

Detecting target elements in multielement arrays: A confusability model

R. A. KINCHLA

Princeton University, Princeton, New Jersey 08540

A "reductive coding" model for the detection of critical elements in multielement arrays originally applied to auditory data is shown to provide an interpretation of "set size" and "redundant critical element" effects in visual letter detection data.

This paper considers the general perceptual problem of evaluating multielement stimulus arrays for the presence of certain critical elements. The theoretical approach is similar to that developed in earlier papers by this author (Kinchla, 1966, 1969) to characterize auditory tasks in which array elements (sounds) were *successively* presented. It was argued then that a similar analysis might be appropriate for visual tasks in which array elements are presented *simultaneously*. Of direct relevance to this argument is a paper by Eriksen and Spencer (1969), which indicated that an O's ability to detect a critical letter in a multiletter visual array was the same whether the letters were presented almost simultaneously (one letter each 5 msec) or successively at a very slow rate (one every 5 sec). This apparent equivalence of certain "simultaneous" and "successive" presentation procedures has been confirmed in subsequent studies by Shiffrin and Gardner (1972). These results seem to undercut the concept of "serial coding" processes in vision, since models incorporating such processes attribute much of an O's difficulty in tachistoscopic letter detection tasks to his inability to process sufficient information from a short-term visual memory before "decay" or "interference" effects terminate the process (Estes & Taylor, 1964; Rumelhart, 1970; Sperling, 1963). Presenting individual letters tachistoscopically but in very slow succession should allow an O almost as much time to process each letter as he has to process all of them when they are presented simultaneously (i.e., the "decay" of the second letter would not begin until ample time had been provided to process the first letter, etc.). The apparently contradictory findings cited here, as well as other results, have led to a critical reappraisal of the concept of "serial coding" processes in vision (Estes, 1972; Gardner, 1973; Shiffrin & Gardner, 1972). Eriksen and Spencer interpreted their results as indicating confusions between "noisy" or partially processed letters, confusions which occurred whether the letters were observed simultaneously or in succession. Thus, the O's problem was to make something like a "statistical decision" based on multiple "noisy" samples, rather than a "race to code a rapidly decaying iconic memory." This is the "multiple observation" problem considered in earlier papers by this author (Kinchla, 1966, 1969). The model developed

in those papers is used here to interpret a particular type of tachistoscopic letter detection task; specifically, one in which Os must decide whether a briefly presented multiletter array did or did not contain a specific "target" letter.

Before proceeding further, it will be useful to introduce some formal notation for describing the task we shall consider. An *n*-element stimulus array will be denoted by the symbol A , and each of its constituent elements (letters) by e_i ($i = 1, 2, \dots, n$). Each e_i will be said to have a *value*, denoted v_i , of 1 if e_i is a target or 0 if e_i is a nontarget. The *number of target elements* in an array will be denoted by t . In the experiments considered here the ordering or positioning of elements within an array is randomized each time the array is presented with all possible orderings equally likely. An O's report that an array contained some target elements will be termed an R_1 response, while a report that no targets were presented will be denoted R_0 . Thus, his performance can be summarized by two proportions: the proportion of R_1 responses to A_1 arrays, denoted $P(R_1 | A_1)$, and to A_0 arrays, denoted $P(R_1 | A_0)$. These are analogous to "hit" and "false-alarm" rates in a conventional signal detection task and can be interpreted as estimates of corresponding conditional probabilities, $P(R_1 | A_1)$ and $P(R_1 | A_0)$.

The object of this paper is to show how the number of elements in an array (n) and the number of target elements (t) might interact with an O's "sensitivity" and "judgmental standards" to determine his performance.

THEORY

The model we shall employ can be interpreted as a series of "coding" operations designed to isolate those aspects of one's sensory activity relevant to a particular judgment. This *reductive coding process* is represented schematically in Fig. 1. For our present purposes, *sensory activity* can be thought of as the ongoing activity of a large number of "receptive units." The influence of a stimulus array on one's sensory system would depend on the physical properties of the stimulus, the nature and current state of the receptive units, and on their "orientation," where *orientation* refers to any efferent action which alters the incidence of external

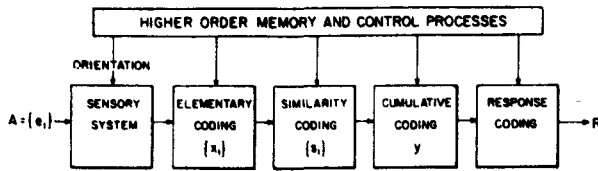


Fig. 1. Schematic of the theoretical reductive coding process.

energy on the neuronal interface (e.g., accommodation, dilation, and eye movements in the case of vision, and head movements, or even muscular control of the bones in the inner ear, in the case of audition).

We shall consider four successive coding operations which might be employed in evaluating an n-element stimulus array for the presence of target elements; these four operations will be termed elementary, similarity, cumulative, and response coding (Fig. 1). *Elementary coding* is the attempt to isolate (segment, chunk) those aspects of the ongoing sensory activity containing information about particular elements in the array. Thus, the product of this coding operation are n "sensory samples," x_1, x_2, \dots, x_n , where x_1 is the sensory activity relevant to element e_i ($i = 1, 2, \dots, n$). *Similarity coding* reduces the stimulus information in each elementary coding to a minimal form relative to the decision required of the O; in the model developed here, each similarity coding, denoted by s_i , characterizes the similarity of x_i to sensory activity typically evoked by the specific values of e_i . *Cumulative coding* integrates the n similarity codings (s_1, s_2, \dots, s_n) into a single coding, denoted by y , to which a decision rule may be applied. It is this translation of y into a response R which constitutes the final coding operation, *response coding*.

As indicated in Fig. 1, it will be assumed that each of the coding operations is influenced by "higher order memory and control processes." The nature of these higher order processes will be considered later, after an explicit quantitative model has been developed.

A Model

The basic assumptions of the model can be expressed in the following three statements: (1) Let each s_i be an independent Gaussian random variable whose distribution depends only on v_i (the value of e_i), for $i = 1, 2, \dots, n$. (2) Let

$$y = \sum_i s_i.$$

(3) Let R_1 occur iff $y > \beta$, where β is a Gaussian random variable.

Note that Statement 1 defines a relation between the value of e_i (v_i) and its similarity coding (s_i) without mentioning the elementary coding x_i . There is an interpretation of s_i in terms of x_i which is of theoretical interest, although it is not testable in the present application of the model; specifically, that

$$s_i = \ln \left\{ \frac{g_1(x_i)}{g_0(x_i)} \right\} \quad (1)$$

where $g_v(x_i)$ is the conditional distribution function on the set of possible elementary codings when v equals v_i (the value of e_i). In other words, s_i is the logarithm of the likelihood ratio $g_1(x_i)/g_0(x_i)$. Thus, it characterizes the relative similarity of x_i to target and nontarget elements and specifies *all* the information in s_i relevant to a decision regarding the value of e_i . While this interpretation of s_i will be considered later in the paper, its simpler definition in Statement 1 is sufficient here.

Statement 2 defines y as a simple sum of the s_i values; thus it is also a Gaussian random variable. This unweighted (or equally weighted) combination of similarity codings is adequate for the data to which the model will be applied in this paper. However, it should be clear that more complex combinatorial rules could be considered, for example, the type of weighted combinations of elementary information which have been proposed by Anderson (1968) in his psychological theory of information integration. Thus y could be defined as a weighted sum of the elementary codings

$$\left(y = \sum_i w_i x_i \right)$$

in situations where an O might "attend to" or "weight" the elementary codings differentially.

Finally, Statement 3 is in the form of a conventional statistical decision rule *except* that the *response criterion* β is a Gaussian random variable rather than a constant. This allows one to distinguish between the expected value and variance of β , with the former interpreted much like a conventional fixed criterion and the latter a measure of inconsistency in judgmental standards from trial to trial.

Some Properties of the Model

Certain properties of the model can be expressed most easily by letting the expected value of s_i given a nontarget element equal 0, and its expected value given a target element equal 1; i.e.,

$$E(s_i) = v_i. \quad (2)$$

This is simply an arbitrary choice of origin and unit for expressing s_i . It will also be useful to denote the variance of s_i when e_i has the value v by α_v ($v = 0, 1$) and the variance of β by π ; i.e.,

$$\text{Var}(s_i | v_i = v) = \alpha_v$$

and

$$\text{Var}(\beta) = \pi.$$

It can then be shown that the expected value and variance of the Gaussian random variable y given any particular array are

$$E(y) = t \tag{3}$$

and

$$\text{Var}(y) = t\alpha_1 + (n - t)\alpha_0, \tag{4}$$

respectively. Note that distribution of y given A_0 is simply the special case for $t = 0$: y is the sum of n Gaussian variables with expected values of 0 and variances of α_0 . In general, each target element ($v_i = 1$) in an A_1 array increases the expected value of y by 1 and its variance by α_1 . These relationships are illustrated graphically in Fig. 2(a and b).

Note that the overlapping distributions of y given A_0 or A_1 in Fig. 2b are similar to the stimulus conditional distributions of likelihood ratio found in Green and Swets (1966), except for π , the variance of β . It can easily be shown that this "variable criterion" decision process is equivalent to a "fixed criterion" one in which $E(\beta)$ is the fixed criterion and the variance of each y distribution is incremented by π (e.g., see Tanner, 1956; or Kinchla, 1969). Thus it is possible to define a *discriminability measure D* analogous to the conventional d' measure used in signal detection theory, specifically

$$D = \frac{E(y | A_1) - E(y | A_0)}{[\text{Var}(y | A_0) + \text{Var}(\beta)]^{1/2}} = \frac{t}{(n\alpha_0 + \pi)^{1/2}} \tag{5}$$

The relation between D and the O's theoretical response probabilities can be expressed most simply in terms of z_0 and z_1 , where z_0 is that value of a standard normal deviate exceeded with probability $P(R_1 | A_0)$ and z_1 is a similar transformation of $P(R_1 | A_1)$; specifically,

$$D = z_0 - Bz_1 \tag{6}$$

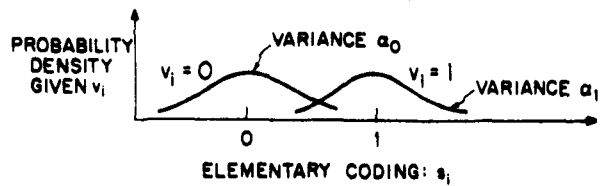
where

$$B = \frac{t\alpha_1 + (n - t)\alpha_0 + \pi^{1/2}}{n\alpha_0 + \pi} \tag{7}$$

[For the details of this derivation, see Kinchla (1966 or 1969).]

The larger the value of D , the more accurately will an O distinguish between A_0 and A_1 arrays, given the appropriate value of $E(\beta)$; i.e., D is a type of signal-to-noise ratio. The numerator in Eq. 5 characterizes the systematic difference in y attributable to the target ("signal") elements, while the denominator

(a) SINGLE OBSERVATION CASE



(b) MULTIPLE (n) OBSERVATION CASE

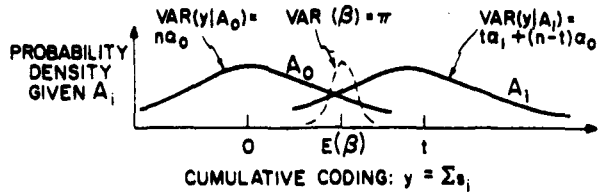


Fig. 2. Illustrative distributions of s_i conditional on the value of element e_i ($v_i = 0, 1$) and of y conditional on the type of array, A_1 or A_0 .

characterizes variability or "noise" in the perceptual process. This noise has two components, variability in y and variability in β . It will be useful to denote this "perceptual noise" by σ^2 , where

$$\sigma^2 = \text{Var}(y) + \text{Var}(\beta) = t\alpha_1 + (n - t)\alpha_0 + \pi \tag{8}$$

for either A_0 or A_1 arrays.

While D characterizes an O's ability to discriminate A_0 and A_1 arrays, a more general expression for an O's response probability given any specific array A can be defined as follows. If $P(R_1 | t)$ denotes the probability of an R_1 response given an n -element array containing t target elements ($0 \leq t \leq n$) and z_R denotes that value of a standard normal deviate exceeded with probability $P(R_1 | t)$, then it can be shown that

$$z_R = \frac{E(\beta) - t}{[\text{Var}(y | t) + \text{Var}(\beta)]^{1/2}} = \frac{E(\beta) - t}{[t\alpha_1 + (n - t)\alpha_0 + \pi]^{1/2}}; \tag{9}$$

i.e., z_R is simply the deviation of $E(y)$ from $E(\beta)$ in units of σ (Eq. 8).

In the subsequent interpretation of experimental data, it will be useful to employ simpler forms of the general model: specifically, a *one-parameter form* in which α_0 and α_1 are equal and π is zero, so that the single theoretical parameter is α ($\alpha_0 = \alpha_1 = \alpha$); and a *two-parameter form* in which α_0 equals α_1 but π is not assumed to equal zero, so that α and π are the

theoretical parameters. In both cases the factor B in Eq. 7 reduces to one, so that

$$D = \frac{t}{(n\alpha + \pi)^{1/2}} \quad (10)$$

An estimate of D, denoted \hat{D} , is therefore defined by

$$\hat{D} = \hat{z}_0 - \hat{z}_1 \quad (11)$$

where \hat{z}_i is that value of a normal deviate exceeded with a probability equal to the obtained proportion of R_1 responses to A_i arrays, $\hat{P}(R_1 | A_1)$ or $\hat{P}(R_1 | A_0)$. Furthermore, the "perceptual noise" (Eq. 8) reduces to

$$\sigma^2 = n\alpha + \pi. \quad (12)$$

An estimate of σ^2 , denoted $\hat{\sigma}^2$, is therefore given by substituting \hat{D} for D in Eq. 5 and rearranging so that

$$\hat{\sigma}^2 = t^2 \hat{D}^{-2}. \quad (13)$$

Finally, Eq. 9 also simplifies when α_0 equals α_1 to

$$z_R = Mt + K \quad (14)$$

where

$$M = -(\sigma^2)^{-1/2} \quad (15)$$

and

$$K = E(\beta)M; \quad (16)$$

i.e., z_R is a linear function of t whose *slope* (M) is a function of σ^2 and whose *intercept* is a function of $E(\beta)$ and σ^2 .

INTERPRETING EXPERIMENTAL DATA

Before describing the application of the model to visual letter detection data, it seems useful to briefly consider its original application in an auditory study.

"Set Size" and "Redundant Targets" Effects in an Auditory Experiment

In this experiment (Kinchla, 1969) the n "elements" (e_1, e_2, \dots, e_n) in each auditory "array" (A) were n , successively defined, 100-msec intervals, during each of which a 1,000-Hz tone was either added ($v_i = 1$) to a continuous level of white noise or not ($v_i = 0$). Each array was temporally defined for the O by n successive, 100-msec illuminations of a light, with 500-msec periods of darkness between illuminations. An A_1 array consisted of t "tone-plus-noise" intervals and $n-t$ "noise-alone" intervals, while an A_0 array consisted of n

noise-alone intervals. The values of n and t were fixed within each testing session at one of the seven possible combinations of n equal to 1, 3, or 6 and t equal to 1, 2, or 3 (note that t cannot exceed n). An O's task in each condition was simply to decide whether at least one tone was presented on each trial (R_1) or not (R_0). Both the noise and tone levels were adjusted during preliminary testing to yield about 90% correct responding when n and t were each one.

The proportions of R_1 responses to each type of array were used to calculate a discriminability measure, \hat{D} , for each n and t condition using Eq. 11 (the "hit" and "false-alarm" values are available in Kinchla, 1969). These measures are presented graphically for each of three Os in Fig. 3; each of the $t = 1$ points is based on 1,000 trials and each of the rest on 600 trials. The two main experimental effects are obvious: discriminability is negatively related to n , a "set size effect," and positively related to t , a "redundant targets effect."

A theoretical interpretation of these data is provided by a simple, one-parameter (α) form of the model. The solid lines in Fig. 3 were defined by selecting that value of $\hat{\alpha}$ in Eq. 10 which minimized the squared discrepancies between the observed (\hat{D}) and predicted (D) discriminability. The three $\hat{\alpha}$ values are presented in Fig. 3. These one-parameter fits account for 93%, 88%, and 84% of the variance of \hat{D} for Os 1, 2, and 3, respectively. While better fits could be obtained by estimating more parameters, the simple model provides a reasonable account of both "set size" (n) and "redundant target" (t) effects: each additional element in the array increases the variance of the cumulative coding y by the amount $\hat{\alpha}$, while each redundant target ("signal") increases the expected value of y by one (Eq. 3).

A "Redundant Targets" Effect in a Visual Experiment

Here the model is applied to some previously unpublished visual letter detection data which are part of a more extensive study by Kinchla and Collyer (in preparation). An O's task was to evaluate briefly presented four-letter arrays for the presence of one or more target letter Fs; i.e., each array consisted of four elements (letters) with A_1 arrays containing at least one F and A_0 arrays none. Os were asked to rate their confidence that the array contained at least one target letter by making one of four responses: r_1 (most confident), r_2 , r_3 , or r_4 (least confident). For our present purposes we shall simply treat the two higher confidence reports (r_1 and r_2) as a "yes" (R_1) response, and the two lower confidence responses (r_3 and r_4) as a "no" (R_0) response (this is a common practice in the analysis of rating data and is discussed in Green & Swets, 1966). Of central interest was how the number of target elements t (0, 1, 2, 3, or 4) influenced the O's response tendencies.

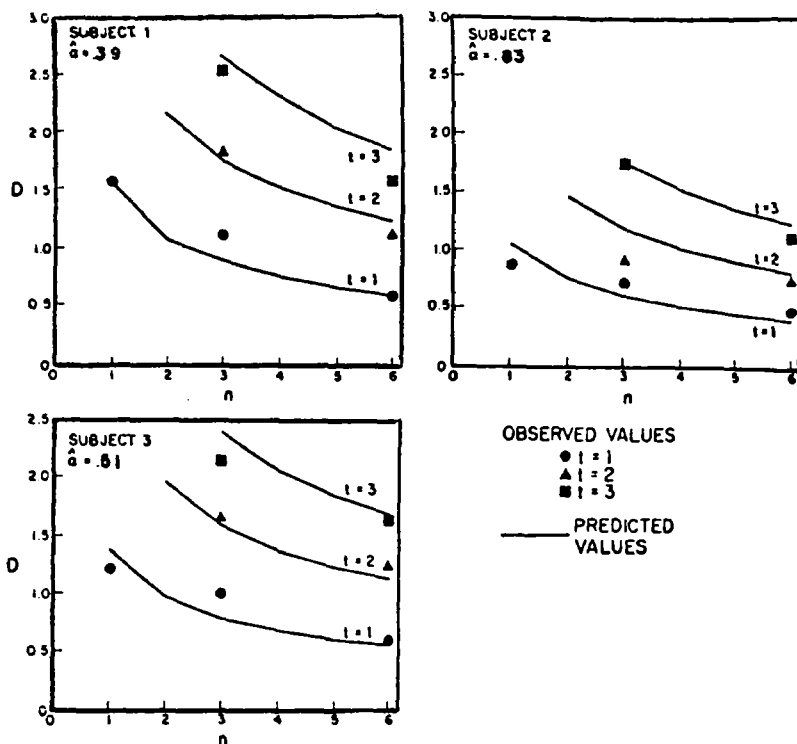


Fig. 3. Application of the model to an auditory task (Kinchla, 1969).

METHOD

Apparatus and Procedure

The stimuli were presented on a rapid-decay cathode ray tube (CRT) driven by a DEC PDP-12 computer. Each letter was defined by a specific subset of a 4 by 6 array of points as is conventional on computers of this type. The luminance of each letter was basically the same as those described in Shiffrin and Gardner (1972). Each O was seated in a darkened acoustical testing chamber and viewed the CRT from about 20 in. Each trial began with a 1-sec illumination of a fixation point in the center of the CRT, followed by a 12-msec illumination of the four-letter array, followed by a 2-sec illumination of a masking field. This masking field consisted of the full 4 by 6 point pattern from which each letter was defined. The letters were positioned in the corners of an imaginary square centered on the fixation point. Each letter subtended .75 deg vertically, .5 deg horizontally, and was centered about 1.2 deg from the fixation point. The O made a pushbutton response while the masking field was on. Each trial concluded with a .5-sec "feedback" period during which the previously presented four-letter array reilluminated. This was followed by a dark .5-sec intertrial period.

Each O was tested for 15 360-trial sessions, or a total of 5,400

Table 1
Observed Proportion of "Detect" Responses, $\hat{P}(R_1|t)$, and Predicted Probability, $P(R_1|t)$, for Each Value of t in the Kinchla and Collyer Experiment

t	Obs 1		Obs 2	
	$\hat{P}(R_1 t)$	$P(R_1 t)$	$\hat{P}(R_1 t)$	$P(R_1 t)$
t = 0	.11	.07	.29	.29
1	.19	.21	.52	.53
2	.40	.44	.75	.75
3	.70	.70	.89	.90
4	.88	.88	.97	.97

trials (not counting 8 prior practice days).

Arrays in which t equalled 0, 1, 2, 3, or 4 occurred with equal frequency in a randomly determined sequence within each block of trials. The (t) target "Fs" were positioned at random within each array along with (4-t) other nontarget letters chosen randomly from the remaining letters of the alphabet.

RESULTS

For our present purposes it will be sufficient to consider representative data from two Os. Since each performed in 15 sessions, the proportion of R_1 responses to each of the five types of array (t values) is based on approximately 1,080 trials. These five proportions for each O are presented in Table 1. Each proportion, $\hat{P}(R_1|t)$, yields a Z value (the normal deviate exceeded with a probability equal to that proportion) which is graphed as a data point in Fig. 4. The general nature of the results is clear: the larger t, the greater the probability of a "detect" (R_1) response.

Theoretical Analysis and Discussion

The solid lines in Fig. 4 indicate the theoretical function defined by Eq. 14 with \hat{M} and \hat{K} chosen so as to minimize the squared discrepancies between z_{R_1} and \hat{z}_{R_1} ; \hat{M} was $-.66$ and $-.60$, while \hat{K} was 1.45 and $.54$ for Os 1 and 2, respectively. More elegant techniques for fitting the model could be employed, including separately estimating α_0 and α_1 , but this simple method seems adequate here. The main point is clear: the model can account for virtually all of the variability in performance.¹ The effect of adding each "redundant"

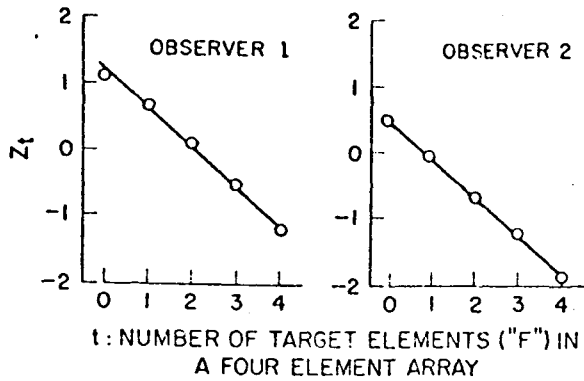


Fig. 4. Data from unpublished study by Kinchla and Collyer (in preparation).

($t > 1$) target element to an array is entirely consistent with the discriminability between single-target ($t = 1$) and no-target ($t = 0$) arrays.

Note that the estimates of M and K in Eq. 14 also yield estimates of the "perceptual noise" σ^2 and the expected response criterion $E(\beta)$, using Eqs. 15 and 16; specifically, $\hat{\sigma}^2 = 1/\hat{M}^2$ and $\hat{E}(\beta) = \hat{K}/\hat{M}$. However, note that these data do *not* allow one to isolate the two components of σ^2 in Eq. 12; i.e., to separate variability in the cumulative coding ($\alpha\pi$) from that in the response criterion (π). One method of separately estimating these two components of perceptual noise will be considered in the next section.

A "Set Size" Effect in a Visual Experiment

The model can also be used to interpret a visual letter detection experiment reported by Eriksen and Spencer (1969), specifically their "Experiment One." The elements (e_i) of their stimulus arrays were individual white-on-black-background letters which subtended .18 deg visual angle. Each letter was an uppercase T, U, or A. The O's task was to detect the occurrence of the letter A. In the present notation, e_i had a value of 0 if it was a T or U and the value of 1 if it was the "critical" letter A. An array was generated by the successive presentation of 1, 3, 5, or 9 letters positioned at various points about the circumference of a 1.3-deg-diam circle, centered on a .25-deg fixation cross. There were 10 possible letter positions equispaced about the circle, and the n (1, 3, 5, or 9) positions utilized on a specific trial were selected randomly so that each possible set of positions was (essentially) equally probable. Each letter was illuminated for about 2 msec. The dark interval between successive letters varied from one session to the next such that the letters were presented at rates of 5, 15, or 30 per sec within any one session. A critical letter was present (A_1) on a randomly selected 50% of the trial and absent (A_0) on the rest. When an array contained the letter A, it occurred with equal probability as the first, middle, or last element. (While this represents a potentially important deviation from a completely

random positioning, it appears to have had no observable effect on performance.) After viewing each array, the O indicated whether he thought an A had been presented (R_1) or not (R_0).

The performances of four Os indicated *no* significant effect of presentation rate, position of critical letter, or differences between Os. Only the number of letters in an array (n) had a statistically significant effect on performance. Table 2 presents values of $\hat{P}(R_1 | A_1)$ and $\hat{P}(R_1 | A_0)$ (the "hit" and "false-alarm" rates) averaged over the three presentation rates, critical letter position, and Os; each of the $n = 1$ proportions are based on 4,880 trials, and each of the rest on 960 trials. The general nature of the "set size" effect is obvious: while $\hat{P}(R_1 | A_1)$ is essentially invariant, $\hat{P}(R_1 | A_0)$ increases with n , indicating a drop in discriminative ability.

The two-parameter form of the model (α and π) provides an interpretation of these data. Estimates of discriminability (\hat{D}) were obtained for each value of n using the data in Table 2 and Eq. 11. These values of \hat{D} were then transformed to estimates of the "perceptual noise," $\hat{\sigma}^2$ using Eq. 13. Theoretically these estimates should be a linear function of n (Eq. 12). The extent to which this is true is indicated in Fig. 5. It is clear that there is no significant deviation from the linear function (solid line) fitted by a least-squares method to the four \hat{D} values. Note that the slope (.15) and intercept (.2) of this line constitute estimates of $\hat{\alpha}$ and $\hat{\pi}$, respectively.

Thus, in terms of the model, the set size effect can be interpreted as an additional square unit of variance (noise) in the cumulative coding y produced by each additional letter in the array. There also appears to be a fixed component of "perceptual noise" ($\hat{\pi}$) which is independent of n and interpretable as variance in the response criterion (or other fixed sources of "noise"). This fixed component was approximately as large as that attributable to each letter. The parameter $\hat{\alpha}$ can be thought of as a measure of "confusability" which makes a progressively larger contribution to the variance in y as n is increased, i.e., the basic difficulty for an O is the increasing number of opportunities for confusing nontargets with targets.

SOME ADDITIONAL CONSIDERATIONS

In the preceding applications of the model, it has been possible to ignore certain aspects of the more general theoretical process represented in Fig. 1; specifically,

Table 2
Data from Eriksen and Spencer (1969) with Estimated Values of σ^2 and \hat{D} and Predicted Values of \hat{D} Given $\alpha = .15$ and $\pi = .2$

N	$\hat{P}(R_1 A_0)$	$\hat{P}(R_1 A_1)$	σ^2	\hat{D}	Predicted D
1	.71	.13	.35	1.68	1.69
3	.74	.27	.64	1.25	1.23
5	.75	.37	1.00	1.00	1.03
9	.75	.44	1.49	.82	.81

“higher order memory and control processes” and “elementary coding.” These will be discussed briefly here, since they suggest important considerations for further application and elaboration of the model. For similar reasons it will be useful to consider some alternative letter detection paradigms and some related theoretical work.

Higher Order Memory and Control Processes

An O’s perception of a stimulus is clearly influenced by information acquired prior to its presentation. Some of this information he brings with him to the laboratory (e.g., familiarity with alphabetic characters), some is specified in his instructions (e.g., “targets will be presented on 50% of the trials”), and some is gained through practice in the actual testing situation (e.g., familiarity with the font, positioning, and timing of stimulus letters). It is the role of such a priori information which is represented by the box labeled “higher order memory and control processes” in Fig. 1. While the acquisition of this prior information (a kind of “perceptual learning” process) is obscured in studies which employ highly practiced Os, its importance is suggested by the large increase in accuracy typically seen during an O’s initial practice sessions.

One type of information an O could acquire during practice sessions is a familiarity with the sorts of “sensory samples” (elementary codes, x_i) typically evoked by target and nontarget elements; i.e., the functions $g_v(x_i)$ in Eq. 1. As he gained a better impression of these distributions, his similarity codings (s_i) would become more accurate, thereby reducing a source of “noise” in his judgments. Practice should also optimize an O’s orientation strategies and stabilize his judgmental standard (i.e., reduce π in Eq. 5, thereby increasing D).

The Concept of “Elementary Coding”

As indicated earlier, the simple model employed in this paper essentially ignores the elementary coding stage, since s_i is defined directly as a Gaussian variable determined by v_i . Why then is “elementary coding” represented in Fig. 1? Many authors have ignored the problem of “segmenting” sensory activity into elementary codes, simply referring to a “perceptual channel” (e.g., Gardner, 1973; Rumelhart, 1970; Shiffrin & Gardner, 1972) corresponding to each element in an array, without considering how these “channels” are defined. They are clearly not anatomical, in any simple sense, since their number varies with the size (n) of an array. In the experiments we have considered there was an obvious “elementary” structure to each stimulus array. Os were essentially trained to perceive this structure, i.e., to “segment” or code the sensory information in the “appropriate” way. Since this coding process was fixed and well practiced, it seemed a

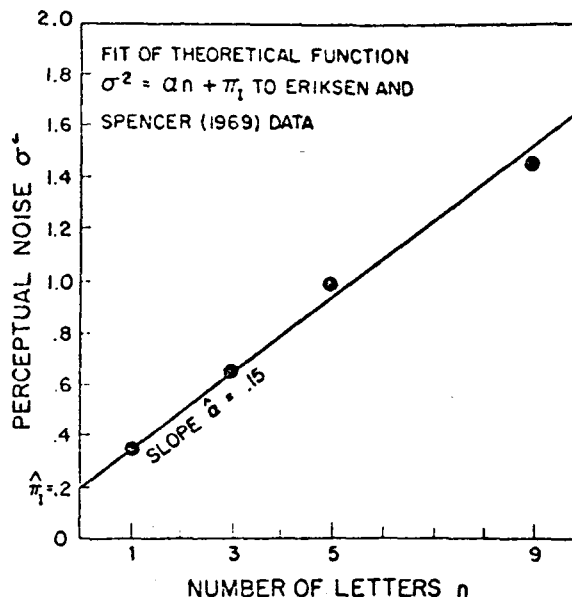


Fig. 5. Data from Eriksen and Spencer (1969).

negligible part of the perceptual process. However, consider the three stimulus patterns (I, II, and III) in Fig. 6. If an O were trained to process stimuli like Pattern I, he might initially “see” Pattern II as a “two-element” array consisting of the letters T and L, whereas if he had been trained to process stimuli like Pattern III, he might “see” Pattern II as 12 small letters. The point of this illustration is that the number of “channels” or “elementary codings” (x_i) employed to evaluate an array is basically defined by the O. This “grouping” or “segmentation” process is often obscured in studies which employ fixed format stimuli and highly practiced Os, yet it may be an important consideration when format (e.g., letter size) is varied (or unknown) or when alternative “elementary” segmentations are possible, as with Pattern II in Fig. 6. The results of some preliminary experimentation in this laboratory suggest that one’s initial organization of stimuli like Pattern II is critical when they are seen tachistoscopically (e.g., 50 msec). A predisposition to organize the stimulus into two large letters seems to reduce one’s ability to “see” the small letters radically, and vice versa, as if it were difficult to “segment” the same array in two different ways simultaneously. One could argue that a similar problem arises in the classic “cocktail party” situation, since the temporal segmentation of acoustic “sensory activity” appropriate for “hearing” one person’s voice would usually be inappropriate for another voice. Thus, there may be selective or “attentional” processes at the level of elementary coding which would not be evident in studies utilizing fixed format stimuli.

Alternative Letter Detection Paradigms

The procedure employed by Neisser (1967) and

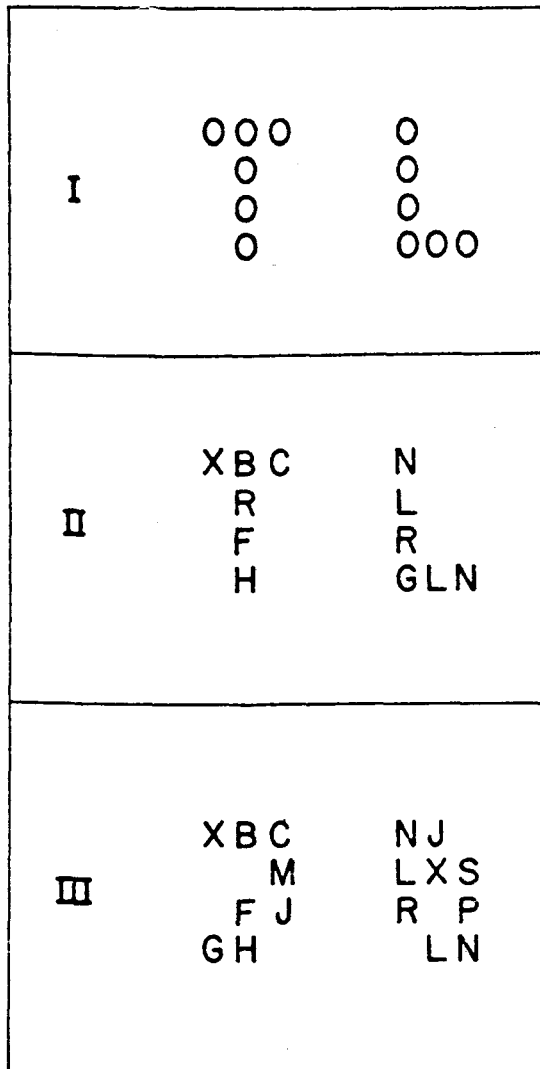


Fig. 6. Illustrative stimulus patterns, where II can be interpreted as either a 2- or 12-element array.

others to study overt visual scanning is formally similar to the tasks we have considered: the O is asked to scan a whole page of characters "from top to bottom," stopping as soon as he detects a target letter or reaches the end of the page. While the logical arguments developed here are relevant to such problems, the lack of E control of orientation (location, duration, and order of fixations) makes the data from such tasks difficult to analyze.

In both visual experiments analyzed in this paper, the stimulus elements were well spaced and equidistant from the fixation point. Care should be taken in applying the model to other situations, e.g., more complex lateral interactions may occur when letters are presented closer together, and it might be necessary to assume different "noise" (α) values when letters are presented at various distances from the fovea.

Another paradigm which has been used extensively in

letter detection studies (e.g., Estes & Taylor, 1966, 1972; Shiffrin & Gardner, 1972) has often been referred to as simply a "forced-choice" form of the "yes-no" tasks considered earlier in this paper. For example, the O could be asked to evaluate tachistoscopically presented letter arrays for the presence of a target T or F, where each array contains either a T or an F (but never both), along with other "nontarget" elements. Thus, every array contains a "target." In our terminology there are three classes of elements, the nontarget elements and the two targets. These could be said to have the "values" 0, 1, and 2, respectively. Note that in contrast to the experiments considered earlier, an O in this task cannot evaluate x_i along a single dimension of similarity (Value 0 vs Value 1), but must consider three possible values for e_i (0, 1, or 2). This "three-valued" decision problem has been referred to as a "recognition" task in the psychoacoustic literature (e.g., Luce, 1963; Shipley, 1960; Tanner, 1956) and is appreciably more complex analytically than the two-valued "yes-no detection" task. It is argued here that a more appropriate "forced-choice detection" task would be simply to present single-target ($t = 1$), multielement ($n > 1$), A_1 arrays (of the sort employed in the preceding "yes-no" experiments) on every trial, and ask the O to specify the position of the target. Since each element could still be evaluated along a single similarity dimension ($v_i = 0$ or 1), the perceptual process would be consistent with the previously defined model up to the cumulative coding stage. In the simplest possible case, the O might simply identify the element evoking the largest s_i as the target. He might also "weight" the s_i values differentially, reflecting some predisposition or "bias" to report the target in one location rather than another. A weighted decision process of this sort would be more difficult to evaluate. In any case, since no suitable data of this sort are currently available, no evaluation of this "forced-choice" variant of our model is possible here. It is described simply to indicate the potential theoretical simplicity of the two-valued "forced-choice" task compared to the three-valued "recognition" problem.

Related Work

While a complete review of all related work is beyond the scope of this paper, some particularly relevant references should be cited. As indicated earlier, one line of theoretical work has emphasized the problem of "coding" stimulus information from a rapidly decaying "iconic" memory system. The assumption that only so much information can be successfully coded before decay (or interference) terminates the coding imposes an information "bottleneck" early in the perceptual process. In contrast, a number of authors have suggested the major problem for an O is the interpretation of confusable or "noisy" subjective representation of the stimulus.

Gardner (1973) develops an "independent channels

confusion" model which is similar in many respects to the one presented here. However, Gardner only derives quantitative expressions for the probability of a correct response in a two-alternative "recognition" task; i.e., a task in which each array contains one of two alternative "critical" letters along with additional "noise" letters. Gardner assumes that the "sensory sample" derived from each stimulus letter is independently identified (as a particular critical letter or a noise letter) before the information is integrated. Basically there are at least nine conditional probabilities implied in such a process: the probabilities of each identification given each type of element. Gardner provides "an intuitive explanation" for these probabilities with a graph of three overlapping bell-shaped distributions on "an underlying T-F (similarity) axis" (T and F being the two critical letters). This axis is assumed to be divided into three regions, corresponding to the three possible "end states," by two symmetrically placed criteria. This "symmetry assumption" allows the nine conditional probabilities to be defined with only three parameters. Then, by assuming that these three parameters are the same in all "channels" and *independent of the number of letters in an array* (n), Gardner derives an expression indicating a negative relation between set size and probability correct. While this relation is of the same ordinal form as the "set size" effect, Gardner does not actually "fit" his quantitative expression to any data, nor does he estimate any theoretical parameters.

Several comments seem appropriate regarding Gardner's analysis. First, processing each sensory sample to the level of an identification prior to "integration" of the information across elements involves a greater information loss than the integration rule used in this author's model (the three-valued "end state" variable carries less stimulus information than the continuous similarity coding). Second, the assumption of symmetry in the overlapping distributions and in the placement of identification criteria is unlikely to be appropriate except in special cases. Finally, there is no a priori reason to suppose that an O would maintain precisely the same identification criteria for different set sizes, and Gardner's derivation of a set size effect depends on this assumption. In any case, Gardner's theoretical arguments are generally consistent with the model presented here, and he presents several excellent experiments supporting the "confusability" point of view.

Two other important papers are those by Estes (1972) and Shiffrin and Geisler (1973). Both emphasize the "confusability" of partially processed letters, although they each suggest a distinction between "primary detection," where enough stimulus information is obtained to allow an unambiguous identification and the processing of ambiguous or "noisy" subjective representations. The Shiffrin and Geisler model is developed in detail and is particularly interesting as a bridge between letter detection and the sort of

"scanning of short-term memory" which is associated with the work of Sternberg (e.g., Sternberg, 1969).

Both Estes and Shiffrin have dealt mainly with "recognition" tasks (the sort of "three-valued" visual task described in the preceding section). As indicated earlier such tasks *may* involve more complex coding and decision-making processes than the "two-valued" detection tasks considered in this paper. Although recognition tasks may be closer to the more complex perceptual problems encountered in reading, the simpler detection paradigms may be more useful for certain types of analysis.

The problem of elementary coding is closely related to the problem of "perceptual grouping" which was studied extensively by early Gestalt psychologists. Excellent contemporary discussions of this topic can be found in Hochberg (1973), Keel (1973), and Kahneman (1973). Of particular relevance is the recent work of Beck (1972) in vision and Massaro (1972) in audition.

REFERENCES

- Anderson, N. H. A simple model for information integration. In R. P. Abelson et al (Eds.), *Theories of cognitive consistency: A sourcebook*. Chicago: Rand McNally, 1968.
- Beck, J. Similarity grouping and peripheral discriminability under uncertainty. *American Journal of Psychology*, 1972, 85, 1-19.
- Eriksen, C. W., & Spencer, T. Rate of information processing in visual perception: Some results and methodological consideration. *Journal of Experimental Psychology Monograph*, 1969, 79(No. 2 Pt. 2).
- Estes, W. K. Interactions of signal and background variables in visual processing. *Perception & Psychophysics*, 1972, 12, 278-286.
- Estes, W. K., & Taylor, H. A. Visual detection in relation to display size and redundancy of critical elements. *Perception & Psychophysics*, 1966, 1, 9-16.
- Gardner, G. T. Evidence for independent parallel channels in tachistoscopic perception. *Cognitive Psychology*, 1973, 4, 130-155.
- Green, D. M., & Swets, J. A. *Signal detection theory and psychophysics*. New York: Wiley, 1966.
- Kahneman, D. *Attention and effort*. New York: Prentice-Hall, 1973.
- Keel, S. W. *Attention and human performance*. Pacific Palisades, Calif: Goodyear, 1973.
- Kinchla, R. A. Temporal uncertainty in multiple-observation detection tasks. Technical Report No. 2, 1966, Department of Psychology, McMaster University, Hamilton, Ont., Canada.
- Kinchla, R. A. Temporal and channel uncertainty in detection: A multiple observation analysis. *Perception & Psychophysics*, 1969, 5, 129-136.
- Luce, R. D. A threshold theory for simple detection experiments. *Psychological Review*, 1963, 70, 61-79.
- Massaro, D. W. Preperceptual images, processing times, and perceptual units in auditory perception. *Psychological Review*, 1972, 79, 124-145.
- Neisser, U. *Cognitive psychology*. New York: Appleton-Century-Crofts, 1967.
- Rumelhart, D. E. A multi-component theory of the perception of briefly exposed visual displays. *Journal of Mathematical Psychology*, 1970, 7, 191-218.
- Shiffrin, R. M., & Gardner, G. T. Visual processing capacity and attentional control. *Journal of Experimental Psychology*, 1972, 93, 72-82.

- Shiffrin, R. M., & Geisler, W. S. Visual recognition in a theory of information processing. *Perception & Psychophysics*, 1973, in press.
- Shipley, E. F. A model for detection and recognition with signal uncertainty. *Psychometrika*, 1960, 25, 273-289.
- Sperling, G. A model for visual memory tasks. *Human Factors*, 1963, 5, 19-31.
- Sternberg, S. The discovery of processing stages: Extension of Donder's method. In W. G. Koster (Ed.), *Attention and performance II*. Acta Psychologica, 1969, 30, 276-315.
- Swets, J. A. Signal detection and recognition by human observers. In J. A. Swets (Ed.), *Contemporary readings*. New York: Wiley, 1964.
- Tanner, W. P., Jr. Theory of recognition. *Journal of the Acoustical Society of America*, 1956, 33, 1233-1244.

NOTE

1. An alternative to the response rule employed in this paper (R_1 if $y \geq \beta$) would be to *respond R_1 if at least one value of s_i exceeds a particular criterion value*. Green and Swets (1966) consider such a rule in a "decision threshold model." They indicate that this decision rule is less efficient than the one employed in this paper, leading to less accurate responding at each value of n than would be predicted from the model used here.

(Received for publication June 28, 1973;
revision received September 21, 1973.)