Psychol Res (1987) 49: 113–121

Psychological Research

Visual attention: Race models for selection from multielement displays

Claus Bundesen

Psychological Laboratory, Copenhagen University, Njalsgade 90, DK-2300 Copenhagen S., Denmark

Summary. A choice model for partial report from briefly exposed visual displays (Bundesen, Shibuya, & Larsen, 1985) is further investigated and related to a general class of selection models called "independent race models." The choice model relates performance to the numbers of targets and distractors in the stimulus display by way of the choice axiom. In race models, the selection process is viewed as a race between items in the choice set toward a state of having been "processed" in that the first items reaching this state are the ones selected. If items are processed independently and processing times are exponentially distributed, selection occurs strictly in accordance with the choice axiom, so the race model is a choice model. The choice model also seems to work as a good approximation for independent race models based on other gamma distributions than the exponential one.

Studies of visual attention deal with limits on our ability, first, to divide attention between simultaneous targets and, second, to focus attention on targets rather than distractors. Theoretically, the two aspects of visual selection may be closely related, since rejection of a distractor requires processing of that distractor, but the relationship between the processing required to reject a distractor and the processing required to attend to a target must be determined empirically. Most empirical studies have addressed one or the other but not both aspects of visual selection. Typically, experiments on selection of targets rather than distractors have been conducted by varying the number of distractors while keeping the number of targets constant at one, whereas experiments on division of attention between multiple targets have been conducted without systematically varying the number of distractors. As recently argued by Duncan (1985), stronger empirical constraints on models for the two aspects of visual selection may be found by studying performance as a joint function of the number of targets and the number of distractors in the stimulus display.

The models for visual selection described in this article are primarily based on experimental studies by Bundesen, Pedersen, and Larsen (1984) and Bundesen, Shibuya, and Larsen (1985) of performance in the partial-report paradigm as a joint function of the numbers of targets and distractors in the stimulus display. In the partial-report paradigm, the subject is instructed to respond to briefly exposed displays by reporting as many targets as possible while ignoring the distractors. In our experiments, trials were blocked by condition such that the subject was informed about the selection criterion before the stimulus was presented. Scores for different selection criteria were compared. Scores for displays without distractors yielded a whole-report baseline.

The work was inspired by data and arguments on whole-report performance provided by Sperling (1960, 1963, 1967). With briefly presented visual displays of letters or digits, Sperling (1960, 1963) found that the number of items correct in whole report was close to the number of items in the stimulus when the display contained four or fewer items, and averaged between four and five items for displays containing five or more items (whole-report limitation). When an effective mask was presented at display offset, the number of items correctly reported increased rapidly from zero to about four as display duration increased from zero up to some value between 50 and 100 ms. With a further increase in display duration, the rate of increase in number of items reported was much smaller, about one item per 100 ms at most (see Coltheart, 1972, 1980; Sperling, 1963, 1967). As argued by Sperling (1967), these findings seem to suggest that the whole-report limitation reflects the limited capacity of a short-term store (recognition buffer memory) with fast read-in and slow read-out.

In the first sections of this article, I describe a choice model for partial report developed by Bundesen et al. (1984, 1985). The model incorporates a limited-capacity short-term store as suggested by Sperling (1967) and a Luce (1959) ratio rule. In later sections, I describe a general class of models for selection from multielement displays, that is, independent race models, and investigate the relationship between such models and the choice model.

A choice model

Our choice model for partial report assumes that, whether partial or whole report is required, performance on any trial reflects the number of targets that enter a limited-capacity short-term memory store; any target that enters the store is correctly reported with probability θ, regardless of the fate of other items. Items entering the store may be targets, distractors, or extraneous noise. The total number of items entering the store, K, is independent of the number of targets and distractors in the stimulus.

Read-in to the store is conceived as selective sampling of items without replacement, such that selection among items occurs in accordance with a constant ratio rule (Clarke, 1957; Luce, 1959; Luce, Bush, & Galanter, 1963). Specifically, for a given selection criterion, there is a ratio scale v such that if R is the set of all items remaining after selection of $k-1$ items ($1 \leq k \leq K$) and $i \in R$, then the conditional probability that item i is the kth to be selected equals

$$\frac{v(i)}{\sum_{j \in R} v(j)}. \quad (1)$$

In words, items are assigned weights or *impacts* such that the probability that any not-yet-selected item will be the next one to be selected equals the impact of that item divided by the sum of impacts for all items not yet selected.

The assumptions involved in Equation 1 may be spelled out as follows. First, kth choices are related to first choices in a way that appears as simple as possible: Given that R is the set of all items remaining after selection of the first $k-1$ items from the choice set U, the conditional probability that item i is the kth to be selected from U is the same as the probability that item i would be the first choice in selection from a choice set consisting exclusively of the members of R.

Second, first choices from different choice sets are interrelated by the choice axiom proposed by Luce (1959; Luce & Galanter, 1963). The axiom can be divided into two parts. Let $P_U(i)$ denote the probability that in selecting from the choice set U, item i is the first to be selected. The first part of the choice axiom says that if $P_U(i) > 0$ for all $i \in U$, then for all i and S such that $i \in S \subset U$,

$$P_S(i) = P_U(i \mid S),$$

where $P_S(i)$ is the probability of selecting item i from the choice set S, and $P_U(i \mid S)$ is the conditional probability of selecting i from the choice set U given that the selected item belongs to S. The second part of the choice axiom says that if $P_U(k) = 0$ for some $k \in U$, then for every $S \subset U$,

$$P_U(S) = P_{U-\{k\}}(S-\{k\}),$$

where $P_U(S)$ is the probability of selecting an item belonging to S from the choice set U, and $P_{U-\{k\}}(S-\{k\})$ is the probability of selecting an item belonging to $S-\{k\}$ from the choice set $U-\{k\}$.

As pointed out by Luce (1959), the first part of the choice axiom guarantees that if $P_U(i) > 0$ for all $i \in U$, then there is a ratio scale v on U such that for all i and S such that $i \in S \subset U$,

$$P_S(i) = \frac{v(i)}{\sum_{j \in S} v(j)},$$

and this ratio scale is given by $v(i) = kP_U(i)$, where k is a constant unequal to zero. The second part of the choice axiom means that any item $i \in U$ such that $P_U(i) = 0$ may be deleted from U without affecting any of the choice probabilities. If the choice axiom holds not only for U itself but also for every subset of U, then the second part of the axiom may be applied repeatedly such that any choice set $S \subset U$ is reduced to one consisting exclusively of items i such that $P_U(i) > 0$. The first part of the choice axiom can then be applied to the reduced choice set.

To justify Equation 1, the choice axiom must hold for the set of all items U, that is, the union of the stimulus ensemble (consisting of all the elements that might appear in a stimulus display) and the set of extraneous noise items – and for every subset of U containing at least one item j such that $P_U(j) > 0$. When considering selections from such sets, any item i for which $P_U(i) = 0$ may be treated as having a v value of zero. Whether the choice axiom holds for subsets of U consisting exclusively of items i such that $P_U(i) = 0$ is immaterial as far as Equation 1 is concerned; the equation cannot be applied to such cases because the denominator must be different from zero.

Four-parameter model

The four-parameter version of the choice model for partial report assumes that only those targets that enter the short-term store are correctly reported, so the model implies that in case n targets enter the short-term store, the conditional probability distribution for the number of targets correctly reported is the binomial distribution for n Bernoulli trials with probability θ for success. Two further simplifications are made. First, all targets have identical impacts and all distractors have identical impacts, so no generality is lost in setting the impact of a target to 1 and the impact of a distractor to α, where α is a constant. Second, the number of extraneous noise items (in the experimental situation or in long-term memory) is large in relation to K, and each one has a small probability of being sampled on a given trial, so the total impact of the not-yet-selected extraneous noise items, ε, is essentially constant during a trial.

The above simplifications leave four parameters: K, the number of items entering the short-term store; α, the impact per distractor with impact per target as the unit; ε, the total impact of extraneous noise with impact per target as the unit; and θ, the probability that a target that has entered the store is reported. Parameter α is a measure for the efficiency of selecting targets rather than distractors. If α equals zero, selection is perfect. If α equals one, sampling is nonselective.

To see how the model works, consider a subject trying to select as many targets as possible from a briefly exposed display containing T targets and D distractors. Let K equal four. Regardless of T and D, a total of four items is transferred to the short-term memory store. As an example, if both T and D are greater than 1, the probability that the first item selected is an extraneous noise item, the second a target, the third a distractor, and the fourth a target is given by the product of $\varepsilon/(T + \alpha D + \varepsilon)$, $T/(T + \alpha D + \varepsilon)$, $\alpha D/[(T-1) + \alpha D + \varepsilon]$, and $(T-1)/[(T-1) + \alpha(D-1) + \varepsilon]$. In the case where two targets enter the short-term store, the conditional probability distribution for the number of targets correctly reported is the binomial distribution for two Bernoulli trials with probability θ for success.

Three-parameter version

The predicted mean score (mean number of targets reported) for a given combination of T and D equals the product of θ and the predicted mean number of targets sampled. Hence, if and when the predicted mean number of targets sampled is proportional to K, then K and θ are not separately identifiable from observed mean scores (as distinct from the underlying frequency distributions of scores). This case is approximated when T and D are large in relation to K so that the predicted mean number of targets sampled is close to the number that would have been obtained if the model had assumed sampling with replacement rather than sampling without replacement. Accordingly, when the analysis is based on observed mean scores and T and D are large in relation to K, the four-parameter model effectively reduces to a three-parameter model with a single parameter K' representing the product of K and θ. Computationally, the three-parameter model is identical to the four-parameter model with parameter θ kept constant at a value of 1 and $K = K'$.

Goodness of fit

Bundesen et al. (1984) tested the three-parameter version of the model against observed mean scores in a variety of conditions with partial reports based on brightness, color, shape, and alphanumeric class. In each condition, targets and distractors were alphanumeric characters positioned at random within a 5 × 5 matrix. Number of targets T and number of distractors D were varied orthogonally from 0 to 20 in steps of 5 with the constraint that $0 < (T + D) \leq 25$. Each display was exposed for either 60 ms (Experiment 1) or 100 ms (Experiment 2) with dark pre- and postfields. As illustrated in Figures 1 and 2, the model gave good fits to the data. Figure 1 displays the group mean of reciprocated individual mean scores as a function of D/T (the ratio of the number of distractors to the number of targets) with T as a parameter for selection by color (Experiment 1). The theoretical fit is indicated by unmarked points connected with straight lines. This fit is close to the fit that would have been obtained if the model had assumed sampling with replacement rather than sampling without, namely, $\mu = K'T/(T + \alpha D + \varepsilon)$ and therefore

$$\mu^{-1} = (\alpha/K')(D/T) + (\varepsilon/K')(1/T) + (1/K') \qquad (2)$$

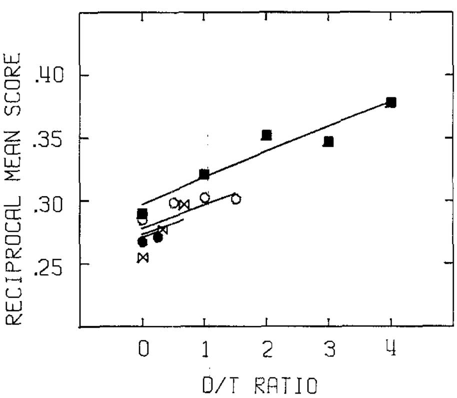

Fig. 1. Group mean of reciprocated mean scores (number of correctly reported items) as a function of D/T ratio (number of distractors to number of targets) with T (number of targets) as a parameter for selection by color. T was 5 (squares), 10 open circles), 15 (hourglasses), or 20 (solid circles). Unmarked points connected with straight lines represent a theoretical fit to the data by the choice model for partial report. From C. Bundesen, L. F. Pedersen, and A. Larsen, 1984, *Journal of Experimental Psychology: Human Perception and Performance, 10,* pp. 329–339, Experiment 1. Copyright 1984 by the American Psychological Association

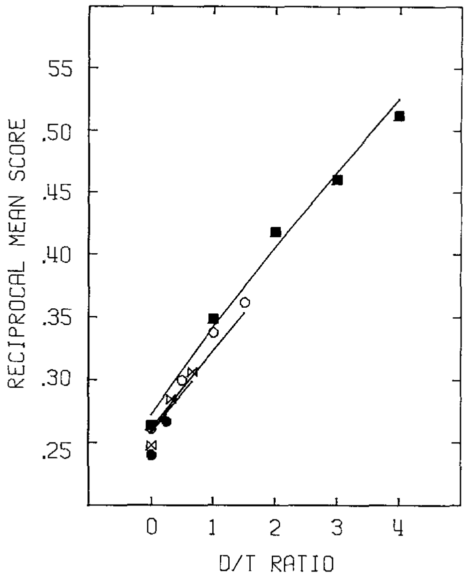

Fig. 2. Group mean of reciprocated mean scores (number of correctly reported items) as a function of D/T ratio (number of distractors to number of targets) with T (number of targets) as a parameter for selection by alphanumeric class. T was 5 (squares), 10 (open circles), 15 (hourglasses), or 20 (solid circles). Unmarked points connected with straight lines represent a theoretical fit to the data by the choice model for partial report. From C. Bundesen, L. F. Pedersen, and A. Larsen, 1984, *Journal of Experimental Psychology: Human Perception and Performance, 10,* pp. 329–339, Experiment 1. Copyright 1984 by the American Psychological Association

Figure 2 shows the corresponding results for selection by alphanumeric class in the same experiment.

The group mean of individual estimates for K' varied little across conditions, with 4.0 and 4.1 items for the color and alphanumeric conditions, respectively, in Experiment 1, and 3.6, 3.9, and 3.3 items for brightness, alphanumeric, and shape conditions in Experiment 2. Estimates for α (impact per distractor with impact per target as the unit) varied widely across conditions with group means ranging from 0.02 in the brightness condition of Experiment 2 to 0.65 in the shape condition of the same experiment. This variation in the efficiency of selecting targets rather than distractors corresponds to the strong variation in slope (i.e., the variation in effect of D/T ratio) between the functions depicted in Figures 1 and 2 (cf. Equation 2). Estimates for ε (total impact of extraneous noise with impact per target as the unit) were rather small, with group means ranging from 0.28 to 1.15. The effect of ε corresponds to the spacing between functions within Figures 1 and 2 (i.e., the effect of parameter T).

Bundesen et al. (1985) tested the four-parameter model against observed frequency distributions of the number of items correctly reported as functions of number of targets (2, 4, or 12), number of distractors (0, 2, 4, 6, 8, or 10), and selection criterion (color or alphanumeric class). Targets and distractors were alphanumeric characters positioned around the perimeter of an imaginary circle centered on fixation. Exposure time was 60 ms and pre- and postfields were dark. The analysis was based on individual data for two extensively tested subjects. Since little was gained by having ε as a free parameter, ε was kept constant near zero.

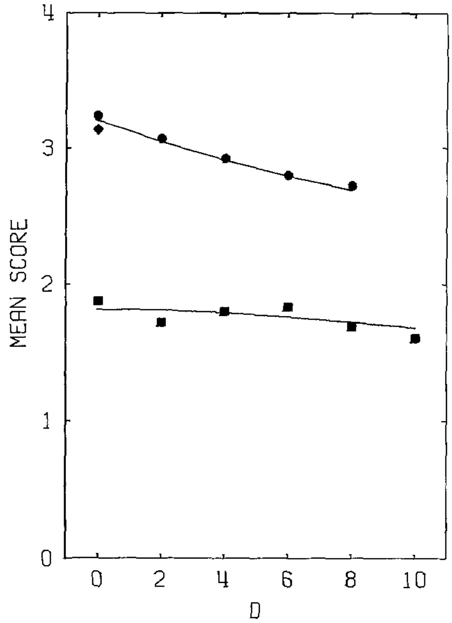

Fig. 3. Mean score (number of correctly reported items) for subject HS as a function of *D* (number of distractors) with *T* (number of targets) as a parameter for selection by color. *T* was 2 (squares), 4 (circles), or 12 (diamond). Unmarked points connected with straight lines represent a theoretical fit to the data by the choice model for partial report. From *Attention and Performance XI* (pp. 631–649) by M. I. Posner and O. S. M. Marin (Eds.), 1985, Hillsdale, NJ: Erlbaum. Copyright 1985 by The International Association for the Study of Attention and Performance

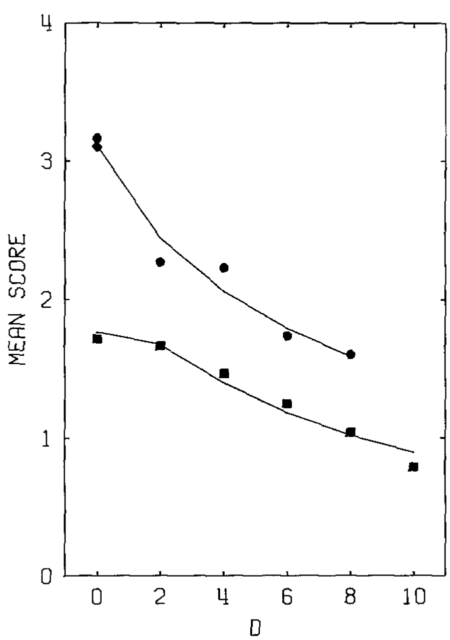

Fig. 4. Mean score (number of correctly reported items) for subject HS as a function of *D* (number of distractors) with *T* (number of targets) as a parameter for selection by alphanumeric class. *T* was 2 (squares), 4 (circles), or 12 (diamond). Unmarked points connected with straight lines represent a theoretical fit to the data by the choice model for partial report. From *Attention and Performance XI* (pp. 631–649) by M. I. Posner and O. S. M. Marin (Eds.), 1985, Hillsdale, NJ: Erlbaum. Copyright 1985 by The International Association for the Study of Attention and Performance

The model did fairly well at describing the frequency distributions of scores for individual subjects. Out of 192 observed frequencies, 163 were less than two standard deviations from the predicted values, and none were more than three standard deviations from predictions. The resulting fits to the observed mean scores are illustrated in Figures 3 and 4. Figure 3 displays the mean score as a function of D with T as a parameter for subject HS and selection by color. Figure 4 shows the corresponding results for selection by alphanumeric class.

Estimates for parameters K and θ varied little with the selection criterion. Across the two subjects, estimates for K averaged 3.57 for the color conditions and 3.52 for the alphanumeric conditions. Estimates for θ averaged .92 for the color conditions and, again, .92 for the alphanumeric conditions. Estimates for parameter α varied widely with the selection criterion, averaging 0.05 for the color conditions and 0.36 for the alphanumeric conditions.[1]

Independence of parameter α from display size

Parameter α was proposed as a measure for the efficiency of selecting targets rather than distractors. The model embedded in Equation 1 implies that α is the same regardless of T and D, and the goodness of fit between the model and the data of Bundesen et al. (1984, 1985) lends some support to this strong assumption.

As a further test for the assumption that α is independent of T and D, I tentatively modified the model by incorporating a new parameter that should reflect any dependence of α (impact per distractor with impact per target as the unit) upon the number of items in the display. Specifically, I let

$$\alpha = a^{(N^b)},$$

where N is the total number of not-yet-selected items in the stimulus display, and $a \geqslant 0$. Clearly, for $N = 1$, α equals a, regardless of b. For $b = 0$, the modified model is equivalent to the old four-parameter model: α remains constant at a regardless of N. For $a < 1$ and $b < 0$, α is an increasing function of N such that α tends to 1 (nonselective sampling) as N tends to infinity. For $a < 1$ and $b > 0$, α is a decreasing function of N such that α tends to 0 (perfect selection) as N tends to infinity.

Table 1 summarizes the least squares fits reported by Bundesen et al. (1984) for the three-parameter (K', α, ε) version of the choice model applied to group data (harmonic means of individual subjects' mean scores) for the color and alphanumeric conditions of their first experiment and the brightness, alphanumeric, and shape conditions of their second experiment. The same fits may be regarded as being generated by the modified model with parameter b kept constant at zero. Table 2 shows the corresponding fits generated by the modified model with b as a

[1] To see how the model parameters relate to the mean scores plotted in Figures 3 and 4, note that when ε is kept constant near zero, the mean score predicted for T:D combination 2:0 $\cong 2\theta$, provided that $K \geqslant 2$, and the mean score predicted for T:D combination 12:0 $\cong K\theta$, provided that $K \geqslant 12$. Hence, given that $2 \geqslant K \geqslant 12$, both K and θ might be estimated from mean scores for T:D combinations 2:0 and 12:0 ("whole-report scores"). The remaining parameter, α, reflects both the effects of D and the interactions between T and D depicted in Figures 3 and 4

Table 1. Fits reported by Bundesen et al. (1984) for three-parameter version of choice model for partial report applied to group mean scores for selection by various criteria

Selection criterion	Parameter			%V[a]	RMSD[b]
	K'	α	ε		
Experiment 1					
Color	3.79	0.067	0.41	91.3	0.103
Alphanumeric class	4.01	0.255	0.22	98.0	0.096
Experiment 2					
Brightness	3.49	0.024	0.72	83.0	0.108
Alphanumeric class	3.88	0.502	1.19	98.5	0.096
Shape	3.14	0.626	0.58	97.9	0.103

$K' = K\theta$, where K is the number of items entering short-term store and θ is the probability that a target that has entered the store will be reported; α = impact per distractor with impact per target as the unit; ε = total impact of extraneous noise with impact per target as the unit

[a] Percentage of variance (with number of targets and number of distractors) in group mean score (number of items correctly reported) accounted for by the fit

[b] Square root of the mean squared deviation between observed and theoretical mean scores

Table 3. Fits reported by Bundesen et al. (1985) for four-parameter version of choice model for partial report applied to individual frequency distributions of scores for selection by various criteria

Selection criterion	Parameter			%V[a]	RMSD[b]
	K	α	θ		
Subject HS					
Color	3.53	0.061	.906	99.3	0.052
Alphanumeric class	3.52	0.419	.883	98.5	0.086
Subject MJ					
Color	3.61	0.043	.942	98.3	0.087
Alphanumeric class	3.53	0.304	.965	98.5	0.090

K = number of items entering short-term store; α = impact per distractor with impact per target as the unit; θ = probability that a target that has entered the store will be reported. Parameter ε representing the total impact of extraneous noise with impact per target as the unit was kept constant at 10^{-10}

[a] Percentage of variance (with number of targets and number of distractors) in observed mean score (number of items correctly reported) accounted for by the fit

[b] Square root of the mean squared deviation between observed and theoretical mean scores

free parameter. As can be seen, the gain in goodness of fit by introducing b as a free parameter was very small. Averaged across the five conditions, the proportion of variance with number of targets T and number of distractors D accounted for by the model changed from 93.7% to 94.2%. Estimates for parameter b ranged from -0.04 to $+0.31$. Estimates for parameters K' and ε were nearly the same for the modified as for the original model.

Table 3 summarizes the maximum likelihood fits reported by Bundesen et al. (1985) for the four-parameter (K, α, ε, θ) choice model (with ε kept constant near zero) applied to individual data (frequency distributions of scores as functions of T and D) for two subjects tested in color and alphanumeric conditions. Table 4 summarizes the corresponding fits for the modified model. The gain in goodness of fit by introducing b as a free parameter was

Table 2. Fits of modified choice model for partial report applied to group mean scores reported by Bundesen et al. (1984) for selection by various criteria

Selection criterion	Parameter				%V[a]	RMSD[b]
	K'	a	b	ε		
Experiment 1						
Color	3.77	0.189	0.159	0.33	92.0	0.099
Alphanumeric class	4.00	0.362	0.103	0.19	98.1	0.093
Experiment 2						
Brightness	3.46	0.244	0.307	0.59	84.5	0.103
Alphanumeric class	3.87	0.603	0.107	1.16	98.5	0.095
Shape	3.14	0.591	–0.040	0.59	97.9	0.103

$K' = K\theta$, where K is the number of items entering short-term store and θ is the probability that a target that has entered the store will be reported; a and b determine how α (impact per distractor with impact per target as the unit) varies with N (number of not-yet-selected display items) as $\alpha = a^{(N^b)}$; ε = total impact of extraneous noise with impact per target as the unit

[a] Percentage of variance (with number of targets and number of distractors) in group mean score (number of items correctly reported) accounted for by the fit

[b] Square root of the mean squared deviation between observed and theoretical mean scores

Table 4. Fits of modified choice model for partial report applied to individual frequency distributions of scores reported by Bundesen et al. (1985) for selection by various criteria

Selection criterion	Parameter				%V[a]	RMSD[b]
	K	a	b	θ		
Subject HS						
Color	3.52	0.009	–0.231	.904	99.3	0.052
Alphanumeric class	3.52	0.347	–0.094	.882	98.5	0.086
Subject MJ						
Color	3.61	0.041	–0.005	.942	98.3	0.087
Alphanumeric class	3.50	0.074	–0.379	.963	99.1	0.069

K = number of items entering short-term store; a and b determine how α (impact per distractor with impact per target as the unit) varies with N (number of not-yet-selected display items) as $\alpha = a^{(N^b)}$; θ = probability that a target that has entered the store will be reported. Parameter ε representing the total impact of extraneous noise with impact per target as the unit was kept constant at 10^{-10}

[a] Percentage of variance (with number of targets and number of distractors) in observed mean score (number of items correctly reported) accounted for by the fit

[b] Square root of the mean squared deviation between observed and theoretical mean scores

negligible; averaged across the four data sets, the proportion of variance in the mean score accounted for by the model changed from 98.7% to 98.8%. Estimates for parameter b ranged from -0.005 to -0.38; estimates for K and θ were virtually the same for the modified as for the original model.

The findings that, first, goodness of fit was improved very little by introducing b as a free parameter and, second, estimates for b were fairly small with an overall median near zero (viz., at -0.005) support the four-parameter choice model for partial report: To a good approximation, α seems to be independent of T and D.

Race models

The choice rule expressed in Equation 1 was originally chosen on grounds of simplicity and computational convenience. It is tempting to try to extend the choice model for visual selection in depth by relating the choice rule to plausible process models. Below I describe a general class of process models for selection from multielement displays – independent race models – and investigate the relationship between such models and the choice model.

Let a *race model* for selection be a model in which the selection process is viewed as a race between items in the choice set toward a state "processed" such that, for some $K \geqslant 1$, the first K items reaching the state processed are the ones selected. If processing times for individual items in the choice set (the times at which individual items reach the state processed) are independent random variables, I call the model an *independent* race model.

Let $F_i(t)$ be the (continuous) distribution function for the processing time of item i. An independent race model in which, for any item i, $F_i(t)$ is the same regardless of the choice set in which item i is presented, is said to be *unlimited* in *processing capacity*. Independent race models with unlimited processing capacity are easy to treat: Let W be a subset of the choice set S such that W consists of K items. By the independence assumption, the probability that the K members of W are the first K items reaching the state processed when selection is from S is given by

$$P_S(W) = \sum_{i \varepsilon W} \int_0^\infty \prod_{h \varepsilon W - \{i\}} F_h(t) \prod_{j \varepsilon S - W} [1 - F_j(t)]\, dF_i(t). \tag{3}$$

To analyze independent race models with *limited processing capacity*, we need a quantitative notion of processing capacity. Intuitively, variations in the amount of processing capacity allocated to an item concern the rate at which the item is processed, but not the type of processing that is done. More specifically, I propose that the effect of processing an item from time 0 to time t with a constant capacity of k units should equal the effect of processing the item from time 0 to time kt with a constant capacity of 1 unit. Similarly, the effect of processing an item from time 0 to time t with a capacity of $C(x)$ units at time x ($0 \leqslant x \leqslant t$) should equal the effect of processing the item from time 0 to time $\int_0^t C(x)dx$ with a constant capacity of 1 unit. Thus, if $F_i(t)$ is the conditional distribution function for the processing time of item i given that the capacity allocated to item i at time x is $C_i(x)$ units, and $G_i(t)$ is the conditional distribution function given that the capacity allocated to item i is kept constant at 1 unit, then

$$F_i(t) = G_i\left(\int_0^t C_i(x)\, dx\right). \tag{4}$$

The quantitative notion of capacity expressed in Equation 4 is a simple generalization of the notion of capacity found in Rumelhart's (1970) multicomponent theory for the perception of briefly exposed visual displays.

Rumelhart (1970) introduced a notion of attentional weights to account for the way capacity is spread over the set of items in the choice set: For any items i and j, the ratio between the amount of capacity allocated to item i and the amount of capacity allocated to item j should equal the ratio between the attentional weight of item i (w_i) and the attentional weight of item j (w_j). Whereas the amount of capacity allocated to an item should be quite variable, the attentional weight of the item should be comparatively stable.

Consider an independent race model with limited processing capacity and attentional weights that are constant over time. The assumption that attentional weights are constant over time means that there is a function $V(t)$ such that for any item i in the choice set, the capacity allocated to item i at time t equals $w_i V(t)$, where w_i is the attentional weight of item i. As before, let W be a subset of the choice set S such that W consists of K items and let us calculate the probability $P_S(W)$ that the K members of W are the first K items reaching the state processed when selection is from S.

By the independence assumption, $P_S(W)$ is given by Equation 3 if distribution functions $F_h(t)$, $F_i(t)$, and $F_j(t)$ are interpreted as conditional distribution functions for the processing times of items h, i, and j given attentional weights w_h, w_i, and w_j and the variation in capacity specified by $V(t)$. Since the capacity allocated to item i at time t equals $w_i V(t)$, Equation 4 implies that

$$F_i(t) = G_i\left(\int_0^t w_i V(x)\, dx\right) = H_i\left(\int_0^t V(x)\, dx\right), \tag{5}$$

if $H_i(t)$ is defined as the conditional distribution function for the processing time of item i, given that the capacity allocated to item i is kept constant at w_i units. By substituting Equation 5 in Equation 3 (and assuming that the total capacity spread over items in the choice set never vanishes before K items have been processed) one finds that $P_S(W)$ is given by Equation 3 with $H_h(t)$, $H_i(t)$, and $H_j(t)$ substituted for $F_h(t)$, $F_i(t)$, and $F_j(t)$, respectively. Thus, if attentional weights are kept constant, variations in capacity affect expectations concerning the times at which items are sampled, but expectations concerning the order in which items are sampled are not affected.

The results may be summarized as follows: For independent race models with unlimited processing capacity, selection probabilities $P_S(W)$ are given by Equation 3. For any independent race model with limited processing capacity and constant attentional weights, there are independent race models with unlimited processing capacity that predict the same selection probabilities $P_S(W)$ as the limited capacity model. One of these unlimited capacity models can be constructed by supplying each item i with a distribution function equal to $H_i(t)$; $H_i(t)$ was defined as the distribution function assumed in the limited capacity model for the processing time of item i given that the capacity allocated to item i is kept constant at w_i units, where w_i is the attentional weight of item i.

Exponential and gamma models

Bundesen et al. (1985) showed that the choice rule expressed in Equation 1 is implied by race models in which items are processed independently and processing times are exponentially distributed. The fits reported by Bundesen et al. (1984, 1985) may thus be regarded as fits by independent race models with unlimited processing capacity and exponential distribution functions for processing times of targets, distractors, and extraneous noise elements. In this interpretation, the impact of an element is proportional to the exponential rate parameter of the distribution function for the element. Setting the rate at which a target is processed at 1 per unit of time, α is the rate at which a distractor is processed, and ε is the overall rate of noise processing (the sum of all rate parameters for extraneous noise elements). The way parameter ε is treated in the choice model can be proved correct on the hypothesis that the number of extraneous noise elements processed between time 0 and time t is Poisson distributed with parameter εt, and this hypothesis can be derived from the assumption that distribution functions for extraneous noise elements are exponential with individual rate parameters that are vanishingly small in relation to ε.

Alternatively, the fits reported by Bundesen et al. (1984, 1985) may be regarded as fits by independent race models with limited processing capacity and attentional weights that are constant over time. Parameter α might thus be regarded as the ratio of the attentional weight of a distractor to the attentional weight of a target, and parameter ε might be regarded as the ratio of the sum of the attentional weights of all noise elements to the attentional weight of a target.

The exponential distribution with rate parameter μ is a special case of the gamma distribution with rate parameter μ and convolution parameter r, namely, the case in which r equals 1. As r increases, the gamma distribution changes in shape such that, in the limit, it becomes a normal distribution. To see how the behavior of independent race models based on gamma distributions depend upon the choice of parameter r, I refitted the data collected by Bundesen et al. (1985). Specifically, I assumed that processing times for targets and distractors were gamma distributed with different rate parameters, but with common convolution parameter r. Considering the results of the analysis by Bundesen et al. (1985), effects of extraneous noise on the observed scores were assumed to be negligible. The resulting gamma model contained four parameters: K, the number of items entering the short-term store; r, the convolution parameter of the gamma distributions; α, the ratio of the rate parameter for a distractor to the rate parameter for a target; and θ, the probability that a target that has entered the short-term store will be reported. In fitting the model to the data, I followed the procedures used by Bundesen et al. (1985). Fits approached in the limit as $r \to \infty$ (referred to as "fits for $r = \infty$") were determined by treating processing times for targets and distractors as independent, normally distributed random variables with equal dispersions.

Table 5 summarizes maximum likelihood fits by the independent gamma race model with parameter r kept constant at 1, 2, 4, and ∞, respectively. The fits by the gamma model with parameter r kept constant at 1 are identical to the original (Bundesen et al., 1985) fits by the choice model. To facilitate comparison between the four models (versions of the gamma model with $r = 1, 2, 4$, and ∞), each model is described in terms of parameter p_d in addition to parameters r, K, and θ; p_d is the theoretical probability that in a race between a single target and a single distractor, the target reaches the goal state before the distractor.[2] As can be seen, both goodness of fit and estimates for parameters K, p_d, and θ were nearly the same for the four models.

To sum up, the original fits to our data by the choice model for visual selection may be regarded as fits by an independent exponential race model with unlimited processing capacity or with limited processing capacity and attentional weights that are constant over time. Highly similar fits were produced by assuming that processing times for targets and distractors were gamma distributed with common convolution parameter r equal to 2, 4, or ∞, rather than to 1. The results suggest that the choice model works as a good approximation for independent race models based on gamma distributions with convolution parameter r, regardless of the value of r.

Table 5. Fits of gamma race models with various convolution parameters applied to individual frequency distributions of scores reported by Bundesen et al. (1985) for selection by various criteria

Convolution parameter r	Parameter			%V[a]	RMSD[b]
	K	p_d	θ		
Subject HS					
Selection by color					
1	3.53	.943	.906	99.3	0.052
2	3.54	.945	.905	99.3	0.051
4	3.54	.946	.905	99.3	0.051
∞	3.54	.947	.903	99.2	0.057
Selection by alphanumeric class					
1	3.52	.705	.883	98.5	0.086
2	3.53	.701	.884	98.5	0.087
4	3.53	.698	.884	98.4	0.088
∞	3.53	.691	.885	98.3	0.092
Subject MJ					
Selection by color					
1	3.61	.959	.942	98.3	0.087
2	3.61	.961	.941	98.1	0.091
4	3.62	.963	.940	97.9	0.096
∞	3.62	.965	.937	97.3	0.108
Selection by alphanumeric class					
1	3.53	.767	.965	98.5	0.090
2	3.53	.764	.965	98.3	0.096
4	3.53	.762	.965	98.2	0.101
∞	3.54	.757	.964	97.7	0.112

K = number of items entering short-term store; p_d = probability that a target is processed faster than a distractor; θ = probability that a target that has entered the store will be reported

[a] Percentage of variance (with number of targets and number of distractors) in observed mean score (number of items correctly reported) accounted for by the fit

[b] Square root of the mean squared deviation between observed and theoretical mean scores

[2] By use of Equation 3, it can be shown that

$$p_d = \sum_{i=0}^{r-1} \binom{r-1+i}{i} \alpha^i (1+\alpha)^{-i-r}$$

Discussion

Choice model

The choice model for visual selection has provided highly accurate accounts for the joint effects of the numbers of targets and distractors on the number of items correct in partial reports with a variety of selection criteria. Estimates obtained for the parameters appear psychologically plausible, and variation in a single parameter, α, has accounted for the strong effects in performance generated by varying the selection criterion. The model seems generally consistent with the literature on visual search through briefly exposed displays, and it has clearly testable implications regarding which of its parameters ought to be influenced by variables such as practice and target-distractor discriminability (see Bundesen et al., 1985).

Analysis of partial reports in terms of the choice model has suggested two fundamental measures for attentional limitations: Our ability to *divide* attention between simultaneous stimuli is measured by parameter K, the total number of items entering the short-term store. In the experiments of Bundesen et al. (1984, 1985), estimates for K ranged between three and four items, regardless of the selection criterion. Our ability to *focus* attention on targets rather than distractors is measured by parameter α, the impact per distractor with the impact per target as the unit. In the experiments of Bundesen et al., estimates for α ranged from about 0.02 to about 0.65, depending upon the selection criterion.

Race models

The choice model relates performance to the numbers of targets and distractors in the stimulus display by way of the choice axiom. In race models, the selection process is viewed as a race between items toward a certain state such that the first K items reaching this state are the ones selected. On the basis of the selection probabilities predicted, an independent race model with unlimited processing capacity is indistinguishable from a model with limited processing capacity and time-invariant attentional weights. If processing times for individual items are exponentially distributed, both models predict selection probabilities that satisfy the choice axiom. The choice model also seems to work as a good approximation for independent race models based on other gamma distributions than the exponential one.

Other models

There are other process models for selection than race models. In *comparison models* based on signal detection theory (Tanner & Swets, 1954), each item is associated with a strength variable taking values on a one-dimensional continuum and items are assumed to be selected in order of decreasing (or increasing) strength. Comparison models for visual selection have been proposed by Hoffman (1978) and Shaw (1980, 1984). Clearly, for any race model, there is a comparison model yielding the same selection probabilities as the race model, and this comparison model may be constructed from the race model by reinterpreting the probability density $f_i(t)$ that the processing time for item i equals t as the probability density that the strength variable associated with item i takes the value t. As the comparison model requires a process of comparison between stimulus items with respect to strength, the comparison model is noticeably more complex than the corresponding race model.

For analysis of partial reports both simplicity and face validity favor race models rather than comparison models; *prima facie,* reporting as many targets as possible from a stimulus display requires no comparison between stimulus items within the display. For some other paradigms, however, comparison models have face validity. One example is a localization paradigm used by Shaw (1984), in which the task is to select the stimulus location providing the strongest impression that it contains a target; *prima facie,* this task requires a comparison between stimulus items within the display with respect to target-likeness.

Shaw (1984) found that with increasing display size, the decrease in accuracy for locating a target letter among other letters was too large to be accommodated by any independent comparison model assuming that, first, strengths of items are independent of display size and, second, regardless of display size, the subject always chooses the location with the largest strength. Maintaining the second assumption, she suggested that for letter detection, division of attention causes increasing overlap between the strength distributions for targets and distractors as display size is increased. An alternative interpretation that maintains the first assumption is that with increasing display size, the selection of the location of the stimulus item generating the most target-letter-like impression becomes less and less accurate due to increasing load in the comparison stage. In this interpretation, comparisons between stimulus items with respect to a criterion such as degree of similarity to a target letter are subject to capacity limitations. Modeling such comparison processes and explaining their limitations goes beyond the scope of the race models I have developed for visual selection from multielement displays.

References

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: Erlbaum.

Clarke, F. R. (1957). Constant-ratio rule for confusion matrices in speech communication. *Journal of the Acoustical Society of America, 29,* 715–720.

Coltheart, M. (1972). Visual information-processing. In P. C. Dodwell (Ed.), *New horizons in psychology* (Vol. 2, pp. 62–85). Harmondsworth, England: Penguin Books.

Coltheart, M. (1980). Iconic memory and visible persistence. *Perception & Psychophysics, 27,* 183–228.

Duncan, J. (1985). Visual search and visual attention. In M. I. Posner & O. S. M. Marin (Eds.), *Attention and performance XI* (pp. 85–105). Hillsdale, NJ: Erlbaum.

Hoffman, J. E. (1978). Search through a sequentially presented visual display. *Perception & Psychophysics, 23,* 1–11.

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

Luce, R. D., Bush, R. R., & Galanter, E. (Eds.) (1963). *Handbook of mathematical psychology: Vol. 1.* New York: Wiley.

Luce, R. D., & Galanter, E. (1963). Discrimination. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, pp. 191–243). New York: Wiley.

Rumelhart, D. E. (1970). A multicomponent theory of the perception of briefly exposed visual displays. *Journal of Mathematical Psychology, 7,* 191–218.

Shaw, M. (1980). Identifying attentional and decision-making components in information processing. In R. S. Nickerson (Ed.), *Attention and performance VIII* (pp. 277–296). Hillsdale, NJ: Erlbaum.

Shaw, M. (1984). Division of attention among spatial locations: A fundamental difference between detection of letters and detection of luminance increments. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 109–121). Hillsdale, NJ: Erlbaum.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74* (11), 1–29.

Sperling, G. (1963). A model for visual memory tasks. *Human Factors, 5,* 19–31.

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

Tanner, W. P. Jr., & Swets, J. A. (1954). A decision making theory of visual detection. *Psychological Review, 61,* 401–409.

Received November 29, 1985/November 26, 1986