ABSTRACT

The present chapter offers a systematic method for the analysis of selective information processing within the framework of the independent-decisions model. First, simple, parameter-free consequences derived from the model provide a way of using detection data to distinguish among three different attentional strategies. Second, it is shown how localization data can be used to answer the question: Does attention affect the quality or clearness of the internal stimulus representation?

INTRODUCTION

Many theories of information processing postulate two internal stages. First the stimulus (which may be visual or auditory) is transformed into some internal representation (iconic or echoic storage). Second, a decision is made regarding which response to make. From the viewpoint of such theories an important issue is whether the quality or clearness of any one part of the representation depends, in part, on the amount of attention allotted to it.

One of the most popular variables in studies of visual attention is the number of signal locations (set size). Interest in this variable arises from the assumption that increasing the number of signal locations means less attention can be paid to each. In fact, error rate and response time typically do increase with increases in the number of signal locations. This set-size effect cannot, however, be attributed simply to attentional factors. Increasing the number of signal locations also increases the complexity of the decision-
making process. Thus, attentional determinants of the stimulus representation are confounded with the characteristics of the decision-making process. In fact, an analysis of the statistical considerations involved in decision making may provide a complete account of the effect of number of signal locations on performance; it may not be necessary to postulate an effect of attention on the quality of the stimulus representation. This point has been made by a number of authors, including Eriksen and Spencer (1969), Gardner (1973), Kinchla (1969b), and Shiffrin and Gardner (1972).

The present chapter is concerned with the separation of attentional factors from decision-making processes. The analysis will focus on a simple experimental paradigm: a trial-by-trial detection task, in which the subject monitors two or more locations for the presence of a target stimulus and makes a binary response on each trial. In particular, first I examine the situation in which the subject’s response choices are yes, a target occurred, or no, a target did not occur.

The problem of separating attentional and decision-making processes is complicated by the fact that the relevant decision-making considerations depend on which response-selection rule is assumed (Shaw, 1979). The present discussion operates within the framework of an independent-decisions model. This type of model has a long history in psychology (e.g., Gardner, 1973; Graham, in press; Green & Swets, 1966; Starr, Metz, Lusted, & Goodenough, 1975).

**The Independent-Decisions Model.** In the independent-decisions model described by Green and Swets (1966), the subject makes a decision on the basis of statistically independent observations or sources of information (e.g., several signal locations or several observations of the same location). For each observation a separate, binary decision (positive or negative) is made. The response is based on all the observations; it is yes if any one of the separate decisions is positive and no if all are negative. Thus, the overall probability of a yes response is

\[
\prod_{k=1}^{n} (1 - P_k),
\]

where \( P_k \) is the probability of a positive decision on the \( k \)th observation.

**EXPERIMENTAL PARADIGM**

We shall consider the application of the independent-decisions model to a hypothetical visual-detection experiment. In this paradigm, the stimulus consists of two or more signals, each of which may appear in two or more
signal locations. One or more of the possible signals are designated as targets. This subset is termed the target set; its complement is the distractor set. The subject's response options consist of a simple binary choice. The first part of this discussion will concentrate on the yes–no case, in which the subject judges whether or not a target was present on a given trial. Later, a second response option will also be considered: which of several possible target locations contained a target.

Consider an experiment in which there are two signal locations, denoted $L_a$ and $L_b$. There is one item in the target set; there may be one or several items in the distractor set. Because two types of signal (target or distractor) may occur in each location, for every trial there are four possible stimulus patterns. These will be denoted by $S_{ij}$. The first subscript, $i$, indicates the signal in location $L_a$. $i = 1$ indicating a target and $i = 0$, a distractor. Similarly, the second subscript, $j$, indicates the signal in location $L_b$. Thus, for example, $S_{10}$ denotes a target in position $L_a$ and a distractor in position $L_b$. Notice that it is possible for a target to appear in both locations at once ($S_{11}$).

The Dependent Measure. The basic dependent variable is the conditional response probability $P_{ij}$, which is defined here as the probability of a no response given stimulus pattern $S_{ij}$. Thus, $P_{10}$ refers to the probability of a no response when there is a target in location $L_a$ and a distractor in $L_b$. The probability of a no response is used—rather than its complement, the probability of a yes response—because the expression of predictions is thereby greatly simplified.

The probability $P_{ij}$ should be understood to be conditional, not only upon a given stimulus pattern but also upon a given set of attentional instructions, a priori stimulus probabilities, and criterial instructions.

Attentional instructions are of two kinds, specifying undivided or divided attention. In the undivided condition, subjects may be instructed to respond yes only when a target is detected in location $L_a$ or, alternatively, to respond only to a target in $L_b$. The divided attention condition directs the subject to respond yes when a target is detected in either $L_a$ or $L_b$, or in both. The subjects may be directed to divide their attention equally or unequally between the two locations. The division of attention can also be influenced by the a priori stimulus–probability distribution.

The a priori stimulus–probability distribution may or may not be symmetrical. When it is symmetrical, targets appear at $L_a$ or $L_b$ with equal likelihood, and it is expected that subjects will pay equal attention to the two locations unless otherwise directed. When it is not symmetrical, subjects will generally attend more to the location with the higher target probability. The probability that a given stimulus pattern will occur is determined by the separate target probabilities at the two locations: For example, the probability of $S_{10}$ is the product of the probability of a target at $L_a$ and the probability of a distractor at $L_b$. 
Clearly, the target probabilities will influence the subjects' response tendencies, as well as their division of attention. For example, as the probability of a target decreases, it is expected that the subjects will become more inclined to say no—that is, more conservative. Such response tendencies are also directly influenced by the criterial instructions.

Criterial instructions involve telling the subject to (1) "respond yes even when you are only slightly sure you saw the target"; or (2) "respond yes only when you are very sure you saw the target." These are termed, respectively, liberal instructions and conservative instructions.

THE FORMAL MODEL

It is assumed that each presentation of a stimulus pattern evokes some internal or sensory response within the subject. This internal state will be specified by two random variables, one for each target location. Their means and variances depend on the signal event in that location: For each variable there is one set of parameters corresponding to a target, another set for a distractor. The four probability distributions are assumed to have finite means and variances. Otherwise, no assumptions are made about these distributions—they may be Gaussian or non-Gaussian. The random variables are denoted by the terms $X_{ai}$ and $X_{bj}$, with the first subscript indicating the location ($La$ or $Lb$) and the second subscript, the objective signal event in that location ($I = a$ target, $O = a$ distractor).

The subject is seen as making two independent decisions based on the values of the random variables $X_{ai}$ and $X_{bj}$ and the values of two corresponding decision criteria $\beta_a$ and $\beta_b$. Specifically, the decision regarding location $La$ will be positive if $X_{ai} \geq \beta_a$ and negative if $X_{ai} < \beta_a$. Similarly the decision regarding location $Lb$ will be positive if $X_{bj} \geq \beta_b$ and negative if $X_{bj} < \beta_b$. The subject is represented as making an overt yes response if the decision about either location is positive and a no response otherwise.

If we wish to separate decision-making and attentional effects on performance, it is necessary to specify more precisely what is meant by attention. The issue is not whether there is such a thing as attention—obviously there is—but whether it influences the quality and clearness of the stimulus representation. Introspection suggests that the less attention we pay to a part of the visual or auditory world, the less sensitive we are to that part. This notion can be represented within the framework of the independent-decisions model by the assumption that the mean and/or variance of the internal random variables $X_{ai}$ and $X_{bj}$ are, in part, functions of how much attention is paid to the stimulus components they represent. For example, the variances ("noise levels") of $X_{ai}$ and $X_{bj}$ may be assumed to increase monotonically with decreases in the amount of attention paid: the less attention, the more noise.
Alternatively, attention may not affect the clarity of the stimulus representation but only the parameters of the decision process, such as the decision criteria. Then attentional instructions will have no effect on the parameters of the internal random variables.

In order to decide between these two possibilities it will be useful first to distinguish among several alternative versions of the independent-decisions model. This will be done by generating parameter-free predictions from the different models—predictions that do not require estimation of the models' parameters. Because their accuracy does not rest on the accuracy of estimated model parameters, these predictions hold whether or not attention influences the means or variances of the internal random variables.

Two Kinds of Attention Allocation

This section examines several ways in which attention may influence the decision-making process. Two major variations of the independent-decisions model will be considered.

The Sharing Model. The first will be termed the sharing model. It can be interpreted in one of two ways: Attention may be given to both locations simultaneously, or attention may be switched rapidly back and forth between the two locations within a single trial. For the present purposes, these two alternatives are considered to be equivalent.

A principle feature of this model is that the allocation of attention does not change from trial to trial. Thus, the conditional probability of a no response given stimulus $S_{ij}$, written $P_{ij}$, remains constant—it is simply the product of the probabilities that each of the two random variables does not exceed its corresponding criterion:

$$P_{ij} = P(X_{ai} < \beta_a)P(X_{bj} < \beta_b)$$

(Bear in mind that the equations in this section are neutral regarding the issue of whether or not attention affects the parameters of the random variables $X_{ai}$ and $X_{bj}$. Attention may affect only the decision criteria $\beta_a$ and $\beta_b$ or only the parameters of the random variables or it may affect both.

The Mixture Model. The second kind of attention-allocation model is termed the mixture model, because processing is characterized as a mixture of two discrete allocation rules. (This has been called an "all or none" process by Kinchla, 1969a, and "attention switching" by Sperling & Melchner, 1978.) On each trial the observer allocates attention primarily to one location or primarily to the other, in a probabilistic fashion. Thus, on a given trial the observer is in one or the other of two different states, which differ in that each
has its own set of criteria. (Again, the parameters of the random variables $X_{ai}$ and $X_{bj}$ may be the same or different in the two states.) Overall performance will be a weighted mixture of what would occur if each of the attentional states were used exclusively.

Two types of mixture model will be considered. In the first, information on a given trial is obtained either from one location or from the other. This will be called the type 1 mixture model, because information comes from only one location. The type 2 mixture model, on the other hand, allows for some impression from the unattended location: Information is obtained from two locations (as it is with the sharing model).

Let the probability that attention is directed primarily to $L_a$ be $\alpha$ and the probability that attention is directed primarily to $L_b$ be $1 - \alpha$. Then for the type 1 mixture model, the conditional probability of a no response is simply

$$P_{ij} = \alpha P(X_{ai} < \beta_a) + (1 - \alpha) P(X_{bj} < \beta_b).$$

With this model, a no response is as likely to occur when there is a target in both locations as when there is a target only in the attended location.

For the type 2 mixture model, the conditional probability of a no response becomes

$$P_{ij} = \alpha P(X_{ai} < \beta_a)P(X_{bj} < \beta'_b) + (1 - \alpha) P(X_{ai} < \beta'_a)P(X_{bj} < \beta_b),$$

where $\beta_k$ is the criterion when attention is directed primarily to location $L_k$ and $\beta'_k$ is the criterion when attention is not directed primarily to $L_k$. The attentional state in which criteria $\beta_a$ and $\beta'_b$ are used occurs with probability $\alpha$, and the state in which $\beta'_a$ and $\beta_b$ are used occurs with probability $1 - \alpha$.

There are several different ways to characterize the difference between the type 1 and type 2 mixture models. The failure to make use of information from the unattended location, in the type 1 model, may be attributed to one of the following causes:

1. The parameters of the internal random variable corresponding to the unattended location may be such that the information coming from that location is practically worthless. If the signal and noise distributions overlap almost completely, then the criterion for that location ($\beta'$) may be set so high that it is virtually never exceeded.

2. The parameters of the random variable may be unaffected by the allotment of attention, but the criterion for the unattended location may nevertheless be set high enough that it is unlikely ever to be exceeded.

3. The internal random variable and the criterion may both be unaffected by attention, but the information about whether the criterion has been exceeded may simply not be used in the later stages of the decision-making
process. The message may be lost or ignored and play no role in determining the overt response.

Predictions

Table 14.1 summarizes the three types of independent-decisions model, as well as the parameter-free predictions that will now be derived from these models.

Consider first the attention-sharing model. From Equation 2 it follows that

\[ P_{01} = P(X_{a0} < \beta_a)P(X_{b1} < \beta_b) \]

and

\[ P_{11} = P(X_{a1} < \beta_a)P(X_{b1} < \beta_b). \]

Then

\[ \frac{P_{01}}{P_{11}} = \frac{P(X_{a0} < \beta_a)P(X_{b1} < \beta_b)}{P(X_{a1} < \beta_a)P(X_{b1} < \beta_b)} = \frac{P(X_{a0} < \beta_a)}{P(X_{a1} < \beta_a)} \]

Similarly,

\[ \frac{P_{00}}{P_{11}} = \frac{P(X_{a0} < \beta_a)P(X_{b0} < \beta_b)}{P(X_{a1} < \beta_a)P(X_{b0} < \beta_b)} = \frac{P(X_{a0} < \beta_a)}{P(X_{a1} < \beta_a)} \]

Assuming that none of the above probabilities are equal to zero, the sharing model predicts equality between the two ratios:

\[ \frac{P_{01}}{P_{11}} = \frac{P_{00}}{P_{11}} \] (5)

The mixture models, on the other hand, do not predict the equality of Equation 5. Whether or not the unattended location is assumed to have an effect, the mixture models predict response probabilities by a sum rather than a product, so the cancellations just shown do not occur.

Returning to the sharing model, the second parameter-free prediction concerns the two sums of conditional response probabilities, \( P_{01} + P_{10} \) and \( P_{00} + P_{11} \). From Equation 2 it follows that:

\[ \begin{align*}
(P_{01} + P_{10}) - (P_{00} + P_{11}) &= P(X_{a0} < \beta_a)P(X_{b1} < \beta_b) + P(X_{a1} < \beta_a)P(X_{b0} < \beta_b) \\
&- P(X_{a0} < \beta_a)P(X_{b0} < \beta_b) - P(X_{a1} < \beta_a)P(X_{b1} < \beta_b) \\
&= [P(X_{a0} < \beta_a) - P(X_{a1} < \beta_a)] \times [P(X_{b1} < \beta_b) - P(X_{b0} < \beta_b)].
\end{align*} \] (6)
<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
<th>Equation</th>
<th>Predictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sharing model</td>
<td>a single allocation rule</td>
<td>$P_y = P(X_{ai} &lt; \beta_a)P(X_{b_j} &lt; \beta_b)$</td>
<td>$\frac{P_{01}}{P_{10}} = \frac{P_{00}}{P_{11}}$</td>
</tr>
<tr>
<td>Mixture model, type 1</td>
<td>No information from unattended location</td>
<td>$P_y = \alpha P(X_{ai} &lt; \beta_a) + (1-\alpha)P(X_{b_j} &lt; \beta_b)$</td>
<td>$P_{01} + P_{10} = P_{00} + P_{11}$</td>
</tr>
<tr>
<td>Mixture model, type 2</td>
<td>Some information from unattended location</td>
<td>$P_y = \alpha P(X_{ai} &lt; \beta_a')P(X_{b_j} &lt; \beta_b')$ + (1-\alpha)P(X_{ai} &lt; \beta_a')P(X_{b_j} &lt; \beta_b)$</td>
<td>$\frac{P_{01}}{P_{10}} = \frac{P_{00}}{P_{11}}$</td>
</tr>
</tbody>
</table>

*Note: Two locations are assumed, with attention divided equally between them.*
But the probability that a variable will fall below the criterion obviously must be greater in the case of a distractor than in the case of a target, so

\[ P(X_{a0} < \beta_a) > P(X_{a1} < \beta_a) \]  \hspace{1cm} (7) \]

and

\[ P(X_{b0} < \beta_b) > P(X_{b1} < \beta_b). \]  \hspace{1cm} (8) \]

Thus, of the two multiplicands in Equation 6, the first is positive and the second is negative, so the expression in Equation 6 is negative and

\[ (P_{01} + P_{10}) - (P_{00} + P_{11}) < 0. \]

Therefore

\[ P_{01} + P_{10} < P_{00} + P_{11}. \]  \hspace{1cm} (9) \]

The same inequality shown in Equation 9 is predicted for the type 2 mixture model, which uses the assumption that there is some information gained from the unattended location. For this kind of model, the probability of a no response is given by Equation 4, from which it follows that:

\[ (P_{01} + P_{10}) - (P_{00} + P_{11}) = \]
\[ \alpha[\{P(X_{a1} < \beta_a) - P(X_{a0} < \beta_a)\} \times [P(X_{b0} < \beta_b') - P(X_{b1} < \beta_b')] + (1 - \alpha)[P(X_{a1} < \beta_a') - P(X_{a0} < \beta_a')] \times [P(X_{b0} < \beta_b) - P(X_{b1} < \beta_b)]. \]

From the inequalities in Equations 7 and 8, the first two multiplicands are, respectively, negative and positive; the second two are also negative and positive. Thus both of the summed terms are negative, and as in Equation 9

\[ P_{01} + P_{10} < P_{00} + P_{11}. \]  \hspace{1cm} (10) \]

The above inequality does not hold for the mixture model that uses the assumption of no information gained from the unattended location. On the contrary, the type 1 mixture model predicts that the two sums will be equal. From Equation 3 it follows that

\[ P_{01} + P_{10} = \alpha P(X_{a0} < \beta_a) + (1 - \alpha)P(X_{b1} < \beta_b) \]
\[ + \alpha P(X_{a1} < \beta_a) + (1 - \alpha)P(X_{b0} < \beta_b) \]  \hspace{1cm} (11) \]

and
\[ P_{00} + P_{11} = \alpha P(X_{a0} < \beta_a) + (1 - \alpha)P(X_{b0} < \beta_b) \\
+ \alpha P(X_{a1} < \beta_a) + (1 - \alpha)P(X_{b1} < \beta_b) \]  

(12)

Inspection of Equations 11 and 12 shows that both \( P_{01} + P_{10} \) and \( P_{00} + P_{11} \) are equal to:

\[ \alpha[P(X_{a0} < \beta_a) + P(X_{a1} < \beta_a)] + (1 - \alpha)[P(X_{b0} < \beta_b) + P(X_{b1} < \beta_b)]. \]

Therefore

\[ P_{01} + P_{10} = P_{00} + P_{11}. \]  

(13)

It is important to remember that the preceding parameter-free predictions depend on neither the parameters of the internal random variables nor the subjects’ decision criteria. It should also be noted that these statistics distinguish among the models only when attention is divided. When attention is directed exclusively to one location, the models make the same predictions. For example, when attention is given solely to location \( L_a \) the prediction for \( P_{00} \) is the same as for \( P_{01} \); namely,

\[ P_{00} = P(X_{a0} < \beta_a). \]

For both \( P_{10} \) and \( P_{11} \) the prediction is

\[ P_{10} = P(X_{a1} < \beta_a). \]

Thus

\[ \frac{P_{00}}{P_{10}} = \frac{P_{01}}{P_{11}} \]  

(14)

and

\[ P_{01} + P_{10} = P_{00} + P_{11}. \]  

(15)

These predictions for the case when attention is not divided provide an additional test of the models.

**Quantitative Predictions.** Table 14.2 illustrates how these statistics allow one to distinguish between the models. Quantitative predictions are generated here for the sharing model and for the type 1 mixture model, in which no information is gained from the unattended location. Exact predictions are not possible for the type 2 mixture model, because inequalities are predicted both
<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>$P_{00}$</th>
<th>$P_{11}$</th>
<th>$P_{01}$</th>
<th>$P_{10}$</th>
<th>$P_{01}/P_{11}$</th>
<th>$P_{00}/P_{10}$</th>
<th>$P_{01} + P_{10}$</th>
<th>$P_{00} + P_{11}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sharing model</td>
<td>High</td>
<td>.90</td>
<td>.20</td>
<td>.42</td>
<td>.42</td>
<td>2.12</td>
<td>2.12</td>
<td>.84</td>
<td>1.11</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>.90</td>
<td>.42</td>
<td>.61</td>
<td>.61</td>
<td>1.46</td>
<td>1.46</td>
<td>1.23</td>
<td>1.32</td>
</tr>
<tr>
<td>Mixture model, type 1</td>
<td>High</td>
<td>.90</td>
<td>.20</td>
<td>.55</td>
<td>.55</td>
<td>2.75</td>
<td>1.64</td>
<td>1.11</td>
<td>1.11</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>.90</td>
<td>.42</td>
<td>.66</td>
<td>.66</td>
<td>1.57</td>
<td>1.36</td>
<td>1.32</td>
<td>1.32</td>
</tr>
</tbody>
</table>

TABLE 14.2
Quantitative Predictions for the Probability of a No Response
for the ratios $P_{01}/P_{11}$ and $P_{00}/P_{10}$ and for the sums $P_{01} + P_{10}$ and $P_{00} + P_{11}$. Thus if experimental results indicate that the ratios are not equal and also that $P_{01} + P_{10}$ is less then $P_{00} + P_{11}$, these data will lend support to the type 2 mixture model, in which some information is obtained from the unattended location.

The predictions shown in Table 14.2 are computed under two conditions: (1) high accuracy, where $P_{00} = .90$ and $P_{11} = .20$; and (2) low accuracy, where $P_{00} = .90$ and $P_{11} = .42$. Given the response probabilities for the two-target and two-distractor stimulus patterns, one can easily compute the response probabilities for the one-target patterns, $P_{01}$ and $P_{10}$. (For these examples it will be assumed that the attentional instructions are for equal attention and that the a priori probability of $S_{01}$ equals that of $S_{10}$, so we expect $P_{01}$ to equal $P_{10}$.

In the case of the sharing model, from Equation 5 we know that

$$P_{01}P_{10} = P_{00}P_{11}.$$  \hspace{1cm} (16)

Because $P_{01} = P_{10}$, both are equal to $\sqrt{P_{00}P_{11}}$, or .42 in the high-accuracy condition and .61 in the low-accuracy condition.

For the type 1 mixture model, from Equation 13 it follows that $P_{01}$ and $P_{10}$ are computed from $(P_{11} + P_{00})/2$. This equals .55 with high accuracy and .66 with low.

From the values for $P_{00}, P_{11}$, and $P_{01}$ (or $P_{10}$), the other predictions in Table 14.2 are readily derived. Notice that the predicted ratios for $P_{01}/P_{11}$ and $P_{00}/P_{10}$ are equal in the case of the sharing model, whereas with the mixture model $P_{01}/P_{11}$ is larger. This is true for both levels of accuracy; however, the effect is greater at the higher accuracy level. The models’ predictions for the two sums, $P_{01} + P_{10}$ and $P_{00} + P_{11}$, also differ more at higher accuracies. These sums are equal in the case of the sharing model, whereas the mixture model predicts that $P_{00} + P_{11}$ will exceed $P_{01} + P_{10}$. Clearly, any experiment designed to test these models must involve sufficiently high levels of accuracy, or these statistics will not discriminate among them.

A final cautionary note concerns the matter of individual differences. It is quite possible for Subject A to behave according to the sharing model, Subject B according to the type 1 mixture model, and Subject C according to the type 2 mixture model. Averaging data would, of course, obscure this fact. Moreover, it is also possible that a given subject may operate in different ways at different times, depending on the type or complexity of the task.

The Set-Size Effect

At this point we return to the question of why the error rate goes up with set size, which in this context refers to the number of possible signal locations. When there are more locations to be monitored, less attention is presumably
available for each location. Thus, if attention does affect the quality of the stimulus representation, four locations will appear less clear than two—or, in terms of the model, the parameters of the internal random variables will be influenced by the number of signal locations. This would result in a greater decline in accuracy, as set size increases, than would be expected from decisional factors alone.

Before attacking this problem directly, it was necessary to distinguish among the three models of attention allocation, which are variations of the independent-decisions model. This was important because the interpretation of the second stage of the analysis depends on the outcome of the preceding investigation.

**Location Judgments.** The paradigm of the hypothetical experiment will now be expanded. There will be two display conditions: one with two target locations and one with four. The target can appear in only one location on a given trial. The subject will now be asked not only to decide whether a target appeared but to give its location as well. The percentage of correct location judgments becomes the dependent measure of interest.

The independent-decisions model is extended to location judgments by the following postulate: that the signal location chosen is the one evoking the strongest impression of the presence of the target (the largest value of $X_k$). Assume for a moment that increasing the number of attended locations from two to four does not change the parameters (mean and/or variance) of the internal random variable corresponding to a given display location. Given the probability distribution of such an internal random variable, it is possible to compute the expected change in the probability of a correct location judgment as the number of locations increases from two to four.

I have explored the expected change in probability correct under a number of different assumptions about the shape of the underlying probability distributions. The predictions for two of these assumed distributions, the Gaussian and the exponential, are shown in Fig. 14.1. In this figure the predicted probability of a correct location judgment in the four-location condition ($P_4$) is plotted against the probability of a correct two-location judgment ($P_2$). Each of the four solid lines in Fig. 14.1 was generated by a different set of assumptions. The curve labeled *Gaussian* is based on the assumptions of normally distributed random variables, equal variance for the target and distractor distributions, and no effect of set size on the parameters of these distributions. The *exponential* curve is based upon the same assumptions except for the postulated shape of the distributions.

As a predicted function moves to the right and downward in Fig. 14.1, increasingly greater drops in accuracy are predicted for the four-location task. The exponential model predicts the least decline in accuracy as the number of target locations goes from two to four; the Gaussian model predicts a slightly greater decline. For both of these models, there is no change
in the parameters of the random variables when set size increases. Thus, a model based on either Gaussian or exponential that *does* call for a change in the parameters with set size would produce a greater drop in four-location accuracy than predicted by the Gaussian and exponential models shown in Fig. 14.1.
A set of points for the pairs of probabilities \((P_2, P_4)\) can be generated empirically by varying, for example, display energy. Given such a set of data points, when is it possible to say that these points cannot be generated by a model that assumes no effect of set size on the parameters of the internal random variables; that is, when can we reject the null hypothesis of no effect of attention on the quality or clearness of the internal representation?

If the analysis of the two-location yes–no data indicates that the subject used either the sharing model or the type 2 mixture model, an answer is readily available. It consists of the boundary curve shown in Fig. 14.1. Any empirical function lying significantly below and to the right of that boundary must derive from a model in which set size does affect the parameters of the random variables.

The boundary function is based on the following arguments. The signal location chosen is assumed to be the one evoking the strongest impression of the target’s presence, in other words, the largest value of \(X_{kl}\). Thus, location \(k\) is chosen out of \(N\) locations if the value of \(X_{kl}\) is greater than the values of all the other \(N - 1\) random variables. Suppose that each \(X_{ko}\) (distractor variable) is distributed as \(F_d(x)\) with density function \(f_d(x)\) and that \(X_{k1}\) (the target variable) is distributed as \(F_t(x)\) with density function \(f_t(x)\). Suppose also that as \(N\) (the number of random variables) increases, the functions \(F_d\) and \(F_t\) retain the same mean and variance. Then the probability \(P_N\) that the target random variable is larger than any of the \(N - 1\) distractor variables is

\[
P_N = \int_{-\infty}^{\infty} f_t(x) F_d^{N-1}(x) \, dx.
\]  

Using Holder’s inequality (Royden, 1963) it is easily shown (see Appendix) that

\[
\int_{-\infty}^{\infty} f_t(x) F_d^{N-1}(x) \, dx \geq \left[ \int_{-\infty}^{\infty} f_t(x) F_d(x) \, dx \right]^{N-1}.
\]  

This inequality simply says that \(P_N\) is always larger than or equal to the result of raising \(P_2\) (probability correct for two locations) to the power \(N - 1\):

\[
P_N \geq P_2^{N-1}.
\]  

Thus, \(P_2^{N-1}\) forms the lower boundary for \(P_N\), given that the parameters of the random variables do not change. The boundary curve for \(P_4\) shown in Fig. 14.1 is \(P_2^{4-1}\), or \(P_2^3\).

The two curves in Fig. 14.1 that lie mostly below and to the right of the boundary are based on the capacity-allocation model and the sample-size model. Both of these models use the assumption that attention does influence the parameters of the random variables. The capacity-allocation model
(Shaw, 1978; Shaw & Shaw, 1977) is discussed by Professor Kinchla in this volume, so it is not necessary to elucidate it here. Notice, however, that the curve for this model crosses the boundary curve where \( P_2 \) equals .82. What this means is that an experiment using the predictions shown in Fig. 14.1 should aim for a \( P_2 \) greater than .82. It would be best if the task were in the .85 to .95 range. This, of course, can be achieved through control of display energy.

The curve labeled \textit{sample size} is based on a model that is similar to the sample-size model proposed by Luce and Green (Green & Luce, 1974; Luce, 1977). With this model each internal random variable \( X_{ki} \) is a sample mean based on \( N_k \) observations of that location, where the total number of observations over all locations remains constant. Thus, the variance of each \( X_{ki} \) depends directly upon \( N_k \):

\[
\sigma_k^2 = \frac{\sigma^2}{N_k}
\]

where \( \sigma^2 \) is the variance of a single observation of one random variable.

For the reader's convenience, the numerical values used to construct Fig. 14.1 are given in Table 14.3.

<table>
<thead>
<tr>
<th>( P_2 )</th>
<th>\textit{Exponential}</th>
<th>\textit{Gaussian}</th>
<th>\textit{Boundary} ((P_2^*))</th>
<th>\textit{Capacity} Allocation</th>
<th>\textit{Sample Size}</th>
</tr>
</thead>
<tbody>
<tr>
<td>.69</td>
<td>.49</td>
<td>.46</td>
<td>.33</td>
<td>.40</td>
<td>.39</td>
</tr>
<tr>
<td>.71</td>
<td>.52</td>
<td>.49</td>
<td>.36</td>
<td>.43</td>
<td>.41</td>
</tr>
<tr>
<td>.74</td>
<td>.57</td>
<td>.52</td>
<td>.41</td>
<td>.45</td>
<td>.43</td>
</tr>
<tr>
<td>.76</td>
<td>.59</td>
<td>.55</td>
<td>.44</td>
<td>.48</td>
<td>.46</td>
</tr>
<tr>
<td>.78</td>
<td>.62</td>
<td>.58</td>
<td>.47</td>
<td>.50</td>
<td>.48</td>
</tr>
<tr>
<td>.80</td>
<td>.66</td>
<td>.61</td>
<td>.51</td>
<td>.52</td>
<td>.50</td>
</tr>
<tr>
<td>.82</td>
<td>.69</td>
<td>.64</td>
<td>.55</td>
<td>.55</td>
<td>.53</td>
</tr>
<tr>
<td>.84</td>
<td>.72</td>
<td>.67</td>
<td>.59</td>
<td>.58</td>
<td>.55</td>
</tr>
<tr>
<td>.85</td>
<td>.73</td>
<td>.70</td>
<td>.61</td>
<td>.59</td>
<td>.57</td>
</tr>
<tr>
<td>.87</td>
<td>.77</td>
<td>.73</td>
<td>.65</td>
<td>.62</td>
<td>.59</td>
</tr>
<tr>
<td>.88</td>
<td>.78</td>
<td>.75</td>
<td>.68</td>
<td>.63</td>
<td>.61</td>
</tr>
<tr>
<td>.90</td>
<td>.82</td>
<td>.78</td>
<td>.73</td>
<td>.66</td>
<td>.64</td>
</tr>
<tr>
<td>.91</td>
<td>.84</td>
<td>.80</td>
<td>.75</td>
<td>.68</td>
<td>.65</td>
</tr>
<tr>
<td>.92</td>
<td>.85</td>
<td>.82</td>
<td>.78</td>
<td>.70</td>
<td>.67</td>
</tr>
<tr>
<td>.93</td>
<td>.87</td>
<td>.84</td>
<td>.80</td>
<td>.72</td>
<td>.70</td>
</tr>
<tr>
<td>.94</td>
<td>.89</td>
<td>.86</td>
<td>.83</td>
<td>.74</td>
<td>.73</td>
</tr>
<tr>
<td>.95</td>
<td>.91</td>
<td>.88</td>
<td>.86</td>
<td>.76</td>
<td>.75</td>
</tr>
<tr>
<td>.96</td>
<td>.93</td>
<td>.90</td>
<td>.88</td>
<td>.78</td>
<td>.78</td>
</tr>
<tr>
<td>.97</td>
<td>.95</td>
<td>.92</td>
<td>.91</td>
<td>.81</td>
<td>.80</td>
</tr>
<tr>
<td>.98</td>
<td>.96</td>
<td>.95</td>
<td>.94</td>
<td>.85</td>
<td>.82</td>
</tr>
</tbody>
</table>
The Type I Mixture Model. It was pointed out previously that the arguments used to construct Fig. 14.1 hold only if information is gained from more than one location on a given trial. What if the analysis of attention- allocation strategy implicates the type I mixture model? In this case attention appears to have a dramatic effect, because the information from the unattended location plays no role in the decision-making process.

Suppose that a subject using this strategy is given the task of choosing the target location. Because there is information from only the attended location, the subject must decide whether the information from that location is sufficiently favorable to suggest that a target was present. If it is, then that location is chosen; if it is not, another location is chosen at random. How will a subject fare with such a strategy?

There is an answer to this question (Shaw, in preparation) in the case of the capacity allocation model. Given a fixed amount of resources (processing capacity or attention) to be distributed among \( N \) locations, the strategy that maximizes the number of correct location judgments is as follows: Ignore one location and allocate the resource among \( N - 1 \) of the locations; should it appear that the target is not in one of the \( N - 1 \) locations, then guess the ignored location. For the capacity allocation model, it is not possible to find a strategy that will maximize both location and detection judgments. When there are only two locations, the mixture strategy can achieve a higher level of correct location judgments than the sharing strategy. Suppose that when four locations are searched only one location is attended, with failure to detect the target resulting in a guess among the three ignored locations. Then, overall percent correct location judgments will be lower than for the sharing strategy. Now suppose four locations are searched and attention is shared among three of the four locations with the ignored location being guessed if detection search fails. This strategy involves both the sharing and mixture strategies and will achieve a higher level of percent correct location judgments than a strategy in which attention is shared among all locations and response selection is based on the maximum rule. Furthermore, the optimal strategy will give values of \( P_2 \) and \( P_4 \) that may fall above the boundary curve. These results would give the impression that set size does not influence the parameters of the internal random variables, despite the fact that performance is based upon a model that does involve an effect on the parameters. For this reason, it is extremely important to determine whether or not subjects share attention between locations. (A more detailed discussion of this topic is presented in Shaw, in preparation.)

To summarize, given evidence for the sharing model of decision making, location data lying below and to the right of the boundary curve of Fig. 14.1 provide evidence that attention does influence the quality of the internal representation. The converse, however, is not true. Empirical functions that lie above and to the left of the boundary curve do not prove that attention has no effect on the parameters of the random variables. With some models of
attention allocation the underlying distribution of the random variables may
be such that \( P_N \geq P_2^{N-1} \). If this is the case, it will be necessary to determine the
underlying distributions directly, in order to test whether set size influences
the parameters of the internal random variables.

Finally, it might seem that a quick answer to the questions explored in this
chapter could have been provided by the measure \( d' \) based on detection data.
Although \( d' \) is a measure of "sensitivity" and would thus be affected by any
change in the parameters of the random variables, an answer to the set-size
question does not emerge from simply plotting an ROC curve for different set
sizes. There are two reasons for this.

First, the usual \( d' \) is based on the assumption of normally distributed
random variables with equal variances. Although this assumption is worthy
of consideration, other assumptions are also possible and have been consid-
ered here.

Second, plotting an ROC for different set sizes requires that the criteria for
the different locations be equal. If they are not equal, the resulting ROC
curves cannot be expressed in a two-dimensional graph: The points lie on a
surface. This makes it impossible to discriminate between the effects of
criterial changes and an effect of set size on the parameters of the internal
random variables.

The previous explanation applies only to the independent-decisions model,
which was the assumption underlying this entire chapter. Under a different
assumption—for example, the integration model (Green & Swets, 1966)—it is
possible to make use of ROC curves.

Conclusions

The present chapter has illustrated how the independent-decisions model can
be used to generate testable predictions regarding the effect of dividing
attention on the quality of the stimulus representation. Two kinds of data are
required for a complete analysis: detection and localization. Detection data
are used to distinguish among three different strategies of attention alloca-
tion. Two of these strategies allow the subject to utilize information, within a
given trial, from more than one location at once. Should the detection data
favor one of these strategies, localization data can then be used to answer the
question: Does attending to more locations at once decrease the quality (i.e.,
increase the noise level) of the internal representation?

ACKNOWLEDGMENTS

I would like to thank Arnold Glass, Tom Wallsten, Mark Alton, and an anonymous
reviewer for their comments on an earlier draft of this paper. Special thanks are due
Ron Kinchla and Judith R. Harris for their extensive editorial assistance. I would like
to thank Duncan Luce for pointing out that Equation 18 is easily proved using Holder's inequality. To Roy C. Milton, I am indebted for preparing extensions of tables published in his book Rank Order Probabilities: Two Sample Normal Shift Alternatives (1970). These were used in making the predictions of the Gaussian and sample-size models.

REFERENCES


Kinchla, R. A. An attention operating characteristic in vision. Paper presented at the International Conference on Attention and Performance, Institute for Perception RVO-TNO, Solsterberg, Netherlands, 1969. (a)


Shaw, M. L. On the difference between detecting and locating targets in visual search. *Psychological Review*, in press.


APPENDIX

Theorem. Let $f_i$ and $f_d$ be two probability density functions, then

$$\int_{-\infty}^{\infty} f_i F_d^N \, dx \geq \left( \int_{-\infty}^{\infty} f_i F_d \, dx \right)^N.$$

Holder's Inequality. Suppose $p, q > 0, p + q = pq$, and $f$ and $g$ are functions such that

$$\int |f|^p \, dx \quad \text{and} \quad \int |g|^q \, dx$$

exist. Then,

$$|\int fg \, dx| \leq \left( \int |f|^p \, dx \right)^{1/p} \left( \int |g|^q \, dx \right)^{1/q}.$$

Because $f_i$ and $f_d$ are densities, we can drop the absolute value signs. Let $q = N, p = N/(N - 1), f = f^{N-1/N},$ and $g = f_i^{1/N}F_d$. Then,

$$\int f_i F_d \, dx = \int f_i^{N-1/N} f_i^{1/N} F_d \, dx$$

$$\leq \left[ \int (f_i^{N-1/N})^{N-1/N} \, dx \right]^{N-1/N} \left[ \int (f_i^{1/N} F_d)^N \, dx \right]^{1/N}$$

$$= 1 \times \left[ \int f_i F_d^N \, dx \right]^{1/N}.$$

Therefore,

$$\left[ \int f_i F_d \, dx \right]^N \leq \int f_i F_d^N \, dx.$$