PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

A computational theory of visual attention

Claus Bundesen

Phil. Trans. R. Soc. Lond. B 1998 **353**, 1271-1281
doi: 10.1098/rstb.1998.0282

The Royal Society

A computational theory of visual attention

Claus Bundesen

Centre for Visual Cognition, Psychological Laboratory, University of Copenhagen, Njalsgade 90, DK-2300 Copenhagen S, Denmark
(bundesen@axp.psl.ku.dk)

A computational theory of visual attention is presented. The basic theory (TVA) combines the biased-choice model for single-stimulus recognition with the fixed-capacity independent race model (FIRM) for selection from multi-element displays. TVA organizes a large body of experimental findings on performance in visual recognition and attention tasks. A recent development (CTVA) combines TVA with a theory of perceptual grouping by proximity. CTVA explains effects of perceptual grouping and spatial distance between items in multi-element displays. A new account of spatial focusing is proposed in this paper. The account provides a framework for understanding visual search as an interplay between serial and parallel processes.

Keywords: visual perception; selectivity; psychology; attention

1. INTRODUCTION

This paper describes and further develops a computational theory of visual attention. The theory is based on a race model of selection from multi-element displays and a race model of single-stimulus recognition. In race models of selection from multi-element displays, display elements are processed in parallel, and attentional selection is made of those elements that first finish processing (the winners of the race). Thus, selection of targets (elements to be selected) instead of distractors (elements to be ignored) is based on processing of targets being faster than processing of distractors. In race models of single-stimulus recognition, alternative perceptual categorizations are processed in parallel, and the subject selects the categorization that first completes processing.

The first race models of selection from multi-element displays appeared in the 1980s (Bundesen *et al.* 1985; Bundesen 1987, 1996). The most successful among the models was the fixed-capacity independent race model (FIRM) of Shibuya & Bundesen (1988). In this model, a stimulus display is processed as follows. First an attentional 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 rather than a distractor. Then the available processing capacity is distributed across the elements in proportion to their weights. The amount of processing capacity that is allocated to an element determines how fast the element can be encoded into visual short-term memory (VSTM). Finally the encoding race between the elements takes place. The elements that are selected (i.e. stored in VSTM) are those elements whose encoding processes complete before the stimulus presentation terminates and before VSTM has been filled up.

In a generalization of FIRM called TVA (theory of visual attention; Bundesen 1990), selection depends on the outcome of a race between possible perceptual categorizations. The rate at which a possible categorization ('element *x* belongs to category *i*') is processed increases with: (i) the strength of the sensory evidence that supports the categorization; (ii) the subject's bias for assigning stimuli to category *i*; and (iii) the attentional weight of element *x*. When a possible categorization completes processing, the categorization enters VSTM if memory space is available there. The span of VSTM is limited to about four elements. Competition between mutually incompatible categorizations of the same element is resolved in favour of the first-completing categorization.

TVA accounts for many findings on single-stimulus recognition, whole report, partial report, search, and detection. Recently the theory has been extended by Gordon Logan (1996). The extended theory, CTVA (CODE theory of visual attention), combines TVA with a theory of perceptual grouping by proximity (van Oeffelen & Vos 1982). CTVA explains a wide range of spatial effects in visual attention.

The formal assumptions of TVA and CTVA are presented in the first main section of this paper (§ 2). The presentation includes a new account of spatial focusing, which provides a framework for understanding visual search as an interplay between serial and parallel processes. The following main sections of the paper treat applications of the theory to single-stimulus recognition (§ 3) and selection from multi-element displays (§ 4).

2. GENERAL THEORY

(a) *Basic TVA*

In TVA, both visual recognition and attentional selection of elements in the visual field consist in making perceptual categorizations. A perceptual categorization has the form 'element *x* has feature *i*', or equivalently, 'element *x* belongs to category *i*'. Here element *x* is an object (a perceptual unit) in the visual field, feature *i* is a perceptual feature (e.g., a certain colour, shape, movement, or spatial position), and category *i* is a perceptual category (the class of all elements that have feature *i*).

A perceptual categorization is made if and when the categorization is encoded into visual short-term memory (VSTM). When the perceptual categorization that element x belongs to category i has been made (i.e. encoded into VSTM), element x is said to be selected and element x is also said to be recognized as a member of category i. Thus, attentional selection of element x implies that x is recognized as a member of one or other category. Element x is said to be retained in VSTM if and when one or other categorization of the element is retained in VSTM.

Once a perceptual categorization of an element completes processing, the categorization enters VSTM, provided that memory space for the categorization is available in VSTM. The capacity of VSTM is limited to K different elements, where K is about 4 (cf. Sperling 1960). Space is available for a new categorization of element x, if element x is already represented in the store (with one or other categorization) or if less than K elements are represented in the store (cf. Luck & Vogel 1997). There is no room for a categorization of element x if VSTM has been filled up with other elements.

Consider the time taken to process a particular perceptual categorization, 'element x belongs to category i'. This processing time is a random variable that follows a certain distribution. In TVA, the distribution is defined by specifying the instantaneous tendency (probability density) that the processing completes at time t, given that the processing has not completed before time t. This instantaneous tendency (hazard rate) is a measure of the speed at which the perceptual categorization is processed. In TVA, the measure is called the v-value of the perceptual categorization that x belongs to i, $v(x,i)$, and $v(x,i)$ is determined by two basic equations. By equation (1),

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

where $\eta(x,i)$ is the instantaneous strength of the sensory evidence that element x belongs to category i, β_i is a perceptual decision bias associated with category i, $\mathcal{S}$ 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.

By equation (1), both perceptual decision biases and attentional weights multiply strengths of sensory evidence, but they do so in very different ways. Parameter β_i multiplies not only $\eta(x,i)$, but every η-value of which perceptual category i is the second argument. Parameter w_x (or $w_x/\Sigma_{z\in\mathcal{S}}w_z$) multiplies not only $\eta(x,i)$, but every η-value of which element x is the first argument. Thus, decision bias parameters are used for manipulating classes of v-values (processing speeds) defined in terms of perceptual categories (values of i), whereas weight parameters are used for manipulating classes of v-values defined in terms of perceptual elements (values of x). In this sense, perceptual decision biases and attentional weights are complementary parameters.

The attentional weights are derived from perceptual processing priorities. Every perceptual category is associated with a certain processing priority (pertinence value). The processing priority associated with 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

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

where $\mathcal{R}$ 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 perceptual processing priority of category j.

By equation (2), perceptual processing priorities can be used for manipulating attentional weights. The attentional weight of an element depends on the perceptual features of the element, and π_j determines the importance of feature j in setting the attentional weights of elements.

By equations (1) and (2), v-values can be expressed as functions of η-, β-, and π-values. When η-, β-, and π-values are kept constant, processing times for different perceptual categorizations are assumed to be stochastically independent.

In most applications of the theory to experimental data, η-, β-, and π-values have been assumed to be constant during the presentation of a stimulus display. When η-, β-, and π-values are constant, v-values are also constant. The v-values were defined as hazard rates, and when these are kept constant, categorization 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.

(b) *Filtering and pigeonholing*

(i) *Filtering*

Basic TVA contains two mechanisms of selection: filtering and pigeonholing (cf. Broadbent 1970). The filtering mechanism is represented by perceptual processing priorities and attentional weights derived from processing priorities. Consider how the mechanism works. Suppose one searches for something that belongs to a particular category, say, something that is red. Selection of red elements in the visual field is favoured by letting the processing priority of the class of red elements be high. For, equation (2) implies that if the processing priority (the π value) of red is increased, then the attentional weight of an element x gets an increment which is directly proportional to the strength of the sensory evidence that the element is red. Thus, if the priority of red is increased, then the attentional weights of those elements that are red increase in relation to the attentional weights of any other elements. By equation (1) this implies that the v-values for perceptual categorizations of red elements increase in relation to the v-values for perceptual categorizations of other elements. Thus, the processing of red elements is speeded up in relation to the processing of other elements so that the red ones get a higher probability of winning the processing race and becoming encoded into VSTM.

(ii) *Pigeonholing*

The pigeonholing mechanism is represented by perceptual bias parameters. Consider how the mechanism works. Suppose one wishes to categorize objects with respect to colour. One can prepare oneself for categorizing elements in the visual field with respect to colour by giving higher values to perceptual bias parameters associated with colour categories than to other perceptual bias parameters. For, equation (1) implies that if the perceptual bias parameter (the β-value) for a particular category is increased, the tendency to classify elements into that category gets stronger: the v-values for perceptual categorizations of elements as members of the category are increased, but other v-values are not affected.

(iii) *Combined filtering and pigeonholing*

Consider how filtering and pigeonholing can be combined. To be specific, consider a partial-report experiment. Let the stimulus displays consist of mixtures of red and black digits, and let the task be to report as many as possible of the red digits and ignore the black ones. A plausible strategy for doing this task is as follows. To select red rather than black elements, the processing priority of the class of red elements is set high, but other processing priorities are kept low. The effect is to speed up the processing of red elements in relation to the processing of black elements. To perceive the identity of the red digits rather than any other attributes of the elements, ten perceptual bias parameters, one for each type of digit, are set high, but other perceptual bias parameters are kept low. The effect is to speed up the processing of categorizations with respect to digit types in relation to the processing of other categorizations. The combined effect of the adjustments of priority and bias parameters is to speed up the processing of categorizations of red elements with respect to digit types in relation to the processing of any other categorizations.

(iv) *Processing priorities against decision biases*

Processing priorities (π-values) and decision biases (β-values) are different concepts. A perceptual system in which processing priorities can be varied independently of decision biases is inherently more powerful than a system in which the two are bound to covary (Bundesen 1990, pp. 525–526). For example, when the task is to report the identity of the red digits from a mixture of red and black digits, the ideal observer should set π high for red and β high for each of the ten types of digits, but π-values for the ten types of digits should be zero. When π is high for red but not for types of digits, then the attentional weights of the black digit distractors may be close to zero. But if π were high for both red and types of digits, performance should deteriorate because the black digit distractors would get appreciable attentional weights.

Consider the consequences of setting β high for red, when π is high for red. If both π and β are high for red, then any red element (relevant or irrelevant to the current task) will tend to be categorized with respect to colour and take up storage space in VSTM, regardless of whether the identity of the element has been determined. Because storage capacity is limited, this may be detrimental to performance. However, if the number of elements in the display is less than storage capacity K, then no loss should be incurred by letting β-values be high.

Basic TVA is neutral on whether all types of perceptual categories can be given positive-valued processing priorities (rather than having priority values fixed at zero). There is evidence to suggest that only a subset of the class of perceptual categories can have positive priorities. For example, both individual letters and short multiletter words are assumed to be perceptual categories, and both letters and words can be associated with positive biases (β-values). Furthermore, demonstrations of automatic attention attraction to particular types of individual letters (after extended consistent training in detecting these letters; cf. Shiffrin & Schneider 1977) suggest that individual letter types can be associated with positive processing priorities. However, a recent study by Bundesen *et al.* (1997) suggests that the initial allocation of attention to items in a visual display may be insensitive to words.

Bundesen *et al.* (1997) presented subjects with briefly exposed visual displays of words, which were short, common first names. In the main experiment, each display consisted of four words: two names shown in red and two shown in white. The subject's task was to report the red names (targets), but ignore the white ones (distractors). In some trials the subject's own name appeared as a display item (target or distractor). Presentation of the subject's name as a distractor caused no more interference with report of targets than did presentation of other names as distractors. Apparently, visual attention was not automatically attracted by the subject's own name.

If priority learning could occur for visual words, so that a visual word could attract attention automatically, one would expect a subject's attention to be attracted automatically by his or her own name (cf. Moray 1959). The contrast between findings with single letters and digits and findings with multiletter words suggests that visual attention can be attracted by individual alphanumeric characters, but not by shapes as complex as multiletter words. Multiletter words may be too complex in shape to have positive processing priorities.

(c) *CTVA*

Logan (1996) has proposed a theory that integrates space-based and object-based approaches to visual attention (Logan & Bundesen 1996; Bundesen 1998). The theory was made by linking TVA together with van Oeffelen & Vos' (1982, 1983) COntour DEtector (CODE) theory of perceptual grouping by proximity. The integrated theory is called the CODE theory of visual attention (CTVA).

(i) *Perceptual grouping*

In the theory of van Oeffelen & Vos (1982, 1983), grouping by proximity is modelled as follows (see figure 1). First, each stimulus item is represented by a distribution centred on the position that the object occupies in one- or two-dimensional space. Van Oeffelen & Vos originally used normal distributions, but Compton & Logan (1993) found that Laplace distributions worked just as well. Thus, in the one-dimensional case (e.g. a linear array of items positioned along a u-axis), item y may be represented by the Laplace distribution

$$f_y(u) = \frac{1}{2}\lambda_y \exp\left(-\lambda_y|u-\theta_y|\right), \qquad (3)$$

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, and a

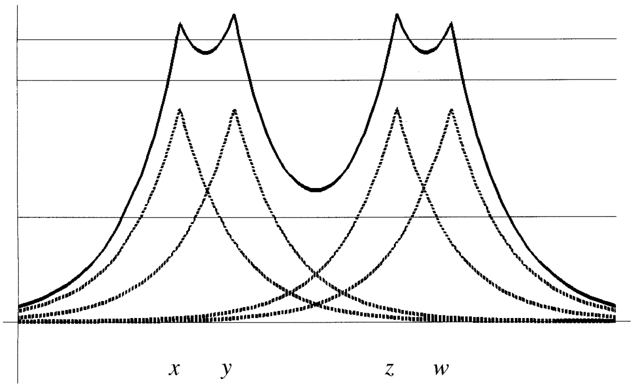

Figure 1. Perceptual grouping by proximity. Laplace distributions (broken curves) and a CODE surface (solid curve) are shown for four items (x, y, z, and w) arrayed in one dimension. Thresholds applied to the CODE surface are shown by crossing horizontal lines. The low threshold includes all four items in one group. The middle threshold generates two groups with two items in each. The high threshold separates all four items.

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. In terms of TVA, a perceptual group is the same as an element in the visual field.

Groups of different sizes can be defined 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. The smaller groups are nested within the larger groups, so the grouping is hierarchical.

(ii) *Spatial focusing*

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 (3), $f_y(u)$ 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 and no information at all from the remaining regions. Here I propose a revision of this assumption.

At any point in time, there is a certain set of elements (above-threshold regions) in the visual field that forms the focus of attention, $\mathcal{F}$. The focus of attention is also called the field of spatial attention (cf. Logan & Bundesen 1996). Processing of elements in the focus of attention is faster than processing of elements outside the focus of attention, because effective η-values for elements in the focus of attention are greater than effective η-values for elements outside the focus of attention. Formally the effect of attentional focusing at $\mathcal{F}$ is to multiply η-values for any element x by an attenuation factor $a_{\mathcal{F}}(x)$ such that

$$a_{\mathcal{F}}(x) = \begin{cases} 1 & \text{if} \quad x \in \mathcal{F} \\ k & \text{if} \quad x \notin \mathcal{F}, \end{cases} \tag{4}$$

where $0 \leqslant k < 1$. If $a_{\mathcal{F}}(x) = 1$, processing of x is said to be unattenuated.

Spatial focusing of attention is assumed to be constrained as follows. First, the focus of attention, $\mathcal{F}$, can be widened to encompass all elements in the visual field. That is, $\mathcal{F}$ can be set equal to $\mathcal{S}$.

Second, the focus of attention, $\mathcal{F}$, can be restricted to any element x found in VSTM. If $\mathcal{F}$ is restricted to x, and x is a group with several members, then the members of x are processed in parallel. Thus, when the focus of attention is directed to a particular perceptual group, a parallel search through the group is performed, and if focusing is strong (so that $a_{\mathcal{F}}(x) \approx 0$ for $x \notin \mathcal{F}$), then the search may occur without any noticeable effects of elements outside the focus of attention.

Finally, if a perceptual group x is found in VSTM, and element y is a member of the group, then the focus of attention, $\mathcal{F}$, can be narrowed down to element y. Thus, a serial search through a perceptual group x represented in VSTM can be performed by shifting $\mathcal{F}$ around among the members of the group. If the members of x themselves are perceptual groups with several members, then the search through x consists in a series of parallel searches through subsets of x.

(iii) *Feature catch*

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 figure 2). It equals the area or volume of the distribution for the item that falls within the limits of the above-threshold region. The feature catch is positive for all items in the display, but it decreases as the spatial distance of the item from the given region is increased.

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. 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,

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

where $f_y(u)$ is given by equation (3), and the integration is done over the above-threshold region formed by item x (cf. figure 2).

(iv) *Effective η-values*

Both spatial focusing of attention and feature catch relations in the display modulate the information input to TVA. Formally this is represented by replacing η-values $\eta(x,i)$ by effective η-values $\eta_e(x,i)$ in equations (1) and (2) of TVA. The effective η-value for the categorization that item x is a member of category i (i.e. x has feature i) is given by

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

where $\mathcal{S}$ is the set of all items in the display, and $a_{\mathcal{F}}(x)$ and $c(x,y)$ are given by equations (4) and (5), respectively. By the summation in equation (6), the effective evidence that item x has feature i depends upon the evidence that item y

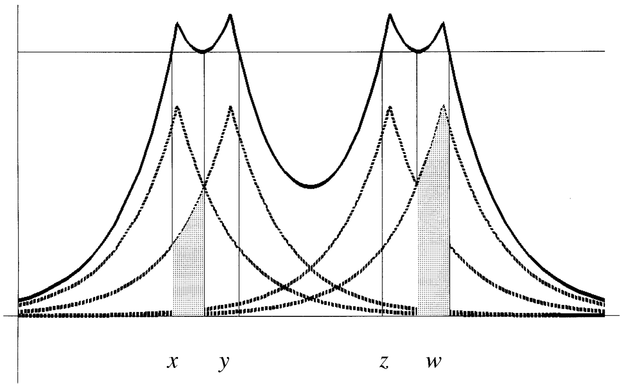

Figure 2. Feature catch. Laplace distributions (broken curves) and a CODE surface (solid curve) are shown for items x, y, z, and w. A threshold (horizontal line) applied to the CODE surface generates four above-threshold regions (separated by vertical lines). The feature catch from item y in the above-threshold 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.

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 equations (1) and (2) of TVA yields the CTVA equations

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

and

$$w_x = \sum_{j \in R} \eta_e(x,j)\pi_j. \qquad (2')$$

Thus, CTVA becomes identical to TVA when $\eta_e(x,i) = \eta(x,i)$ for every element x and every perceptual category i. This is the case when: (i) $\mathcal{F} = S$ (i.e. the focus of attention coincides with the set of items in the display); and (ii) $c(x,x) = 1$ for every item x, but $c(x,y) = 0$ when x is different from y. For example, in many partial-report experiments, it seems plausible that: (i) the focus of attention encompasses all items in the display; and (ii) interitem distances are so long that feature catches from adjacent items can be neglected. In such cases, an analysis based on CTVA reduces to an analysis based on TVA. Thus, CTVA can be viewed as a generalization of TVA, and TVA can be viewed as a special case of CTVA.

3. APPLICATIONS. I. SINGLE-STIMULUS RECOGNITION

(a) *Biased-choice model*

TVA has been applied to experimental findings from a broad range of paradigms concerned with single-stimulus recognition and selection from multi-element displays. For single-stimulus recognition, the theory provides a simple derivation of a classical model of effects of visual discriminability and bias: the biased-choice model of Luce (1963).

Consider a single-stimulus recognition experiment with n distinct stimuli and n appropriate responses, one for each stimulus. In each trial, one of the n stimuli is exposed, and the subject attempts to identify the stimulus by giving the appropriate response. The presentation of the stimulus continues until the subject responds. With a single element x in the visual field, equation (1) implies that for every perceptual category i,

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

Assume that η- and β-values are constant during the period of stimulus exposure. Then the processing time of the perceptual categorization that x belongs to i is exponentially distributed with a rate parameter equal to the v-value, $v(x,i)$. Suppose the subject's choice among the n responses is based on the winner of the processing race between n corresponding perceptual categorizations, one for each response. Then the probability that the subject chooses the ith response can be written and rewritten as follows:

$$\begin{aligned} P &= \int_0^\infty v(x,i)\exp[-v(x,i)t] \prod_{\substack{j=1 \\ j \neq i}}^{n} \exp[-v(x,j)t]\,\mathrm{d}t \\ &= \int_0^\infty v(x,i)\exp\left[-\sum_{j=1}^{n} v(x,j)t\right]\mathrm{d}t \\ &= \frac{v(x,i)}{\sum_{j=1}^{n} v(x,j)} \\ &= \frac{\eta(x,i)\beta_i}{\sum_{j=1}^{n} \eta(x,j)\beta_j}. \end{aligned}$$

The last expression for P is identical to the basic representation of choice probabilities in the biased-choice model of Luce (1963).

The biased-choice model has been successful in explaining many experimental findings on effects of visual discriminability and bias in single-stimulus recognition. For example, in a thorough test of ten mathematical models of visual letter recognition against data from a letter confusion experiment, Townsend & Ashby (1982) found that the biased-choice model consistently provided the best fits.

(b) *Processing time distributions*

The derivation of the biased-choice model presented here presupposes that v-values are constant during stimulus exposure, which means that processing times are exponentially distributed. The biased-choice model can also be derived on the weaker assumption that the v-values are mutually proportional functions of time (cf. Bundesen 1990, footnote 4; Bundesen 1993). However, the available evidence suggests that the strong assumption that v-values are constant during stimulus exposure is true to a first approximation.

To test the assumption that v-values are constant over time, Lisbeth Harms and I investigated single-letter recognition as a function of the exposure duration of the stimulus. Our subjects were presented with one stimulus letter (a randomly chosen consonant) on each trial. The letter appeared at one of 12 equiprobable positions that were equally spaced around the circumference of an imaginary circle centred on fixation. Exposure duration was varied from 10 ms up to 200 ms, and the stimulus was followed by

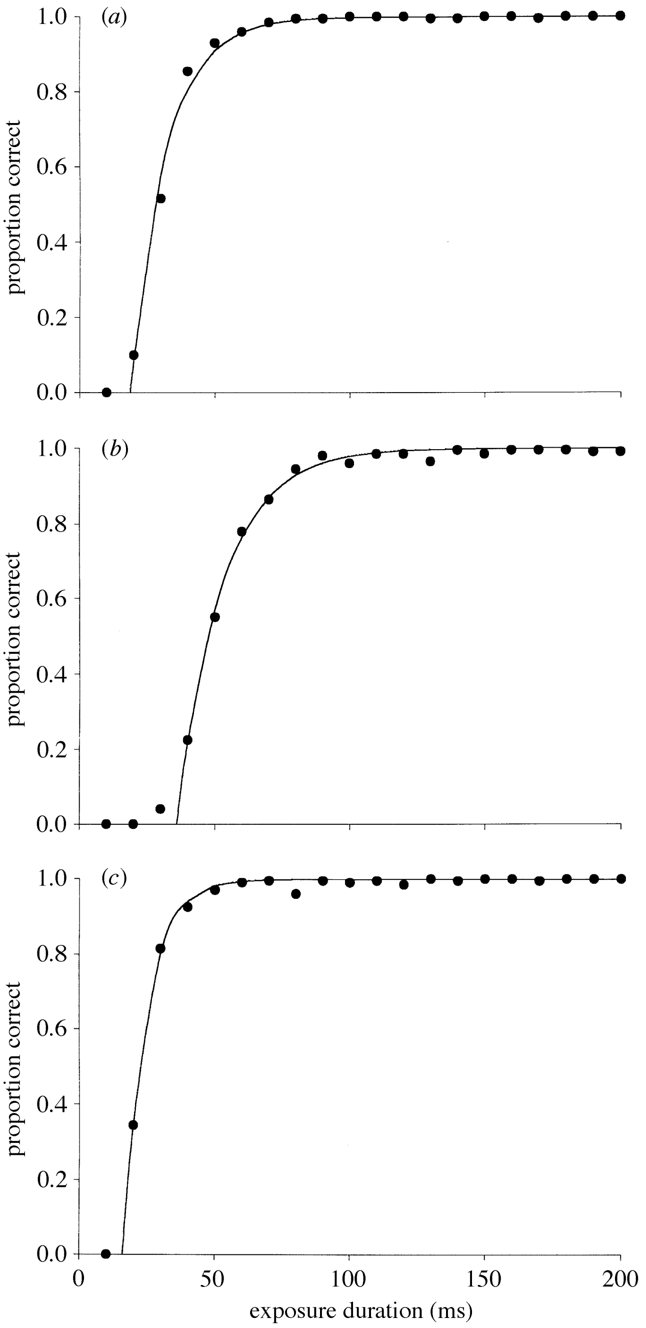

Figure 3. Proportion of correct reports of the identity of a single, postmasked stimulus letter as a function of the exposure duration of the letter. (Individual data for three subjects: subjects EA (*a*), MK (*b*), and AO (*c*). Theoretical fits are indicated by smooth curves.)

a pattern mask. The subject's task was to report the identity of the stimulus letter, but refrain from guessing.

Figure 3 shows the observed proportion of correct reports as a function of the exposure duration of the stimulus letter for each of the three subjects. Smooth curves show least squares fits to the data by the exponential distribution function

$$F(t) = \begin{cases} 0 \text{ for } t < t_0 \\ 1 - \exp[-v * (t - t_0)] \text{ for } t \geqslant t_0, \end{cases}$$

where $F(t)$ is the probability that the stimulus is correctly identified as a function of exposure duration t, parameter v is the constant v-value of the correct categorization of the stimulus, and parameter t_0 is the minimum effective exposure duration. As can be seen, the exponential distribution function provided reasonable approximations to the data.

4. APPLICATIONS. II. SELECTION FROM MULTI-ELEMENT DISPLAYS

(a) *Applications of TVA*

Bundesen (1990) applied TVA to experimental findings from a broad range of paradigms stemming from a number of different research traditions. The findings included effects of object integrality in selective report (see, for example, Duncan 1984), number and spatial position of targets in studies of divided attention (Sperling 1960, 1967; Posner *et al.* 1978; van der Heijden *et al.* 1983), selection criterion and number of distractors in studies of focused attention (Estes & Taylor 1964; Treisman & Gelade 1980; Treisman & Gormican 1988), joint effects of numbers of targets and distractors in partial report (Bundesen *et al.* 1984, 1985; Shibuya & Bundesen 1988), and consistent practice in search (Schneider & Fisk 1982). We describe two of these applications here.

(i) *Partial report*

Shibuya & Bundesen's (1988) fixed-capacity independent race model (FIRM) for selection from multi-element displays can be derived as a special case of TVA. Basically, the notion of a fixed processing capacity (C) can be derived from the normalization of attentional weights assumed in equation (1) (see Bundesen 1990, pp. 524–525). The remaining parameters of FIRM are the storage capacity of VSTM (K), the ratio between the attentional weight of a distractor and the attentional weight of a target (α), and the minimum effective exposure duration (t_0).

Although FIRM has only four free parameters (C, K, α, and t_0), the model has provided highly accurate accounts of effects of the number of targets, the number of distractors, and the exposure duration on the number of targets that can be reported from briefly presented displays. To illustrate, figure 4 shows a fit to observed frequency distributions of scores for a subject tested by Shibuya & Bundesen (1988). The subject was required to report as many digits as possible from briefly presented mixtures of digits (targets) and letters (distractors) followed by pattern masks. Let F_j ($j = 1, 2, \ldots$) be the relative frequency of scores of j or more (correctly reported targets). Each panel in the figure shows F_1, F_2, and so on, as functions of exposure duration for a given combination of number of targets T and number of distractors D. Hence, within each panel, the distance in the direction of the ordinate between 1 and F_1 equals the relative frequency of scores of exactly 0, the distance between F_1 and F_2 equals the frequency of scores of exactly 1, and so on. The theoretical fit is shown by smooth curves, which were generated by FIRM with processing capacity C at 49 elements per second, storage capacity K at 3.7 elements, weight ratio α at 0.40, and minimum effective exposure duration t_0 at 19 ms.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

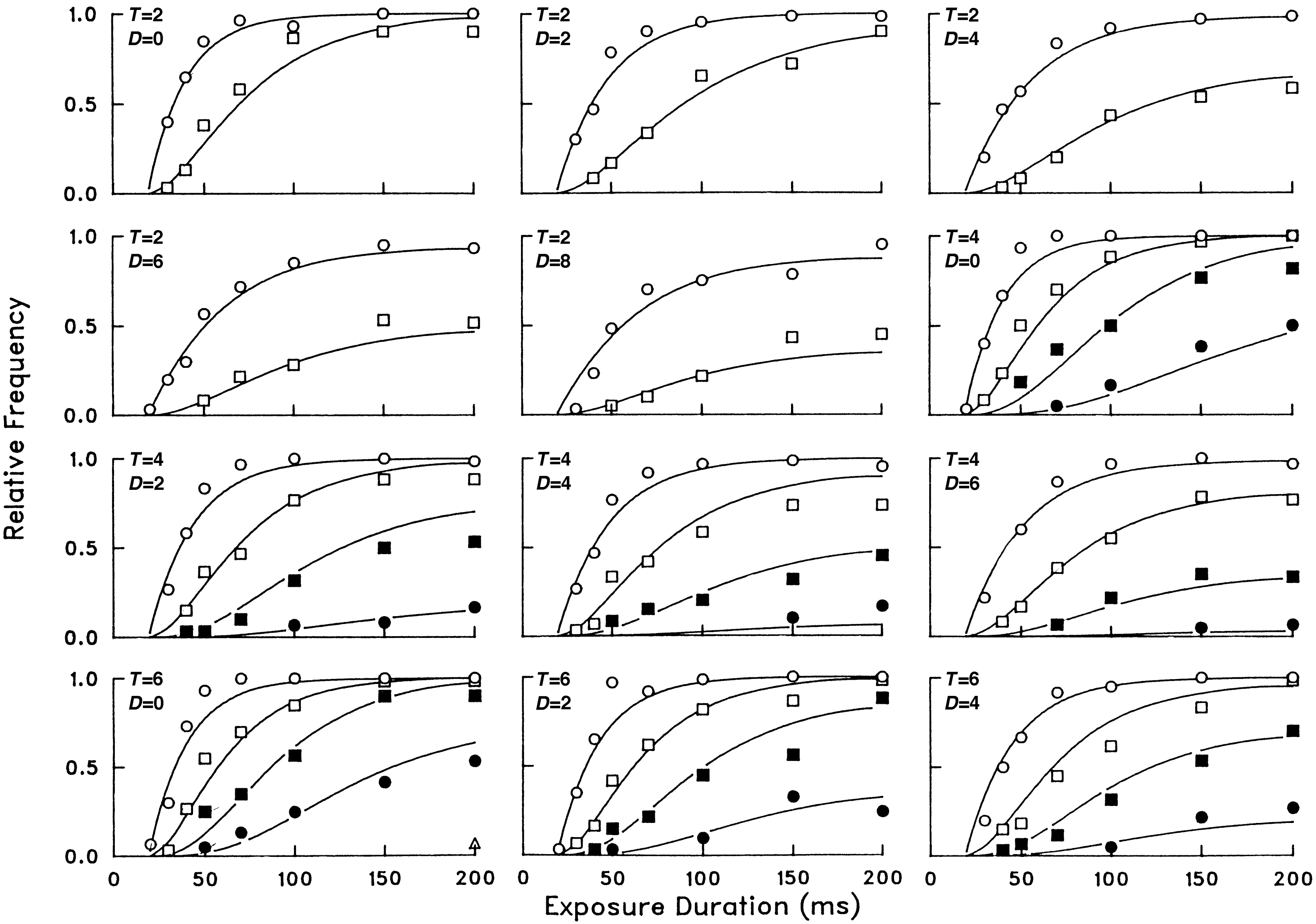

Figure 4. Relative frequency of scores of *j* or more (correctly reported targets) as a function of exposure duration with *j*, number of targets *T*, and number of distractors *D* as parameters in the experiment of Shibuya & Bundesen (1988). (Data for subject M.P. Parameter *j* varies within panels; *j* is 1 (open circles), 2 (open squares), 3 (solid squares), 4 (solid circles), or 5 (triangle). *T* and *D* vary between panels; their values are indicated on the figure. Smooth curves represent a theoretical fit to the data. For clarity, observed frequencies less than 0.02 are omitted from the figure. From Shibuya & Bundesen (1988, p. 595). Copyright 1988 by the American Psychological Association.)

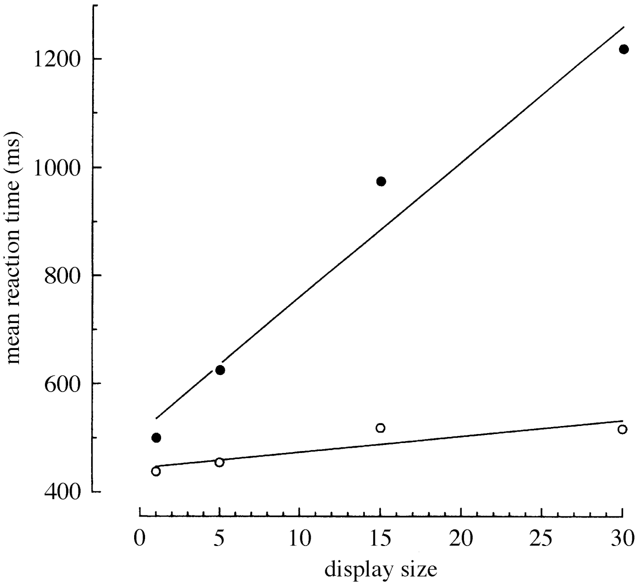

Figure 5. Positive and negative mean reaction times as functions of display size in feature search condition of Treisman & Gelade (1980, experiment 1). (Group data for six subjects. Positive reaction times are shown by open circles, negative reaction times by solid circles. A theoretical fit is indicated by unmarked points connected with straight lines. The observed data are from Treisman & Gelade (1980, p. 104). The figure is from Bundesen (1990, p. 535). Copyright 1990 by the American Psychological Association.)

(ii) *One-view search*

Figure 5 illustrates an application of TVA to a case of highly efficient visual search studied by Treisman & Gelade (1980, experiment 1, feature search condition). In this case, subjects searched for a target that was equally likely to be a blue element (a blue T or a blue X) or an S (a brown S or a green S). The distractors were brown Ts and green X s, and the display was exposed until a positive ('target present') or negative ('target absent') response was made.

The reaction time data in figure 5 are fitted by two straight lines, one for positive and one for negative responses. The fit was made on the assumption that positive responses were based on positive categorizations, whereas negative responses were made by default when a temporal deadline *d* was reached, but no positive categorization had been made (deadline model of one-view search). A positive categorization was assumed to be a categorization of the form '*x* is blue' or '*x* is an S', and processing priorities (π-values) and decision biases (β-values) were assumed to be high for blue and S, but low for any other perceptual categories.

For any deadline *d*, there is a certain probability *r* of missing a target, because the deadline may be reached before a positive categorization has been made even if a target is present in the display. Consistent with error rates observed by Treisman & Gelade, the deadline *d* was assumed to increase with display size in such a way that the miss rate *r* was kept constant.

The assumptions left four free parameters: *r*, the ratio α/C, a positive base reaction time *a*, and a negative base reaction time *b* (cf. Bundesen 1990, pp. 534–535). The least squares fit shown in figure 5 was obtained with *r* at 0.0002, α/C at 2.93 ms, *a* at 448 ms, and *b* at 536 ms. The estimate for α/C seems plausible; it is consistent with a hypothesis that, say, $C = 49$ elements per second (as in the fit shown in figure 4) and $\alpha = 0.14$.

(b) *Applications of CTVA*

(i) *Spatial effects*

Logan (1996) applied CTVA to many findings of effects of perceptual grouping and spatial distance between items on reaction times and error rates in visual attention tasks. The findings 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). Effects of distance between items on occurrence of illusory conjunctions and effects of distance in the flankers task were explained by the assumption that the feature catch factor for a particular item in a region around another one increases if the distance between the two items is decreased. The finding that conjunction search is slowed down when distances between items are decreased was explained by assuming that the threshold applied to the CODE surface is raised to prevent formation of illusory conjunctions when distances between items are decreased.

Logan & Bundesen (1996) reanalysed the data of Mewhort *et al.* (1981) on location errors in the bar-probe partial-report task introduced by Averbach & Coriell (1961). In this task, the subject is presented with an array of items and instructed to report a single one, which is the item adjacent to a bar marker (probe). A response is required on each trial. When the bar probe is presented at various delays relative to the array, decay functions similar to those observed in the partial-report paradigm of Sperling (1960) are observed.

Mewhort *et al.* (1981) distinguished correct reports from two types of errors: location errors in which the subject reports a distractor that has been presented in the array and item errors in which the subject reports an item that has not been presented in the array. In the standard Averbach & Coriell condition, Mewhort *et al.* found a nearly perfect trade-off between correct reports and location errors. As probe delay increased, the frequency of correct reports decreased, but the frequency of location errors increased in a compensatory fashion. Thus, the frequency of item errors remained nearly constant over probe delays. Mewhort *et al.* also analysed the spatial distribution of location errors and found that they primarily came from items immediately adjacent to the target. These results led Mewhort *et al.* to conclude that decay in sensory memory (iconic memory; Neisser 1967) after the offset of a stimulus display is a decay of location information rather than item information.

According to the reanalysis by Logan & Bundesen (1996), attention is spread evenly over the stimulus array until the bar probe is presented (i.e. all items in the array have the same attentional weight). When attention is reallocated in response to the probe, all attentional weight is concentrated on the target (i.e. attentional weights of distractors are set to zero). Processing of the target is speeded up once attention is concentrated on the target, so the frequency of correct reports is inversely related to probe delay. Also, the longer the time that the array is processed with equal attention to each item and the shorter the time the array is processed with attention concentrated on the target, the greater the probability that VSTM will contain distractors from the array without containing the target. Hence, assuming that distractor items in VSTM are reported with greater probability than items not in VSTM, the frequency of location errors must increase with probe delay. Thus, the trade-off between correct reports and location errors found by Mewhort *et al*. is perfectly compatible with the traditional assumption that sensory memory decay reflects the loss of item information rather than the loss of location information proposed by Mewhort *et al*.

The finding that location errors primarily come from items immediately adjacent to the target was also explained by CTVA. The finding is predicted by the assumption that attention is focused on an above-threshold region of the CODE surface at the location of the target, where feature catch factors are particularly high for items adjacent to the target.

Logan & Bundesen (1996) presented detailed quantitative fits of CTVA to the data of Mewhort *et al*. (1981). Other spatial effects in the partial-report paradigm were explained at a qualitative level. These effects included Snyder's (1972) finding that errors in a single-target partial-report task with selection based on colour were likely to be correct reports of items adjacent to the target; Fryklund's (1975) finding that performance in a multitarget partial-report task was better when the targets were clustered together than when they were spread at random throughout the display; and Merikle's (1980) finding that performance in a multitarget partial-report task was better when the targets formed a row than a column if the display was organized (by proximity) as a set of rows, whereas performance was better when the targets formed a column than a row if the display was organized as a set of columns. The findings of Fryklund (1975) and Merikle (1980) were explained by noting that in CTVA, intrusions from near neighbours on the feature catch of an attended target tend to generate correct reports when the near neighbours are targets, but errors when the near neighbours are distractors.

(ii) *Many-view search*

Many experiments on visual search have yielded positive and negative mean reaction times that are essentially linear functions of display size with a positive-to-negative slope ratio of about 1:2 (cf. Grossberg *et al*. 1994). For example, in experiment 1 of Treisman & Gelade (1980), conjunction search for a green T among brown Ts and green X s generated a positive reaction time function with a slope constant of 29 ms per item, a negative reaction time function with a slope constant of 67 ms per item, and a slope ratio of 0.43. In experiment 2 of Treisman & Gelade (1980), search for a red O among green O s and red N s generated a positive reaction time function with a slope constant of 21 ms per item, a negative function with a slope constant of 40 ms per item, and a slope ratio of 0.52. Nearly the same positive and negative slope constants and a slope ratio of 0.53 were found by Treisman & Gormican (1988) as means across 37 conditions of feature search with low target–distractor discriminability.

Wolfe (1994) and his associates examined 708 sets of positive (target present) and negative (target absent) search slopes from subjects tested on a wide variety of different search tasks in their laboratory. Among these 708 sets, 167 had positive slopes greater than 20 ms per item. This subset showed a (harmonic) mean positive-to-negative slope ratio of 0.50. Another 187 had positive slopes less than 5 ms per item. For this subset, the (harmonic) mean positive-to-negative slope ratio was 0.53.

The pattern of approximately linear reaction time functions with positive-to-negative slope ratios of about 1:2 suggests search with (overt or covert) reallocation of attention (many-view search). The pattern conforms to predictions from simple self-terminating serial models in which attention is shifted from element to element until a target has been found (respond present) or the display has been searched exhaustively, but no target has been found (respond absent; cf. Treisman & Gelade 1980). The pattern also conforms to predictions from the assumption that attention is shifted among groups (subsets) of elements in the display so that processing is parallel within groups but serial between groups, and shifting is random (blind) with respect to the distinction between target and distractors (cf. Pashler 1987; Treisman & Gormican 1988; also see Bundesen & Pedersen 1983; Duncan & Humphreys 1989; Treisman 1982). (The guided search model of Wolfe *et al*. (1989) and Cave & Wolfe (1990) predicts slow serial search with a 1:2 slope ratio when activations caused by targets and distractors are identically distributed so that the serial order in which items are scanned is independent of their status as targets compared with distractors. However, when target activations are stronger than distractor activations, search is guided by the activations so that any target in the stimulus display is likely to be among the first items that are scanned. Thus, when search becomes guided, both the positive search slope and the positive-to-negative slope ratio should decrease. The results of Wolfe's (1994) study of 708 sets of search slopes went counter to this prediction. To accommodate the results, Wolfe (1994) suggested a modification of the guided search model based on the assumption that as signal strength increases, the mean of the distribution of target activations increases, but the standard deviation of the distribution decreases (an inverted Weber's law).)

In TVA, reallocation of attention is assumed to be slow, but serial search through a display should occur when the time cost of shifting (reallocating) attention is outweighed by gain in speed of processing once attention has been shifted (cf. Bundesen 1990, pp. 536–537). Serial search is based on selection by location. Specifically, serial search is performed by first using a spatial selection criterion for sampling elements from one part of the display, then (with or without eye movements) shifting the selection criterion to sample elements from another part of the display, and so on, until a target has been found or the entire display has been searched exhaustively.

CTVA also assumes that serial search is based on selection by location, but in CTVA selection by location is 'special' (cf. Nissen 1985; also see Bundesen 1991). Selection by criteria other than location must be done by filtering, that is, by raising the processing priority (say, π_j) of the class of elements to be selected. By equations (2) and (2′), the attentional weight of an element is a sum of weight components, one for each perceptual category, and the effect of increasing π_j is to increment the weight component that corresponds to category *j*. The summation (addition) of weight components permits efficient search for feature disjunctions (e.g. search for elements that are blue or S-shaped; Treisman & Gelade 1980, experiment 1), but not for conjunctions.

By contrast, selection by location can be done by spatial focusing, that is, by restricting the focus of attention $\mathcal{F}$ to a perceptual group at the target location. The effect is to attenuate effective η-values for elements outside the target location by multiplication with a factor $k<1$. If $k \approx 0$, and if feature catches from elements outside the target location are negligible, then a parallel search through the members of the perceptual group at the target location should be just as efficient as it would have been if no elements had been presented outside the target location.

Thus, in CTVA, a parallel search for a target defined by a feature, *f*, can be restricted to a perceptual group at a certain location with no loss in efficiency. If the perceptual group is the set of elements with feature *g*, then the process as a whole is a search for a feature conjunction of *f* and *g* (for examples, see Egeth *et al*. 1984; Kaptein *et al*. 1995).

If target–distractor discriminability is high with respect to feature *f*, and the processing priority (π-value) is high for feature *f*, and only for feature *f*, then the distractor-to-target weight ratio α must be low. In this case, the perceptual group can be rapidly searched for an element with feature *f* by processing the group in parallel in accordance with the deadline model of one-view search. When processing is done in accordance with the deadline model, the time taken to process the perceptual group varies directly with weight ratio α (cf. Bundesen 1990, pp. 534–535). In the limiting case in which $\alpha = 0$, the search time is independent of the number of elements in the search set. Hence, if feature *g* defines a strong perceptual group, and detection of feature *f* is easy ($\alpha \approx 0$), then search for a conjunction of *f* and *g* should show little effect of display size (for examples of fast conjunction search, see Nakayama & Silverman 1986; Wolfe *et al*. 1989).

Our considerations of the implications of CTVA suggest a simple explanation for the frequently observed pattern of positive and negative search reaction times that are essentially linear functions of display size with a wide range of slopes but positive-to-negative slope ratios of 1:2. Such search functions can be explained by assuming that the stimulus displays are processed by shifting the focus of attention $\mathcal{F}$ among groups of elements so that processing is parallel within groups but serial between groups. The parallel processing of members of the same group can be done in accordance with the deadline model of one-view search, so that the time taken to process a group varies with target–distractor discriminability. But the shifting among groups is random (blind) with respect to the distinction between target and distractors, and it is this randomness that generates the 1:2 slope ratios.

5. CONCLUDING REMARKS

TVA (Bundesen 1990) provides a unified account of single-stimulus recognition and selection from multi-element displays. It integrates the biased-choice model for single-stimulus recognition (Luce 1963) with the fixed-capacity independent race model for selection from multi-element displays (Shibuya & Bundesen 1988). Mathematically the theory is tractable, and it organizes a large body of experimental findings on performance in visual recognition and attention tasks. CTVA (Logan 1996) combines TVA with a theory of perceptual grouping by proximity (van Oeffelen & Vos 1982). The combined theory explains a wide range of effects of perceptual grouping and spatial distance between items on performance in attention tasks. It also provides a useful framework for describing visual search as an interplay between serial and parallel processes.

The reported research was supported by grants from the Danish Ministry of Education and Research and the International Human Frontier Science Program Organization. I thank Jon Driver and John Duncan for constructive comments on the draft.

REFERENCES

Averbach, E. & Coriell, A. S. 1961 Short-term memory in vision. *Bell Syst. Tech. J.* **40**, 309–328.

Banks, W. P. & Prinzmetal, W. 1976 Configurational effects in visual information processing. *Percept. Psychophys.* **19**, 361–367.

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

Bundesen, C. 1987 Visual attention: race models for selection from multi-element displays. *Psychol. Res.* **49**, 113–121.

Bundesen, C. 1990 A theory of visual attention. *Psychol. Rev.* **97**, 523–547.

Bundesen, C. 1991 Visual selection of features and objects: is location special? A reinterpretation of Nissen's (1985) findings. *Percept. Psychophys.* **50**, 87–89.

Bundesen, C. 1993 The relationship between independent race models and Luce's choice axiom. *J. Math. Psychol.* **37**, 446–471.

Bundesen, C. 1996 Formal models of visual attention: a tutorial review. In *Converging operations in the study of visual selective attention* (ed. A. F. Kramer, M. G. H. Coles & G. D. Logan), pp. 1–43. Washington, DC: American Psychological Association.

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

Bundesen, C. & Pedersen, L. F. 1983 Color segregation and visual search. *Percept. Psychophys.* **33**, 487–493.

Bundesen, C., Pedersen, L. F. & Larsen, A. 1984 Measuring efficiency of selection from briefly exposed visual displays: a model for partial report. *J. Exp. Psychol. Hum. Percept. Perf.* **10**, 329–339.

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

Bundesen, C., Kyllingsbæk, S., Houmann, K. J. & Jensen, R. M. 1997 Is visual attention automatically attracted by one's own name? *Percept. Psychophys.* **59**, 714–720.

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

Cohen, A. & Ivry, R. 1989 Illusory conjunctions inside and outside the focus of attention. *J. Exp. Psychol. Hum. Percept. Perf.* **15**, 650–663.

Cohen, A. & Ivry, R. B. 1991 Density effects in conjunction search: evidence for a coarse location mechanism of feature integration. *J. Exp. Psychol. Hum. Percept. Perf.* **17**, 891–901.

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

Duncan, J. 1984 Selective attention and the organization of visual information. *J. Exp. Psychol. Gen.* **113**, 501–517.

Duncan, J. & Humphreys, G. W. 1989 Visual search and stimulus similarity. *Psychol. Rev.* **96**, 433–458.

Egeth, H. E., Virzi, R. A. & Garbart, H. 1984 Searching for conjunctively defined targets. *J. Exp. Psychol. Hum. Percept. Perf.* **10**, 32–39.

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

Estes, W. K. & Taylor, H. A. 1964 A detection method and probabilistic models for assessing information processing from brief visual displays. *Proc. Natn. Acad. Sci. USA*, **52**, 446–454.

Fryklund, I. 1975 Effects of cued-set spatial arrangement and target-background similarity in the partial-report paradigm. *Percept. Psychophys.* **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. *Psychol. Rev.* **101**, 470–489.

Kaptein, N. A., Theeuwes, J. & van der Heijden, A. H. C. 1995 Search for a conjunctively defined target can be selectively limited to a color-defined subset of elements. *J. Exp. Psychol. Hum. Percept. Perf.* **21**, 1053–1069.

Logan, G. D. 1996 The CODE theory of visual attention: an integration of space-based and object-based attention. *Psychol. Rev.* **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 *The psychology of learning and motivation*, vol. 35 (ed. D. L. Medin), pp. 243–282. San Diego, CA: Academic Press.

Luce, R. D. 1963 Detection and recognition. In *Handbook of mathematical psychology*, vol. 1 (ed. R. D. Luce, R. R. Bush & E. Galanter), pp. 103–189. New York: Wiley.

Luck, S. J. & Vogel, E. K. 1997 The capacity of visual working memory for features and conjunctions. *Nature* **390**, 279–281.

Merikle, P. M. 1980 Selection from visual persistence by perceptual groups and category membership. *J. Exp. Psychol. Gen.* **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. *Mem. Cogn.* **9**, 50–67.

Moray, N. 1959 Attention in dichotic listening: affective cues and the influence of instructions. *Q. J. Exp. Psychol.* **11**, 56–60.

Nakayama, K. & Silverman, G. H. 1986 Serial and parallel processing of visual feature conjunctions. *Nature* **320**, 264–265.

Neisser, U. 1967 *Cognitive psychology.* New York: Appleton-Century-Crofts.

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

Pashler, H. 1987 Detecting conjunctions of color and form: reassessing the serial search hypothesis. *Percept. Psychophys.* **41**, 191–201.

Posner, M. I., Nissen, M. J. & Ogden, W. C. 1978 Attended and unattended processing modes: the role of set for spatial location. In *Modes of perceiving and processing information* (ed. H. L. Pick & E. Saltzman), pp. 137–157. Hillsdale, NJ: Lawrence Erlbaum.

Prinzmetal, W. 1981 Principles of feature integration in visual perception. *Percept. Psychophys.* **30**, 330–340.

Schneider, W. & Fisk, A. D. 1982 Degree of consistent training: improvements in search performance and automatic process development. *Percept. Psychophys.* **31**, 160–168.

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

Shiffrin, R. M. & Schneider, W. 1977 Controlled and automatic human information processing. II. Perceptual learning, automatic attending, and a general theory. *Psychol. Rev.* **84**, 127–190.

Snyder, C. R. R. 1972 Selection, inspection, and naming in visual search. *J. Exp. Psychol.* **92**, 428–431.

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

Sperling, G. 1967 Successive approximations to a model for short-term memory. *Acta Psychol.* **27**, 285–292.

Townsend, J. T. & Ashby, F. G. 1982 Experimental test of contemporary mathematical models of visual letter recognition. *J. Exp. Psychol. Hum. Percept. Perf.* **8**, 834–864.

Treisman, A. M. 1982 Perceptual grouping and attention in visual search for features and for objects. *J. Exp. Psychol. Hum. Percept. Perf.* **8**, 194–214.

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

Treisman, A. M. & Gormican, S. 1988 Feature analysis in early vision: evidence from search asymmetries. *Psychol. Rev.* **95**, 15–48.

van der Heijden, A. H. C., La Heij, W. & Boer, J. P. A. 1983 Parallel processing of redundant targets in simple visual search tasks. *Psychol. Res.* **45**, 235–254.

van Oeffelen, M. P. & Vos, P. G. 1982 Configurational effects on the enumeration of dots: counting by groups. *Mem. Cogn.* **10**, 396–404.

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

Wolfe, J. M. 1994 Guided search 2.0: a revised model of visual search. *Psychon. Bull. Rev.* **1**, 202–238.

Wolfe, J. M., Cave, K. R. & Franzel, S. L. 1989 Guided search: an alternative to the feature integration model for visual search. *J. Exp. Psychol. Hum. Percept. Perf.* **15**, 419–433.