Detecting a target letter in briefly presented arrays:
A confidence rating analysis in terms of a weighted additive effects model

R. A. KINCHLA and C. E. COLLYER
Princeton University, Princeton, New Jersey 08540

A visual experiment is reported in which Os rated their "confidence" that briefly presented four-letter arrays contained one or more target-letter Fs. Their responses are interpreted as the product of a weighted "additive" combination of partial or "noisy" subjective representations of the four letters in each array.

An O's report that he "saw" a specific target letter in a briefly presented letter array can be interpreted in various ways. One conception is that the perceptual processing ("coding") of the individual letters is essentially "all-or-none," so that the response reflects either a "true detection" or a "pure guess." Another conception is that each letter may be partially processed so that the response is a sort of statistical decision based on partially processed or "noisy" subjective representations of each letter. The "all-or-none" conception, which was quite influential during the early research on this problem, did not encourage the use of "confidence rating" responses in letter-detection tasks, since the O ostensibly has only two levels of confidence, "completely certain" or "just guessing." (Rating responses have been used in an attempt to distinguish "high-confidence" and "guessing" responses; e.g., Wolford, Wessel, & Estes, 1968.) On the other hand, the use of confidence ratings is quite consistent with the "statistical decision" point of view (e.g., Green & Swets, 1966). This paper presents the results of a study in which Os indicated their "confidence" that a tachistoscopically presented array of four letters contained one or more target-letter Fs. A statistical decision model for multiple "noisy" observations (Kinchla, 1969, 1974) is used to interpret the data.

Before proceeding further, it will be useful to introduce some notation for describing the type of letter-detection task we shall consider. An n-element (n-letter) array will be denoted by A, and each of the constituent elements by $e_i$, $i = 1, 2, \ldots, n$. Each Element $e_i$ will be said to have a "value," $v_i$, of one if $e_i$ is a target element, or a value of zero if it isn't ($v_i = 0, 1$). An O's task is to distinguish between arrays containing only nontarget elements, termed $A_0$ arrays, and those containing at least one target element, termed $A_1$ arrays. He does this by making a confidence rating $R$ equal to 1, 2, 3, or 4 each time an unknown array is presented, with larger values of $R$ indicating progressively more confidence that the array contained some targets; i.e.,

that it was an $A_1$ rather than an $A_0$ array.

In some instances, it will be useful to distinguish the specific ordering of the $t$ target and $n-t$ nontarget elements in an array, A. Since there are exactly $t!$ possible orderings of these elements, a specific ordered array can be denoted by $(A_t^j)$ with $j = 1, 2, \ldots, t!$. Furthermore, $v_{ij}$ can denote the value of Element $e_i$ in Array $(A_t^j)$.

**THEORY**

The quantitative model to be utilized in the analysis can be interpreted as a "reductive coding process"—a series of coding operations designed to isolate those aspects of one's sensory activity relevant to a particular decision. This theoretical process is developed in detail elsewhere (Kinchla, 1974) and will be only briefly summarized here to provide a conceptual framework for the more abstract assumptions of the model. In general, one could consider four successive coding operations leading to an O's response: an elementary coding, which isolates (segments, chunks, groups) ongoing sensory activity into $n$ "sensory samples" denoted by $x_1, x_2, \ldots, x_n$, where $x_i$ is that component of the sensory activity influenced by Element $e_i$; a similarity coding, which assigns a measure, $s_i$, to each $x_i$ characterizing its similarity to sensory samples typically evoked by each possible value of $e_i$ (something like a likelihood ratio); a cumulative coding, which integrates the similarity codings into a summary code denoted by $y$; and finally, a response coding which translates the cumulative code, $y$, into a value of the response variable, $R$.

For our present purposes, we shall consider only the latter three coding operations (similarity, cumulative, and response coding). Elementary coding would be an important consideration in experiments which employed highly variable display formats or unpracticed Os; for example, an O might have difficulty in "segmenting" an array when he didn't know whether it contained four small letters or one large one. However, in the present experiment, letter arrays had a fixed format (a 2 by 2 matrix of four letters) and the Os were highly practiced.
Thus, we shall assume that the efficiency of elementary coding was essentially "asymptotic" and relatively constant from trial to trial.

The Model

The assumptions of the model can be summarized in the following three statements: (I) Each similarity coding, $s_i$, is a Gaussian random variable with variance $\sigma$ and an expected value determined by the value of $e_i$. (II) The cumulative coding, $y$, is defined as

$$y = \sum w_i s_i,$$

where each $w_i$ is nonnegative and

$$\sum w_i = 1.$$ (III) The confidence response, $R$, will have a value greater than $r$ if $y$ exceeds a "judgmental criterion," $\beta_r$, whose value is a positive monotone function of $r$ for $r = 1, 2, \text{ or } 3$.

Note that although the expected values of $s_i$ are conditional on the values of $e_i$, there is some "noise" or variability in the similarity coding, denoted by $\sigma$. It is this "perceptual noise" which limits the O's ability to discriminate between $A_1$ and $A_0$ arrays.

This model differs from the one presented in Kinchla (1969, 1974) in an assumption of a weighted cumulative coding; an unweighted integration rule was employed in the earlier model. The more general assumption that an O may "weight" one part of an array more than another in determining his response is one way of introducing an "attentional" mechanism into the perceptual process. It is similar to the "weighted information integration" rules considered in Anderson (1968). It will be shown that certain aspects of an O's performance are relatively independent of this weighting assumption, whereas other are not.

Properties of the Model

The following derivations are facilitated by letting the expected value of $s_i$, given $e_i$, $E(s_i \mid e_i)$, equal $v_i$ (0 or 1). This simply represents a convenient choice of origin and unit for $s_i$ and does not determine any testable properties of the model. Furthermore, since many properties of the model have been derived elsewhere (e.g., Kinchla, 1969, 1974), only those derivations unique to the weighted cumulative coding assumption will be developed in detail here.

First of all, it will be useful to consider the distribution of $y$, given an ordered array $(A_i)$ which will be denoted $g_j(y)$, for $j = 1, 2, \cdots (\binom{n}{1})$. Since $y$ is the sum of $n$ Gaussian random variables

$$y = \sum w_is_i,$$

it will also be Gaussian with an expected value and variance defined as follows

$$E(y \mid A_i) = \sum w_i v_i$$

and

$$\text{Var}(y \mid A_i) = \sum w_i^2.$$

where $v_i$ denotes the value of Element $e_i$ in Array $(A_i)$.

The probability of a confidence rating greater than $r$, given a particular ordered array, $P(R > r \mid A_i)$, is simply the probability that the $y$ value evoked by that array is greater than $\beta_r$ for $r$ equal to 1, 2, or 3. Note that in this case $y$ is the sum of $n$ independent Gaussian random variables

$$y = \sum w_is_i$$

whose expected values equal $v_i$ (0 or 1), and which each have variance $\sigma$. Thus, it follows that

$$P(R > r \mid A_i) = \int_{\beta_r}^{\infty} g_j(y) \, dy$$

where $g_j(y)$ is a Gaussian density function on $y$ with parameters $E(y \mid A_i)$ and $\text{Var}(y \mid A_i)$ (Eqs. 1 and 2).

The probability of a confidence rating greater than $r$, given simply the number of targets in the array, $P(R > r \mid t)$, must be defined somewhat differently, since the distribution of $y$ in this case is not necessarily Gaussian (y is a mixture rather than a convolution of the various Gaussian distributions evoked by the different $A_i$). However, an expression for $P(R > r \mid t)$ can be written as a function of $P(R > r \mid A_i)$ and the probabilities of each ordered array, $P(A_i)$:

$$P(R > r \mid t) = \sum P(A_i)P(R > r \mid A_i)$$

$$= \sum P(A_i)\int_{\beta_r}^{\infty} g_j(y) \, dy$$

or, if the $\binom{n}{t}$ orderings of the $t$ targets and $n-t$ nontarget elements are equally likely,

$$P(R > r \mid t) = \left[1/(\binom{n}{t}) \right] \sum_{j} P_j \int_{\beta_r}^{\infty} g_j(y) \, dy,$$

where in each case the parameters of $g_j(y)$ are those defined in Eqs. 1 and 2.

While we have defined the probability of a rating response greater than some value ($R > r$), the probabilities of specific responses ($R = r$) can be defined either by selecting the appropriate limits of integration in Eq. 4a or simply as
Table 1

\( \hat{P}(R = r | t), \) Theoretical \( P(R = r | t) \) Based on the Equal Weighting Model in Parentheses, and Parameter Estimates for Each Observer

<table>
<thead>
<tr>
<th></th>
<th>O R Y.</th>
<th></th>
<th></th>
<th>O S A.</th>
<th></th>
<th></th>
<th>O M R.</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>( t )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>.45</td>
<td>.26</td>
<td>.16</td>
<td>.13</td>
<td>.65</td>
<td>.24</td>
<td>.07</td>
<td>.04</td>
</tr>
<tr>
<td></td>
<td>(.44)</td>
<td>(.27)</td>
<td>(.24)</td>
<td>(.05)</td>
<td>(.63)</td>
<td>(.30)</td>
<td>(.06)</td>
<td>(.01)</td>
</tr>
<tr>
<td>1</td>
<td>.24</td>
<td>.24</td>
<td>.31</td>
<td>.21</td>
<td>.32</td>
<td>.50</td>
<td>.14</td>
<td>.04</td>
</tr>
<tr>
<td></td>
<td>(.22)</td>
<td>(.25)</td>
<td>(.38)</td>
<td>(.15)</td>
<td>(.37)</td>
<td>(.41)</td>
<td>(.20)</td>
<td>(.02)</td>
</tr>
<tr>
<td>2</td>
<td>.07</td>
<td>.18</td>
<td>.38</td>
<td>.37</td>
<td>.18</td>
<td>.43</td>
<td>.33</td>
<td>.06</td>
</tr>
<tr>
<td></td>
<td>(.09)</td>
<td>(.16)</td>
<td>(.42)</td>
<td>(.33)</td>
<td>(.16)</td>
<td>(.39)</td>
<td>(.36)</td>
<td>(.09)</td>
</tr>
<tr>
<td>3</td>
<td>.02</td>
<td>.09</td>
<td>.38</td>
<td>.31</td>
<td>.05</td>
<td>.26</td>
<td>.50</td>
<td>.20</td>
</tr>
<tr>
<td></td>
<td>(.03)</td>
<td>(.07)</td>
<td>(.34)</td>
<td>(.56)</td>
<td>(.05)</td>
<td>(.25)</td>
<td>(.46)</td>
<td>(.24)</td>
</tr>
<tr>
<td>4</td>
<td>.01</td>
<td>.03</td>
<td>.32</td>
<td>.65</td>
<td>.02</td>
<td>.09</td>
<td>.38</td>
<td>.51</td>
</tr>
<tr>
<td></td>
<td>(.01)</td>
<td>(.02)</td>
<td>(.20)</td>
<td>(.77)</td>
<td>(.01)</td>
<td>(.11)</td>
<td>(.40)</td>
<td>(.48)</td>
</tr>
</tbody>
</table>

\( \hat{\alpha} \) = 0.67, \( \hat{\beta}_1 \) = 0.06, \( \hat{\beta}_2 \) = 0.23, \( \hat{\beta}_3 \) = 0.79

\[
P(R = 1) = 1 - P(R > 1) \quad (5a)
\]

\[
P(R = 2) = P(R > 1) - P(R > 2) \quad (5b)
\]

\[
P(R = 3) = P(R > 2) - P(R > 3) \quad (5c)
\]

\[
P(R = 4) = P(R > 3). \quad (5d)
\]

In general, then, in order to calculate specific response probabilities, it is necessary to specify at least \( n + 1 \) theoretical parameters: the "coding noise," \( \alpha \), a "judgmental criterion," \( \beta_1 \), and any \( n - 1 \) of the \( n \) "weighting parameters," \( w_i \) (since \( \sum w_i = 1 \)). It will be useful to consider a simpler form of this general model in which each \( w_i \) is assumed to be equal; i.e., \( w_i = 1/n \) for \( k = 1, 2, \cdots, n \). This will be referred to as the "equal weighting model" and has only the theoretical parameters, \( \alpha \) and \( \beta_1 \). In this special case, the Gaussian distributions of \( y \) values evoked by each \( (A_j) \) are identical, with parameters (by Eq. 1)

\[
E(y | (A_j)) = \sum w_i v_{ij}
= (1/n) \sum v_{ij}
= t/n
\]

and (by Eq. 2)

\[
\text{Var}(y | (A_j)) = \alpha \sum_i w_i^2
= \alpha \sum (1/n)^2
= \alpha/n,
\]

for \( j = 1, 2, \cdots (n_i) \).

THE EXPERIMENT

Three Os evaluated four-letter (2 by 2) arrays for the presence of a target letter, \( F \). Of principle interest was how the number of target letters in each array, \( t \) equal to 0, 1, 2, 3, or 4, would influence their confidence rating, \( R \).

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 characteristics of each letter were 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 one of four pushbutton responses while the masking field was on: R equaled 1 ("lowest confidence target present"), 2, 3, or 4 ("highest confidence"), with the O instructed to "try to use each category about equally often." Each trial concluded with a 3-sec "feedback" period during which the previously presented four-letter array was 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 trials (not counting 8 prior practice days).

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

The stimulus presentation schedule was chosen to provide equal amounts of data at each \( t \) value. This meant that the occurrence of a target in one position was not independent of its occurrence at another on any one trial, although targets did occur equally often at each position over an entire session. To
have obtained such “independence” would have required, for example, \( P(\omega_2 = 1) = .5 \) for all four positions on each trial, with \( t \) then equal to 0 through 4, respectively, on .06, .25, .38, .25, and .06 of the trials. Thus, virtually no data would be obtained for the \( t \) equals 0 or 4 conditions (similar problems arise with other schedules of this type). While the schedule we employed was more satisfactory in this respect, it might be argued that the preceding “independence” schedule would be even more likely to induce an “additive” integration of information, i.e.,

\[ y = \sum_i w_i. \]

Results

The proportions of each rating response given each value of \( t \), \( P(R = r \mid t) \), are presented numerically for each \( O \), in Table 1 and as data points in Fig. 1, with each proportion based on approximately 1,080 stimulus presentations. A chi-square analysis indicated a statistically significant \((p < .001)\) effect of \( t \) for each \( O \), with each proportion positively related to \( t \).

Response tendencies as a function of the \( t \) targets were placed within the array were examined by calculating the proportion of responses in the two higher confidence categories \((R > 2)\) for each of the 16 ordered arrays, i.e., \( P(R > 2 \mid \omega \beta_j) \) for each \( j \) at each value of \( t \). This particular partition of the rating responses was chosen in order to obtain maximally stable probability estimates for each of the three \( O \)s. These are presented in Table 2 and Fig. 2. Note that there is only one ordering for \( t \) equal to 0 or 1, four orderings for \( t \) equal to 1 or 3, and six orderings for \( t \) equal to 2. A chi-square analysis indicated a statistically significant \((p < .001)\) variation in response tendencies as a function of target location for all three \( O \)s. The data points in Fig. 2 indicate how variation in response tendencies within a particular value of \( t \) was substantial for Os R.Y. and S.A., and almost negligible for O.M.H. except for one point (Array F, which contained two targets in the bottom row).

### Theoretical Analysis and Discussion

We shall first interpret the data in terms of the equal weighting model, since it is both simple and representative of a class of models which depict the \( O \) as being equally influenced by each element in an array. The extent to which this model accounts for the variation in \( P(R > r \mid t) \) was evaluated by selecting those values of \( \alpha, \beta_1, \beta_2, \) and \( \beta_3 \) which minimized the chi-square statistics based on the observed proportions in Table 1 and those predicted by Eqs. 3, 4, 6, and 7. These “theoretical” values of \( P(R > r \mid t) \) are presented numerically in Table 1, along with the “optimal” parameter estimates, and as solid curves in Fig. 1. While there is a statistically significant \((p < .01)\) residual variance in each \( O \)’s data, the model clearly accounts for the major share of variation in each performance. All three \( O \)s demonstrated similar levels of discriminability (\( \alpha \) was about the same for each) but somewhat different

### Table 2

<table>
<thead>
<tr>
<th>( O )</th>
<th>( &lt;A_j&gt; )</th>
<th>( V_1, V_4, V_2 )</th>
<th>( \alpha )</th>
<th>( \beta_1 )</th>
<th>( \beta_2 )</th>
<th>( \beta_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>R.Y.</td>
<td>.29</td>
<td>.11</td>
<td>.18</td>
<td>.18</td>
<td>.18</td>
<td>.18</td>
</tr>
<tr>
<td>S.A.</td>
<td>.66</td>
<td>.66</td>
<td>.66</td>
<td>.66</td>
<td>.66</td>
<td>.66</td>
</tr>
<tr>
<td>M.H.</td>
<td>.64</td>
<td>.64</td>
<td>.64</td>
<td>.64</td>
<td>.64</td>
<td>.64</td>
</tr>
</tbody>
</table>

### Parameter Estimates

\[
\begin{align*}
\hat{P} & = .43 \\
\hat{w} & = .11 \\
\hat{w_1} & = .40 \\
\hat{w_2} & = .08 \\
\hat{\alpha} & = .41 \\
\hat{\beta_1} & = .23 \\
\hat{\beta_2} & = .25 \\
\hat{\beta_3} & = .30
\end{align*}
\]
judgmental criteria ($\beta_k$). They all tended to use the higher confidence responses more than the lower, despite our instructions to "use them equally often." This, of course, could be expected, since the "feedback" they received after each trial clearly indicated that "no-target" ($A_0$) arrays were seldom presented.

While the preceding analysis appears to support the "equal weighting model," a different picture emerges if one considers the data in Fig. 2. If each ordering of $t$ targets and $n-t$ nontargets was perceptually equivalent, then $P(R > 2 | A_1)$ would be the same for each $A_1$ containing $t$ targets. Both the chi-square analysis cited earlier and visual inspection of Fig. 2 indicates that there was considerable variation in response tendencies as a function of target placement, particularly for Os R.Y. and S.A.

The general model was fitted to the data presented in Table 2 and Fig. 2 by selecting values of $\alpha$, $\beta_2$, $w_1$, $w_2$, and $w_3$ (this implies $w_4$ since $\Sigma w_i = 1$), which minimized the discrepancies (as measured by a chi-square statistic) between $P(R > 2 | A_i)$ and those predicted by Eq. 3. These parameter estimates together with the predicted values of $P(R > 2 | t)$ are presented in Table 2. The solid points in Fig. 2 (which are connected within a particular $t$ value) indicate the degree to which the more general "unequal weighting" model accounts for each O's performance. The model is most successful with Os R.Y. and S.A., where it reduced the residual variance about the predicted values by 80% and 30%, respectively, compared to the predictions of the equally weighted model; note that $P(R > 2 | A_1)$ equals $P(R > 2 | t)$ for each $A_1$ in the equally weighted model. Each of these Os appeared to "weight" the two elements on the left of the array much more than the other elements in determining their response (their values of $\tilde{w}_1$ and $\tilde{w}_3$ are much larger than $\tilde{w}_2$ and $\tilde{w}_4$ in Table 2). The model did equally well for O.M.H. except for the two discrepant values of $P(R > 2 | A_i)$, which are apparent in Fig. 2 (Arrays A and F). These correspond to arrays which contained no targets ($A$), which the subject seemed to be unusually accurate in classifying, and an array containing only two targets in the bottom row (F).

However, her response tendencies, given other arrays containing two targets in the bottom row ($L$, $M$, and $P$), are entirely consistent with the model. Thus given the total pattern of data from all three Os, only the unusually accurate classification of no-target ($t = 0$) arrays by O.M.H. seems worth noting. While it could represent a meaningful departure from the theoretical perceptual process, it could also be an idiosyncratic feature of one O's performance, an O whose noise level ($\sigma$) was appreciably larger than that of the other two Os.

One implication of our analysis is that statistics such as $P(R > 1 | t)$ plotted in Fig. 1 are relatively insensitive to the effects of differential "weighting"; the values of $P(R > 2 | t)$ predicted by the unequal weighting model [these are available from Table 2 as the average $P(R > 2 | A_i)$ for each $t$] do not provide a significantly better fit to the observed values of $\tilde{P}(R > 2 | t)$ (the data points in Fig. 1) than does the equal weighting model; in fact, they are virtually identical. Thus, some caution should be observed in evaluating data where performance is solely dependent on the number of targets in the array. Only an analysis which considers an O's responses to particular arrangements of target and nontargets would be sensitive to the effects evident in Fig. 2, effects which seem to require something like the model proposed here, in which Os differentially "weight" the information obtained from various elements in the array.

CONCLUSION

The major implication of our analysis is that the simple model we have employed provides a reasonable interpretation of confidence rating data in this sort of letter-detection task. Our assumption of a "weighted integration" rule,

$$y = \sum_{i} w_i s_i,$$

is one way of interpreting the variation in an O's
response tendencies produced by the positioning of targets in the array. There are, of course, a number of alternative possibilities one could explore. The Os may have obtained more information from one part of the array than the other; e.g., the “noise” parameter might not have been equal for each s. Positional variations in sensitivity or acuity could be produced by biased orientation, nonhomogeneity in retinal structure (see Estes & Wolford, 1971), or by “attentional” mechanisms at the level of elementary coding (such as those suggested by the $\Theta_i$ parameters in Rumelhart’s 1970 model). The analysis presented here is not intended as a test of such alternatives, merely a demonstration of one theoretical approach. A distinction between variations in the amount of sensory information obtained from various parts of an array, as opposed to differentially weighting (“paying more or less attention to”) the information from each part, requires explicit models of each process. The present model is an attempt to characterize a differential weighting process.

It seems particularly significant that one could view our experiment as a factorial design with the value of each element defining a factor, i.e., we evaluated an O’s response tendencies, given each combination of these four factors (each combination of $v_1$, $v_2$, $v_3$, and $v_4$). Characterized in this way, our analysis is an example of “conjoin” (Krantz, Luce, Suppes, & Tversky, 1971) or “functional” (Anderson, 1968) measurement; we essentially asked if a specific transformation of the response probabilities (the value of a normal deviate exceeded with that probability) would demonstrate an additive effect of the four stimulus factors. This follows from the assumption that the response is an “additive” function of four Gaussian random variables,

$$y = \sum w_i s_i.$$

While a complete development of this point is beyond the scope of the present paper, it is mentioned here to indicate the rather general nature of our analysis. In fact, most “signal detection” or “Thurstonian” models lead to analyses of this sort, although this is not generally recognized; e.g., “ROC” plots are a way of demonstrating the “additivity” of “signal intensity” and “instructional” (“response biasing”) factors. An excellent discussion of this issue is presented in Anderson (1972).

REFERENCES


(Received for publication January 7, 1974; revision received March 23, 1974.)