OBSERVATIONS

A Model of Automatic Attention Attraction When Mapping Is Partially Consistent

Richard M. Shiffrin and Mary P. Czerwinski
Indiana University

A model is described to account for the data of Durso, Cooke, Breen, and Schvaneveldt (1987). On the basis of the relative frequency of an item's presentation as a target, the item develops an automatic tendency to attract attention. When stimuli are then displayed, each calls the attention system to a degree determined by its present strength. We assume that attention eventually drifts to the strongest stimulus (which is then given as a response), but in a time determined inversely by the difference in strength between the two strongest stimuli. A version of this model in which the strengths were freely estimated parameters predicted the various elements of the data with good accuracy. In other versions of the model, strength values were derived from assumptions concerning the learning of automatism. Two of these models, quite different in character, captured the major qualitative features of the data. Further empirical tests of the models are suggested.

Durso, Cooke, Breen, and Schvaneveldt (1987; see also Pashler & Badgio, 1985) displayed two, three, or four different digits, the largest of which could be any digit from 4 to 9, and trained subjects extensively to press a key corresponding to the largest presented digit (the target). The main results are depicted in Figures 1 and 2.

The left hand panel of Figure 1 gives the latencies as a function of the number of displayed digits (the set size function) for early sessions of training (1-5) or late sessions of training (27-31).

In standard search paradigms, a linear set size function with large slope is usually interpreted in terms of serial search models, in which the slope is taken to reflect a comparison of automatic attraction of attention. When stimuli are then displayed, each calls the attention system to a degree determined by its present strength. A reduction in slope with practice is commonly thought to signal a switch from a more serial, limited capacity process to a more parallel, unlimited capacity process. Although the present “find the largest digit” task differs in significant ways from standard search tasks, Durso et al. (1987) suggested that the reduction of slope with practice signals a switch to a more parallel, automatic process.

Because for digits 4 through 8 the response “largest” is not consistently assigned (sometimes the digit is largest, and sometimes not), the authors concluded that “perfectly consistent mapping is not a necessary precondition for reduction of visual search rate” (Durso et al., 1987, p. 223). Although we agree with this conclusion, the form of the statement may be somewhat misleading. First, previous research had already shown that perfect consistency is not needed for a flattening of search functions to occur (e.g., Schneider & Fisk, 1982). Second, previous presentation of automatism theory (e.g., Shiffrin & Dumais, 1981; Shiffrin & Schneider, 1977) posited that automatism (and hence lowering of slopes) would develop as a function of the degree of consistency. Third, previous research (Dumais, 1979; Shiffrin & Dumais, 1981) suggests a mechanism by which the present results may be the outcome of automatic attraction of attention. Fourth, this mechanism is a parallel one in which search “rate” in the sense of serial comparisons is irrelevant. This article will consider several models based on this mechanism and will evaluate them on their ability to capture the essential aspects of the data of Durso et al. (1987).

Before turning to models based on automatic mechanisms, we shall consider two other notable features of the data and their implications for limited capacity search models. First, the right panel of Figure 1 depicts strong effects of serial position, especially in the early data: digits 4 and 9 produce rapid responses when targets; intermediate target digits produce slower responses, the function resembling an inverted U overall. Second, as depicted in Figure 2, pronounced effects of split are seen, especially in the early data: the larger the difference between the two largest digits (the split), the faster is the reaction time.

These data are not easy to deal with in the framework of limited capacity models. For ease of exposition, these limited-capacity models will be described in terms of serial comparison processes. The first model we consider is a serial check of each item in the display, each being compared with the largest found thus far (and replacing it if larger). The search would terminate when every item had been checked, thereby predicting set size effects. However, this model is not able to
predict serial position effects, or split effects. The model could be modified by assuming search terminates whenever digit 9 is encountered, but for trials not containing a 9, the modified and original models share the same mispredictions.

An alternative serial model would assume subjects search sequentially for 9, then 8, then 7, and so on, each search being serial (and either exhaustive or terminating) through the displayed digits until a match is found. This model would predict set size effects, and also serial position effects. However, the model predicts latency to increase monotonically as the largest digit becomes smaller, a prediction at odds with the data. Furthermore, the model does not account for the split data. Finally, notice that the late data in Figure 1 depict an overall reduction in response times, as well as a general "flattening out" of set size and serial position functions. The various serial models provide no natural account for the changes that occur with practice. It is because these serial models fare so poorly that serious consideration must be given to parallel, automatic search mechanisms. Pashler and Badgio (1985), using a similar task, but manipulating visual display quality and decision factors along with display size, also suggested that a serial model might not be appropriate to characterize such tasks, and discussed in a general way the possibility that parallel mechanisms might be used.

The model we have in mind, based on the suggestions of Dumais (1979) and Shiffrin and Dunais (1981), is quite simple. Owing to the relative frequency of trials on which a given digit is largest, each digit develops a tendency to attract attention. This tendency can be represented as a value on a positive numerical scale. The digit 9 will have the greatest tendency, because it is always a target, and 1, 2, and 3 will have the smallest tendencies, because they are never targets. The other digits will be intermediate, in order of their numerical value. Let each digit have a strength distribution at a given point in training, representing its tendency to attract attention. Let the mean for digit \( i \) be denoted \( M(i) \), and the variance \( V(i) \).

Assume when digits are presented on a trial, that each produces a "call" to a limited attention process, the strength of the call being represented by a sample from the respective strength distribution. Assume that attention eventually drifts to the largest of these samples, but that the time to reach the largest sample is a function solely of the difference in value between the two largest samples. This difference is the theoretical "split" and is denoted \( S \). The response time we assume to be a monotonic function of \( S \).

To apply the model, a number of simplifying and arbitrary assumptions must be made. Assume that all the distributions of strengths are gaussian, with equal variances, \( V \). Somewhat arbitrarily, also assume that

\[
RT = B + a \exp(-bS).
\]

Considering only the early or late data, this model has 12 free parameters (nine means for the nine digits, one of which may be set arbitrarily at any convenient value, the three parameters in Equation 1, and the value of \( V \)). A simulation program was written to produce pseudodata, and allow us to

![Figure 1](image-url)
derive predictions for the accuracy and latency data. Unfortunately, the computer time needed to derive reliable predictions was too long to enable us to search the parameter space for a best fit. We therefore simplified the model further by setting \( V = 0 \). In this case, analytical predictions were easy to obtain. Of course, the simplified model predicts no errors, so only the latency data were fit.

A minimum least-squares criterion was used to find a best fit, with all the points given in Figures 1 and 2 weighted equally. Because no error data were considered, a restriction was placed on the parameters \( M(i) \) such that for \( j \) greater than 2, \( M(i) \) had to be at least 1.0 greater than \( M(j) \) if \( i \) was greater than \( j \). (This restriction ensures that in the more general model with small values of \( V \), attention will usually end up drifting to the largest digit.) The parameters were estimated simultaneously for the early and late data, under the assumption that the reaction time mapping parameters (\( a \) and \( b \)) were the same for the early and late data; \( M(I) \) was set equal to zero, and the base time parameter (\( B \)) and the other strength parameters (\( M(i) \), \( i = 2, 9 \)) were allowed to differ for the early and late data. The data and the predictions are given in Figures 1 and 2, and the best fitting parameter values and least sum of squared deviations are given in Table 1.

It seems clear that the qualitative aspects of the data are captured by this model, as are most of the quantitative details.

A few features of the predictions require comment. First, in the early-session data, serial position functions with fast responses for digits 4 and 9 are predicted for the following reasons. The target 4 must have 1s, 2s, and 3s as competitors; these have small strength values, and so relatively large theoretical splits are guaranteed. The target 5, for example, will quite often have a 4 as a competitor; when this happens the theoretical split is small, and a large reaction time occurs. It is even simpler to explain fast responding for 9: Whenever 9 is a target, the mean theoretical split will be at least that corresponding to the difference between 9 and 8, which is already quite large. Second, the display size predictions arise out of a peculiarity of the experimental design. Predicted response time depends on theoretical split only, and hence on the target and the empirical split, but not on display size per se. However, larger display sizes tended to produce smaller splits on the average, because of the way the study was designed. Although Durso et al., 1987, did not report display size data with split and target held constant, they did mention that display size effects disappeared in the late data when split was greater than two, a result that is predicted by our model with its best fitting parameters. Third, the split predictions are a direct consequence of the model's assumptions, because response speed is assumed to be a monotonic function of theoretical split, which is highly correlated with empirical split.

\[\text{Note, however, that if the strength values for digits 1, 2, and 3 were close to those for digits 4, 5, and so on, then the downturn in the serial position function at the smallest digits would not be predicted.}\]
Table 1

Best Fitting Parameter Values for the Version of the Model With Freely Estimated Strengths

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Early</th>
<th>Late</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>572</td>
<td>511</td>
</tr>
<tr>
<td>a</td>
<td>254</td>
<td>254</td>
</tr>
<tr>
<td>b</td>
<td>.067</td>
<td>.067</td>
</tr>
<tr>
<td>M(1)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>M(2)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>M(3)</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>M(4)</td>
<td>24</td>
<td>240</td>
</tr>
<tr>
<td>M(5)</td>
<td>25</td>
<td>263</td>
</tr>
<tr>
<td>M(6)</td>
<td>26</td>
<td>287</td>
</tr>
<tr>
<td>M(7)</td>
<td>28</td>
<td>302</td>
</tr>
<tr>
<td>M(8)</td>
<td>42</td>
<td>327</td>
</tr>
<tr>
<td>M(9)</td>
<td>66</td>
<td>356</td>
</tr>
<tr>
<td>V</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SSD*</td>
<td>33.6</td>
<td>3.3</td>
</tr>
</tbody>
</table>

*Sum of squared deviations (ms²/1000).

Although our model does a reasonable job of predicting display size effects on the basis of averaging over different splits, it is possible that further research will show display size effects to be present even with split held constant. If so, it might be possible to argue that items lower than the second largest are also affecting the response time. Alternatively, it may be that subjects could have been supplementing a fast, holistic processing of the display with a more serial, controlled search, with the first process reaching completion governing the response. Implementations of these ideas into theory will have to await further experimentation.

The effects of practice are predicted mainly on the basis of the much larger strength values estimated for the late sessions. Because the splits are larger, the reaction time predictions by Equation 1 tend to approach a constant, B, reducing the range of the various predictions.

The main failure of the model occurs in the early data; it revolves around the predictions for the target 4. The failure may be traced to the split data. In particular, the (target 4, split 1) latency is overpredicted. In the data, (target 4, split 1) is much faster than (target 5, split 2) and also faster than (target 6, split 3). Such a finding is qualitatively inconsistent with the model. If the target 4 data are reliable, then we must consider possible explanations outside of the present model. Perhaps target 4 with competitors 1, 2, and 3 occurs often enough that it is learned as a holistic pattern, and the response may not be dependent on split. Another, less likely, possibility requires that the subject be able to process automatically the fact that no digit above 4 is in the display; if so, an inference can be made that the target must be 4. Other possibilities are based on other sorts of alterations of the model (for example, by assuming different target digits are associated with different base time parameters), but we will not discuss these further. We did implement a version of the model that did not include target 4 in any of the predictions, and an excellent fit of all the other aspects of the data was obtained.

The version of the model with $V = 0$ predicts no errors. Although error rates were very low, errors did occur and the model therefore mispredicts the accuracy data. Although we could not obtain a best fit (because of the time needed for simulations) when $V$ was greater than zero, we did examine the fit obtained when we retained all the parameter values other than that for $V$ in Table 1, but allowed $V$ to vary. In this case, both the accuracy data and the latency data contributed to the least-squares measure. We found the various latency predictions were not altered in any important way, and the probabilities of correct choice were also reasonably well predicted.

A Learning Model

It could be argued with at least some justification that we have done little more than map a large number of scale values into reaction times and that we have not linked our model very closely to automatism theory. It is appropriate to ask, How do the scale values become learned? Can learning proceed quickly enough to allow automatic processing in the “early” sessions? Is the learning mechanism based simply on frequencies of occurrence as a target, and occurrence as a distractor? To start to answer these and other questions, we need to describe a few more details of the original study, and then propose some specific learning models from which the strength values may be derived.

In Durso et al.’s (1987) study, the early sessions consisted of Sessions 1 through 5, involving a total of 540 training trials. It may well be the case that the first few trials involve some sort of search and comparison operation until enough learning occurs to allow automatic processing to set in.² We have simplified matters by ignoring this possibility. In fact, for the purposes of modeling, we decided to treat the mean early data as if they were representative of the mean early training trial (Trial 270), and the mean late data (Sessions 27 through 31) as if they were representative of the mean late training trial (Trial 3079).

We believe it is reasonable to assume that automatic processing could develop in just the first session or two of training in the present task. Schneider (personal communication, 1986) observed that automatic detection in memory search tasks can begin to develop in as few as eight trials of training when the targets and distractors are highly discriminable and members of different, already learned categories. In the present case, the known ordering of digits may allow rapid learning, making it not unreasonable that automatic processing develops in the first session or two. Such a hypothesis can be examined in future research, but will simply be assumed for the present exposition.

As a first step toward a learning model, we can make the simple assumption that each presentation of an item (i) as a

²Although Durso, Cooke, Breen, and Schvaneveldt (1987) were kind enough to supply us with their raw data on paper printouts, no computer coding was available, and we could not devote the time to recode their data for analysis. Thus, many interesting analyses could not be examined; these included possible strategy changes in the first few sessions. For example, if the first session involved some serial search process, then the pattern of results might have differed significantly from that seen later in the study.
target increases its strength, \( M(i) \), by a fixed amount, \( e \). Similarly, we can assume that each presentation of an item as a distractor \((i.e., \text{not the biggest in the display})\) decreases its strength by a fixed amount, \( f \). Without loss of generality, we can set the starting value for each item equal to zero. Table 2 gives the number of trials on which each item, \( i \), was presented as a target, \( N_a(i) \), and presented as a distractor, \( N_d(i) \), on the average in any one session of Durso et al.’s (1987) study. As can be seen in Table 2, each of the digits 4 through 9 occurs equally often as a target in any given session, although there is an inverse relation between the magnitude of a digit and its frequency of occurrence as a distractor. According to the assumption just stated, then, digits of larger magnitudes will accumulate larger strength values than digits lower in magnitude.

Using the numbers from Table 2, it is straightforward to derive predictions for the \( M(i) \) values in the early and late sessions. Let \( n \) = the number of sessions of training, and \( M(i,n) \) = the strength for digit \( i \) after \( n \) training sessions. We call this version Model A:

\[
M(i,n) = enN_a(i) - fnN_d(i). \tag{2}
\]

We obtained a best fit of this model to all the reaction time data under the assumption that \( V = 0 \). We allowed \( e, f, \) and \( B \) to attain different values for the early and late data, but \( a \) and \( b \) were set to the same values for the early and late data. The parameter values and sum of squared deviations for this model are given in Table 3. The strength values that result are given in Table 4. A graph of the predictions against the observed data demonstrated that this model did a poor job. The main problem seemed to lie in the following characteristic of this model: The difference between \( M(i) \) and \( M(i+1) \) is predicted to be larger for smaller values of \( i \) (see Table 4). It is easy to see that the model makes this prediction because the pattern of differences in the \( N_a(i) \) values in Table 2 exhibits such an ordering. That is, the target occurrences are equal for all items above 3, and the strength differences are due to the occurrences as a distractor. Unfortunately, the differences in these occurrences are larger near the lowest digits, whereas the data require the reverse pattern (as indicated by the parameter estimates in Table 1).

We therefore decided to examine two further variants of the basic learning model, quite different in character. In Model B, a target gains more strength if it occurs with a larger digit. In Model C, items larger than the target gain strength (even though they are not presented on that trial). Consider Model B first.

Suppose that the ease of the discrimination to be made on a trial determines the magnitude of the gain in strength for the target. To make this notion precise, assume (somewhat arbitrarily) that a split of size \( j \) will cause a gain in strength for the target of \( e^j \), where \( e \) and \( g \) are parameters to be estimated. Let \( p(j|i) \) be the probability that a split of \( j \) occurs when \( i \) is the target. Then we get for Model B,

\[
M(i,n) = -fnN_d(i) + enN_a(i) + \sum_{j=1}^{i-1} j^ep(j \mid i), \tag{3}
\]

split of \( j \) given target \( i \) and set size \( s \). This last quantity can be calculated from

\[
p(j \mid s,i) = \begin{cases} \frac{(i-j-1)!}{(s-1)!} & i \leq j \leq i-s \\ 0 & \text{otherwise} \end{cases} \tag{4}
\]

The best fit parameters for Model B for the reaction time data (again assuming \( V = 0 \)) are given in Table 3, along with the sum of squared deviations. The resulting strength values are given in Table 4, and the predictions are given in Figures 3 and 4 (along with the data and Model C2). The qualitative patterns of the predictions are in accord with the data. Although the fit is not quantitatively as close as for the model with freely estimated strength values, the reduction in the number of estimated parameters is considerable: from 20 to 7. The quantitative discrepancies will be further discussed later.

Model C is motivated by the possibility that items that are not presented as targets may nonetheless gain some measure of strength due to their position in the natural ordering of digits. For example, if a study similar to that of Durso et al. (1987) were carried out in which the digit 8 was never presented during training, subsequent tests of 8 as a target might suggest a strength between that of 7 and 9. If so, one would have to consider models in which nonpresented items gain strength. In one such model, not only the strength of the target on a trial is augmented, but also all digits larger than that target. Perhaps this could be accomplished by a sophisticated rehearsal mechanism.

We considered two variants of this model. In Model C1, all digits larger than the target receive an equal increment in strength, \( g \), whereas the target itself receives an increment, \( e \).

Model C1: \( M(i,n) \)

\[
= -fnN_d(i) + enN_a(i) + gn \sum_{k=4}^{i-1} N(k). \tag{5}
\]

In the second variant, Model C2, the increment for an item, \( i \), greater than the target, \( k \), is a linear function of the difference, \( i-k \).
Table 3

Best Fitting Parameter Values for Models A Through C

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>659</td>
<td>525</td>
<td>569</td>
<td>524</td>
<td>641</td>
<td>528</td>
<td>635</td>
<td>527</td>
</tr>
<tr>
<td>a</td>
<td>74.2</td>
<td>74.2</td>
<td>322</td>
<td>322</td>
<td>257</td>
<td>257</td>
<td>302</td>
<td>302</td>
</tr>
<tr>
<td>b</td>
<td>0.02</td>
<td>0.02</td>
<td>0.33</td>
<td>0.33</td>
<td>0.03</td>
<td>0.03</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>e</td>
<td>20.1</td>
<td>20.1</td>
<td>0.03</td>
<td>0.03</td>
<td>4.8</td>
<td>4.8</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>f</td>
<td>.374</td>
<td>.374</td>
<td>0.04</td>
<td>0.04</td>
<td>-0.4</td>
<td>-0.4</td>
<td>-0.12</td>
<td>-0.12</td>
</tr>
<tr>
<td>g</td>
<td>1.5</td>
<td>1.5</td>
<td>0.42</td>
<td>0.42</td>
<td>0.3</td>
<td>0.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SSD*</td>
<td>257.9</td>
<td>15.6</td>
<td>103.4</td>
<td>21.6</td>
<td>186.8</td>
<td>20.5</td>
<td>131.5</td>
<td>18.4</td>
</tr>
</tbody>
</table>

* Sum of squared deviations (ms²/1000).

Model C2: \( M(i,n) \)

\[
= - fnN_d(i) + enN_i(i) + gn \sum_{k=4}^{(i-1)} (i-k)N_i(k). \tag{6}
\]

Clearly, other assumptions concerning learning could be examined, but we restricted attention to these two variants. Some might question the underlying assumptions of Models C1 and C2 that allow strength to be augmented for nonpresented items higher than the target. At the least, some restrictions on the general model would seem likely to be needed. For example, it may be that augmentation only occurs for items presented only as distractors.

The fitting procedures for Models C1 and C2 were similar to that for Models A and B. Parameter estimates and least squares for each model are given in Table 3; the strength values are given in Table 4. Model C1 does not fare as well as C2, failing for reasons similar to those causing problems for Model A. We will not discuss Model C1 further.

The predictions of the better fitting model, Model C2, are graphed (along with the data and predictions of Model B) in Figures 3 and 4. It is apparent that Model C2 does not do as good a job as the model with freely estimated strengths. However, the model does seem to capture the main characteristics of both the early and late data. Furthermore, it does so with a considerable reduction in the number of freely estimated parameters: 7 total parameters for the combined early and late data for Model C2 compared with 20 total parameters for the original model depicted in Figures 1 and 2.

In the best fit of Models B and C2 to the data, parameter \( f \) was estimated to attain a small negative value. Because \( f \) is the value subtracted each time an item occurs as a distractor, the net effect is to produce a small growth of strength of attention attraction, even for items presented only as distractors. This possibility is consistent with the present data but may not prove acceptable in further research. Also, a growth of attention attracting strength for distractors is not consistent with research by Dumais (1979; reported in Shiffrin & Dumais, 1981): Subjects were given consistent training in a typical visual-search task. Then the targets or distractors were replaced by items that had not been seen during training. In such a task, Dumais showed that consistently mapped distractors attracted attention less than nonpresented items. The largest digit task of Durso et al. (1987) is, of course, different from the visual search task used by Dumais, and perhaps strengths develop differently within the two paradigms. Beyond this point, we would caution the reader not to attach too great a significance to the numerical value of this parameter estimate. For one thing, the parameter space is fairly flat in the region of this parameter, and predictions almost as good can be obtained when \( f \) is set to a small positive value instead (say, .1). Secondly, we suspect that the estimated value of this parameter is strongly affected by the exact form of certain assumptions in Models B and C2 that were chosen somewhat arbitrarily, namely, the last terms in Equations 3

Table 4

Strength Values Derived From the Parameter Estimates in Table 3 for Each Model

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>M(1)</td>
<td>-22</td>
<td>-198</td>
<td>9.1</td>
<td>105</td>
<td>83</td>
<td>746</td>
<td>26</td>
<td>231</td>
</tr>
<tr>
<td>M(2)</td>
<td>-22</td>
<td>-198</td>
<td>9.1</td>
<td>105</td>
<td>83</td>
<td>746</td>
<td>26</td>
<td>231</td>
</tr>
<tr>
<td>M(3)</td>
<td>-22</td>
<td>-198</td>
<td>9.1</td>
<td>105</td>
<td>83</td>
<td>746</td>
<td>26</td>
<td>231</td>
</tr>
<tr>
<td>M(4)</td>
<td>1,752</td>
<td>15,767</td>
<td>11.8</td>
<td>135</td>
<td>496</td>
<td>992</td>
<td>1,391</td>
<td>12,516</td>
</tr>
<tr>
<td>M(5)</td>
<td>1,769</td>
<td>15,919</td>
<td>12.1</td>
<td>140</td>
<td>517</td>
<td>1,335</td>
<td>1,412</td>
<td>12,712</td>
</tr>
<tr>
<td>M(6)</td>
<td>1,782</td>
<td>16,040</td>
<td>13.1</td>
<td>151</td>
<td>541</td>
<td>1,685</td>
<td>1,462</td>
<td>13,161</td>
</tr>
<tr>
<td>M(7)</td>
<td>1,794</td>
<td>16,144</td>
<td>14.6</td>
<td>168</td>
<td>567</td>
<td>2,040</td>
<td>1,540</td>
<td>13,859</td>
</tr>
<tr>
<td>M(8)</td>
<td>1,803</td>
<td>16,230</td>
<td>16.5</td>
<td>190</td>
<td>595</td>
<td>2,398</td>
<td>1,645</td>
<td>14,806</td>
</tr>
<tr>
<td>M(9)</td>
<td>1,812</td>
<td>16,306</td>
<td>18.8</td>
<td>216</td>
<td>624</td>
<td>2,758</td>
<td>1,778</td>
<td>15,999</td>
</tr>
</tbody>
</table>
and 6. For these reasons we think that it would be best to settle this issue empirically in future research (for example, by the testing of a small digit such as 1 that had not been seen in training).

It should be noted that Models B and C2 assume $V = 0$ and were fit only to the latency data. We retained the parameter values of these models, except for $V$, allowed $V$ to vary, and used a computer simulation to fit the combined latency

Figure 3. Left panel: predicted and observed latencies for correct responses as a function of the number of digits in a display (set size), for early and late training sessions; right panel: predicted and observed latencies for correct responses as a function of the numerical value of the largest digit in the display, for early and late training sessions. (The predictions are derived from Models B and C2, and the parameter values giving rise to the predictions are given in columns 3 and 4 and 7 and 8 of Table 3.)

Figure 4. Predicted and observed latencies for correct responses as a function of the numerical difference between the two largest digits for the early and late training sessions. (Predictions are derived from Models B and C2. The curves are graphed as explained in the caption for Figure 2.)
and accuracy data. The latency predictions were not altered in any substantial fashion, and a reasonably good fit of the accuracy data was obtained as well.

Compared with the model with freely estimated strengths, both Models B and C2 fail to capture some of the quantitative features of the data, such as the form of the serial position function (see Figure 1). It seems clear that the discrepancy is located in the particular choice of functions used to produce learning in the two models: the forms of Equations 3 and 6 (and, in particular, the rightmost terms in those expressions). The expressions were chosen for simplicity and convenience rather than in any real attempt to fit the quantitative details of the data. We suspect that we would be able to choose more complex learning algorithms that would closely approximate the strength estimates from the model with freely chosen strength values, and that would therefore fit the data as well as that model. We have not pursued this course because there is little difference between freely choosing strength estimates and freely choosing a learning algorithm. It seems preferable to us to defer such subtle theoretical questions until additional data are available.

On the whole, the proposed models, based on automatic attention attraction, capture the essential properties in the data. The models make a number of interesting and testable predictions. First, a change in the mapping consistency of the task should have immense consequences. For example, after subjects have been trained, a task reversal requiring them to respond to the lowest digit in the display should be extremely difficult. Subjects in this condition would probably need to revert to, or adopt, some sort of attentive search process which would allow them to bypass their automatic search mechanism. For another example, alternating or mixing trials of training on “largest” and “smallest” responding should not allow automatic attention attraction to become a useful strategy. In such a case, some sort of limited capacity search might be needed, perhaps governed by one of the serial models discussed in the introduction to this article. Another prediction is that the present results should not be restricted to responding “largest” to “digits.” Similar results should be obtainable for responding “smallest” instead. Also, with enough training, any collection of symbols could be learned as an ordered set. Thus, the present results should be obtainable in a similar paradigm for other types of characters placed in an initially arbitrary order. Of course, in such a study one might not expect strength increments during learning for items “higher” than the targets (at least not until the arbitrary ordering has become very well learned).

Much interesting data could be obtained by investigating differential training for different symbols, and by investigating transfer to untrained symbols. For example, an item lower than any target (and therefore trained only as a distractor) could be compared with an item that has not been presented at all (both being adjacent digits, say). If the parameter $f$ in Models B and C2 is negative, then the trained item should attract attention more than the novel item, and conversely for positive values of $f$.

Finally, it should be easy to distinguish between models similar to B and those similar to C2, because nonpresented items gain strengths in Model C2 but not in B. Thus, an item near the target end of the ordering could be trained only as a distractor (on trials with a larger target), or not presented at all. If later tested as a target, this item should be very slow and error prone according to models similar to B, but perhaps fairly fast and reasonably accurate according to models similar to C2 (because these models allow items to gain strength when a smaller item is presented as a target). Conversely, an item tested only as a target and never as a distractor should develop very high strength regardless of its position in the ordering according to Model B. This should produce rapid and accurate target responses, but potential errors and slow responses if the item is later tested as a distractor. Similar problems might occur for such an item according to Model C2, but the strengths and hence the results should be very dependent on the item’s position in the ordering. These possibilities, and others, await further research.\(^3\)

\(^3\)A number of these proposed experimental tests have been carried out recently by Dan Fisk at the University of South Carolina (personal communication, July 1986). A report of the findings should be forthcoming shortly.

References


