation should be followed by a movement of attention in the contralesional direction.

A simpler explanation of the extinction-like reaction time pattern found in parietal patients by Posner et al. (1984, 1987) is based on the assumption that the attentional weight of an element was lower, the farther the element was moved in the contralesional direction. This is illustrated by the predicted values in Table 1.

The predictions in Table 1 were based on TVA and the assumption that attentional weights of elements (target or noise) at the ipsilesional stimulus location were normal, but weights of elements at the contralesional location were reduced to a certain proportion, m, of their normal value. Following Bundesen (1990, p. 532), the ratio of the normal value of the weight of an element (target or noise) on the cued side to the normal value of the weight of an element on the uncued side was assumed to be 2:1. For the patients, therefore, the ratio of the attentional weight of a cued target on the contralesional side (*valid*, *contra*) to noise on the ipsilesional side should be $2m$:1. The weight ratio of an uncued target on the ipsilesional side (*invalid*, *ipsi*) to contralesional noise should be 1:$2m$. The weight ratio of a cued target on the ipsilesional side (*valid*, *ipsi*) to contralesional noise should be 2:m, and the weight ratio of an uncued target on the contralesional side (*invalid*, *contra*) to ipsilesional noise should be m:2. The predicted mean reaction times are given by:

$$E(\mathrm{RT} \mid \text{valid, contra}) = [C \times 2m/(2m+1)]^{-1} + b$$
$$E(\mathrm{RT} \mid \text{invalid, ipsi}) = [C \times 1/(2m+1)]^{-1} + b$$
$$E(\mathrm{RT} \mid \text{valid, ipsi}) = [C \times 2/(2+m)]^{-1} + b$$
$$E(\mathrm{RT} \mid \text{invalid, contra}) = [C \times m/(2+m)]^{-1} + b$$

where C is the overall rate of processing and b is a base reaction time.

The fit shown in Table 1 was obtained with processing rate C at 6.48 elements per second, base reaction time b at 273 msec, and parameter m at 0.64. Note that b and C act only as translation and scaling parameters. What captures the observed extinction-like reaction time pattern is the hypothesis that weights at the contralesional stimulus location were reduced by a fixed proportion. The estimated value of 0.64 for m means that attentional weights were reduced by 36% at the location contralateral to the lesion.

CONCLUSION

Results from both electrophysiological and neuropsychological studies seem to suggest that attentional weights are represented in posterior parietal cortex. The evidence seems to suggest that the attentional loop in Fig. 6 takes input from the ventral pathway and feeds back to the ventral pathway, but the loop goes through the posterior parietal cortex. As suggested by Posner and Petersen (1990), transfer of information between the ventral system and the posterior parietal cortex may occur through the thalamus (cf. Petersen, Robinson, & Morris, 1987). A rough sketch of the anatomical localization of the information flows shown in Fig. 6 is thus available. I am currently working on more detailed models of the way in which the operations described in TVA and CTVA may be implemented in the brain.

REFERENCES

Banks, W.P., & Prinzmetal, W. (1976). Configurational effects in visual information processing. *Perception and Psychophysics, 19*, 361–367.

Broadbent, D.E. (1970). Stimulus set and response set: Two kinds of selective attention. In D.I. Mostofsky (Ed.), *Attention: Contemporary theory and analysis* (pp. 51–60). New York: Appleton-Century-Crofts.

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. (1991). Visual selection of features and objects: Is location special? A reinterpretation of Nissen's (1985) findings. *Perception and Psychophysics, 50*, 87–89.

Bundesen, C. (1993a). The notion of elements in the visual field in a theory of visual attention: A reply to van der Velde and van der Heijden (1993). *Perception and Psychophysics, 53*, 350–352.

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

Bundesen, C. (1996). Formal models of visual attention: A tutorial review. In A.F. Kramer, M.G.H. Coles, & G.D. Logan (Eds), *Converging operations in the study of visual selective attention* (pp. 1–43). Washington, DC: American Psychological Association.
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.M. Marin (Eds), *Attention and performance XI* (pp. 631–649). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
Bushnell, M.C., Goldberg, M.E., & Robinson, D.L. (1981). Behavioral enhancement of visual responses in monkey cerebral cortex. I. Modulation in posterior parietal cortex related to selective visual attention. *Journal of Neurophysiology, 46*, 755–772.
Cave, K.R. (1994). Changes in the strength of spatial attention. Poster session presented at the conference on *Converging Operations in the Study of Visual Selective Attention*, University of Illinois at Urbana-Champaign, May.
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.B. (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.
Compton, B.J., & Logan, G.D. (1993). Evaluating a computational model of perceptual grouping by proximity. *Perception and Psychophysics, 53*, 403–421.
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.
Goodale, M.A., & Milner, A.D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences, 15*, 20–25.
Kim, M.-S., & Cave, K.R. (1995). Spatial attention in visual search for features and feature conjunctions. *Psychological Science, 6*, 376–380.
Logan, G.D. (1996). The CODE theory of visual attention: An integration of space-based and object-based attention. *Psychological Review, 103*, 603–649.
Logan, G.D., & Bundesen, C. (1996). Spatial effects in the partial report paradigm: A challenge for theories of visual spatial attention. In D.L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 35, pp. 243–282). San Diego, CA: Academic Press.
Luce, R.D. (1959). *Individual choice behavior*. New York: Wiley.
Luce, R.D. (1963). Detection and recognition. In R.D. Luce, R.R. Bush, & E. Galanter (Eds), *Handbook of mathematical psychology* (Vol. 1, pp. 103–189). New York: Wiley.
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.
Mishkin, M., Ungerleider, L.G., & Macko, K.A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neurosciences, 6*, 414–417.
Nissen, M.J. (1985). Accessing features and objects: Is location special? In M.I. Posner & O.S.M. Marin (Eds), *Attention and performance XI* (pp. 205–219). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
Petersen, S.E., Robinson, D.L., & Morris, J.D. (1987). Contributions of the pulvinar to visual spatial attention. *Neuropsychologia, 25*, 97–105..
Posner, M.I., Nissen, M.J., & Ogden, W.C. (1978). Attended and unattended processing modes: The role of set for spatial location. In H.L. Pick & E. Saltzman (Eds), *Modes of perceiving and processing information* (pp. 137–157). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
Posner, M.I., & Petersen, S.E. (1990). The attention system of the human brain. *Annual Review of Neuroscience, 13*, 25–42.

Posner, M.I., Walker, J.A., Friedrich, F.J., & Rafal, R.D. (1984). Effects of parietal injury on covert orienting of attention. *Journal of Neuroscience, 4*, 1863–1874.

Posner, M.I., Walker, J.A., Friedrich, F.J., & Rafal, R.D. (1987). How do the parietal lobes direct covert attention? *Neuropsychologia, 25*, 135–145.

Prinzmetal, W. (1981). Principles of feature integration in visual perception. *Perception and Psychophysics, 30*, 330–340.

Schneider, W.X. (1995). VAM: A neuro-cognitive model for visual attention control of segmentation, object recognition, and space-based motor action. *Visual Cognition, 2*, 331–375.

Shibuya, H. (1991). Comparison between stochastic models for visual selection. In J.-P. Doignon & J.-C. Falmagne (Eds), *Mathematical psychology: Current developments* (pp. 337–356). New York: Springer-Verlag.

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.

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

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

Treisman, A.M., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review, 95*, 15–48.

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

Ungerleider, L.G., & Mishkin, M. (1982). Two cortical visual systems. In D.J. Ingle, M.A. Goodale, & R.J.W. Mansfield (Eds), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT Press.

van der Heijden, A.H.C. (1993). The role of position in object selection in vision. *Psychological Research, 56*, 44–58.

van der Velde, F., & van der Heijden, A.H.C. (1993). An element in the visual field is just a conjunction of attributes: A critique of Bundesen (1991). *Perception and Psychophysics, 53*, 345–349.

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.