# 35 Visual Search, Visual Attention, and the Attention Operating Characteristic

In: J. Requin (Ed.)
Attention and Performance VII
Hillsdale, N.J.: Erlbaum, 1978

George Sperling
Melvin J. Melchner

*Bell Laboratories*
*Murray Hill, New Jersey, United States*

## ABSTRACT

Visual search experiments, in which subjects search an array of background objects for a target, are difficult to interpret because eye movements occur that are not under experimental control. The present experiments eliminated eye movements by presenting arrays of alphanumeric characters in brief flashes on a cathode ray oscilloscope. In previous experiments of this type we found that subjects searching for a numeral target in a background of letters could search simultaneously for an unknown one-of-ten numeral almost as well as for a single known numeral, and they could scan in parallel 15 to 25 characters in an array. The efficiency of this search was shown to be limited by the local density of characters.

In the attention experiments reported here, two different targets appeared in each sequence of character arrays. When the subject's task was to search simultaneously for an unknown large-size and an unknown small-size numeral, there was considerable interference between the two tasks. Instructions to attend primarily to the large or primarily to the small numeral were highly effective. Two other pairs of visual search tasks also were studied, and data from the three experiments were used to trace out Attention Operating Characteristic (AOC) curves. An AOC is defined as the locus of points on a graph plotting performance on search task 1 against performance on search task 2; it is analogous to the Receiver Operating Characteristic (ROC) of signal detection theory. Using an AOC, it is possible to measure the compatibility of two tasks between which attention is divided. In order to do this one cannot use just one condition of attention for each pair of tasks, for this would be comparing one point from each of two curves instead of comparing two curves. (A similar problem occurs in signal detection with ROC curves).

Attention-switching and attention-sharing models of two-task performance are defined. Our data enabled us to show that, in these experiments, movement along the AOC was primarily due to switching between two attention states, although some sharing of attention also occurred.

## I. VISUAL SEARCH UNDER A SINGLE ATTENTIONAL STATE

The primary mechanisms of visual attention are overt: eye movements and body orientation. But even within a single eye fixation, as we will demonstrate, attentive processes determine what particular kinds of signals are analyzed and from what parts of the visual field they are accepted.

Consider normal visual search. The subject searches an array of objects (background objects) for a critical object (target) by moving his eyes over the array. While the pattern of eye movements is interesting in itself, it complicates the analysis of attention because the eye movements are not under experimental control. Therefore, we eliminate eye movements in our experiments by having the subject keep his eyes fixated on the center of a display, and presenting new stimuli to him every $t$ msec. This method gives the experimenter precise control over the flow of information to the visual system. When $t$ is 240 msec, this display sequence approximates the sequence that the eyes produce for themselves in natural visual search.

A typical paradigm is illustrated in Fig. 1. The subject first sees a fixation field presented for one second, and then a sequence of alphanumeric character arrays. Each character array is presented as a brief flash lasting a fraction of a millisecond. The briefly flashed arrays are clearly visible; in other experiments (Sperling, 1973)[1] we have found no difference in performance between brief flashes and arrays presented continuously for 200 msec. One of the character arrays, the critical array, contains the target character. It is preceded by a random number (from 7 to 12) of noncritical arrays and followed by at least 12 noncritical arrays. The subject does not know which array contains the target character, nor what the particular target will be, nor where in the array it is located. His task is to report the identity and location of the target character, and his degree of confidence in the correctness of his report.

The great advantage in collecting reports of confidence and location – even when we are interested primarily in identification – is that confidence and location enable us to "purify" the data by indicating when the subject is guessing. We have established the following: (1) when our subjects do make location confusions, they nearly always are confusions of two adjacent locations; and (2) when subjects use the lowest confidence category ("guessing"), their responses are, in fact, statistically independent of the stimulus (guessing!). Therefore, most of the effects of guessing are eliminated simply by scoring responses as wrong whenever a location is in error by more than one position, or whenever the lowest confidence category is used (cf. Sperling & Melchner, 1976b).

The data (probability of a correct response) are analyzed in terms of $\hat{l}$, the average number of characters the subject scans in each array, and $\hat{t}$, the average time it takes to scan one character. In previous studies (Sperling, Budiansky,

[1] In this experiment, the arrays tested were those illustrated in Fig. 2 of text, and onset-to-onset times tested were 240 msec and 480 msec.

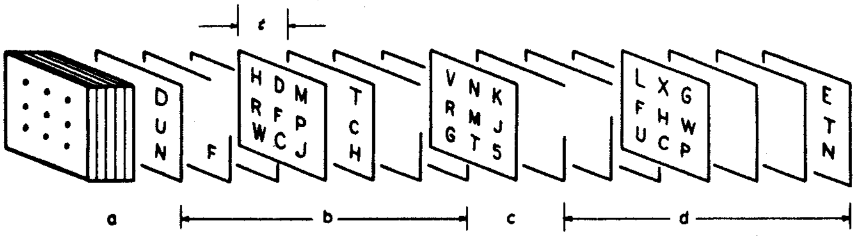

FIG. 1. Sequence of consecutive displays in the search experiment. (a) Fixation field; (b) 7 to 12 noncritical displays; (c) critical display with single target numeral; (d) 12 noncritical displays; ($t$) interval from the onset of one stimulus to the onset of the next. The subject attempts to identify the target numeral (e.g., "5"), its location, and he gives his confidence in the correctness of his reports. (From Sperling, et al., 1971. Copyright 1971 by the American Association for the Advancement of Science.)

Spivak, & Johnson, 1971) we studied the ability of a subject to search for a single target numeral when the background characters were letters. Some salient results and conclusions were:

1. Telling the subject in advance which particular numeral (e.g., "5") will be presented on each one of a long series of trials does not improve the detectability of that numeral, relative to its detectability when the subject is searching for an unknown one-of-ten numerals. A related observation is that when the subject detects a target, he detects both its location and its identity. That is, he does not report seeing an unknown numeral in a particular location, or seeing a particular number in an unknown location. Detection of a numeral among letters implies both identity and location information.

2. The maximum number of characters a subject can scan in an array (max $\hat{l}$, his "span") is about 15–25 characters. Increasing the number of characters in the array beyond a subject's span does not improve his performance.

3. In viewing letters whose size is greater than about .5 deg, subjects approach their asymptotic performance when arrays are presented every 120 msec; increasing $t$ to 240 msec improves performance only slightly; increases in $t$ beyond 240 msec are of no benefit whatsoever. A corollary conclusion is that when a subject searches arrays of many letters naturally by means of eye movements (i.e., corresponding to a new input about every 240 msec), his processing capacity is unused for almost half of the time (between 120 and 240 msec after each eye movement).

4. The most efficient search (min $\hat{t}$) occurs when new arrays occur every 40 to 50 msec (corresponding to 20 to 25 fresh arrays per sec). In these presentations, most subjects achieve a $\hat{t}$ of less than 10 msec, corresponding to scan rates in excess of 100 characters per sec. When arrays are more closely spaced than every 40 msec, performance deteriorates rapidly.

5. The conclusion from these and related studies is that subjects scan in parallel 15 to 25 characters in an array.

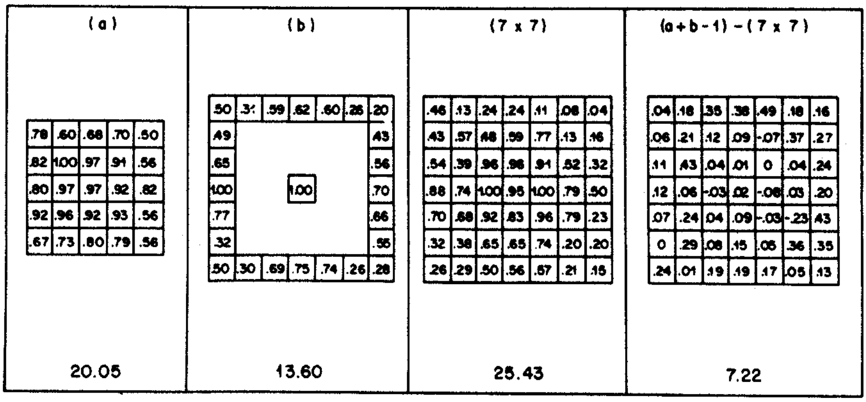


FIG. 2. Three display configurations tested in the search paradigm: 5 x 5, 7 x 7 frame plus one center location, and 7 x 7. The subject's task was to identify the target, an unknown one-of-ten numeral, that occurred at a random location among the letter background characters. The proportion of correct target identifications at each location is indicated at the location itself. The proportion of correct identifications (over all locations and all trials) times the number of stimulus letters ($\hat{l}$) is indicated under each display. The right-most panel indicates the location-by-location superiority of the sum of the parts (a + b − 1) over the whole (7 x 7).

At Attention and Performance V we reported on one reason for the limited capacity: local sharing of processing capacity between nearby locations of the same array. We compared the performance (number of locations scanned, $\hat{l}$) in separate experiments. Each experiment involved identification of an unknown one-of-ten numeral in one of the three arrays shown in Fig. 2: (a) a 5 x 5 array, (b) a 7 x 7 array, and (c), the sum of (a) and (b), a 7 x 7 array. Performance on the whole (c) is clearly less than the sum of performances on the component parts (minus the duplicated central location); $c < a + b - 1$. From a location-by-location analysis of the deficit we can characterize it as follows: The probability of correct identification at a location in (c) will be lower than at the corresponding location in (a) or (b) to the extent that the location in (c) has more neighbors. Thus, the drop in performance is most obvious at the locations that correspond to the corners of the 5 x 5 array. This is local interference (cf. Estes, 1972, 1975).

## II. VISUAL SEARCH UNDER MULTIPLE ATTENTIONAL STATES

In contrast to the local interference by neighbors, there is a global effect of attention in visual search we shall report here. In an attempt to produce an optimum array of characters for visual search, we constructed an array in which we made a modest attempt to match the number of characters in each local area to the density of information-processing capacity in that area. This array consisted of very small characters in the center, surrounded by larger and larger characters in each successively larger ring around the center. To our surprise, we found that performance was not improved. This led us to investigate the question of whether subjects can search in parallel for targets of different sizes. This question has been posed by Kinchla (1977) in a somewhat different paradigm. Kinchla found that instructions to attend to stimuli of a particular size influenced his subjects' decision criteria but not the quality of information they obtained from the stimuli. Nothing in his results would suggest the difficulty our subject had.

## A. Procedure

The experimental paradigm was the same as before except that the critical array now contained two target numerals, a large one and a small one, chosen independently, and the subject's task was to report both numerals, both locations, and both confidences. The spatial arrangement of the small characters in the *inside* and the large characters in the *outside* of an array is illustrated in Fig. 3. The size and number of the *inside* and *outside* characters were chosen so that in control experiments, in which the task of the subject was to report only *inside* or only *outside* numerals, the probability of a correct report was approximately the same for *inside* and *outside* targets.

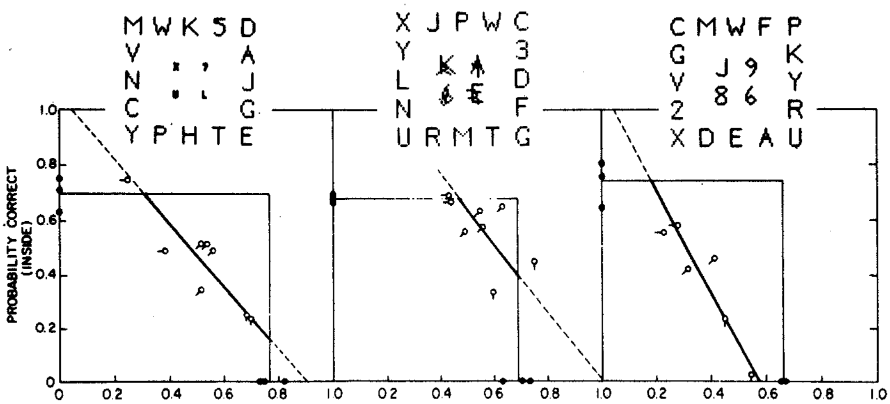


FIG. 3. Data from divided attention experiments. Typical stimuli, each with two targets, are illustrated at top. Each open circle represents data from one block of 30 to 60 trials (report of both targets); each filled circle represents data from a control session (report of only the designated class of targets). The direction of the "tail" on a data point represents the attention instruction: down indicates "give 90% of your attention to the outside," left indicates "90% to inside," diagonal indicates "equal attention." The vertical and the horizontal lines are best-fits to the control data; the diagonal lines are best-fits to the multiple detection data; the darkened portion is the estimated Attention Operating Characteristic (AOC).

In addition to the *Small* condition described above, two other conditions were investigated, *Noise* and *Reversal.* In the *Noise* condition, the *inside* was composed of large characters (the same size as the *outside*), but detection of a critical numeral was made comparably difficult by superimposing a randomly-chosen, squiggly line segment ("noise" segment) on each *inside* character. In the *Reversal* condition, background characters in the *inside* were numbers and the target was a letter.

In all conditions, the *outside* was the same. The background characters were all the letters of English except that the letters B, S, Z, Q, O, and I were omitted because of their similarity, respectively, to the numerals 8, 5, 2, 0, 0, and 1. In the *Reversal* condition, which was studied separately after the others, the numerals 0 and 1 also were omitted.

### B. Results

Some data from a typical subject viewing the display with *Small* characters are illustrated in the leftmost section of Fig. 3 (Sperling & Melchner, 1975, 1976a). The ordinate indicates the probability of correctly identifying the small target numeral (from the *inside*); the abscissa indicates the probability of correctly identifying the large target numeral (from the *outside*). In these experiments both numerals always occurred in the same array.

Each point in Fig. 3 represents data from a different block of trials. In some blocks the subject was instructed to give 90% of his attention to the large *outside* characters, and 10% to the small *inside* characters. In other blocks the percentages were reversed, and in still others he was told to pay equal attention to both. Figure 3 indicates that he was indeed able to follow these instructions, and that he was able to trade off performance on one class of targets against the other. The range of performances of which he is capable, as he varies his attention from being concentrated entirely on the small targets to entirely on the large targets, defines his "Attention Operating Characteristic" (AOC) for this task. In this task, the AOC is approximately a straight line with slope of $-1$, indicating that the subject can exchange a certain amount of probability on one task ($\Delta P_1$) for an equal amount on the other ($\Delta P_2$).

In control conditions, the subject was told to report just one kind of target (e.g., the *outside* target) for an entire block of trials. These data are graphed directly on the axes of Fig. 3. A vertical line is drawn through the mean of the *outside* control data, and a horizontal line through the *inside* mean. The intersection of these two points defines the "independence point," the point at which the subject would operate if he could perform both search tasks simultaneously without any interference, i.e., independently of each other. Insofar as the AOC lies inside the independence point, it represents some degree of interference between the two tasks.

The middle section of Fig. 3 illustrates performance in the *Noise* condition. The *outside* search task was the same as in the *Small* condition; the *inside* search task was matched to be of equal difficulty. Nonetheless, we see here that the AOC curve is closer to the independence point. This subject can carry out these two search tasks (with targets of equal size) with very little mutual interference.

The righthand section of Fig. 3 illustrates performance in the *Reversal* condition, in which the subject searches for a letter target among numerals on the *inside*, and for a numeral target among letters on the *outside*. The data show that the mutual incompatability of these two search tasks is nearly total.

## C. Controls and Further Experiments

Interference between two search tasks does not occur because of any memory deficit. To prove this, the display was altered so that the targets remained the same but each background character was replaced with just a single dot. In this case, subjects gave errorless reports of both targets. Thus, the subject's inability to report both targets in the experimental condition is due to the mutual interference of the two search tasks.

By occasionally putting the inside and outside targets into different arrays (instead of always in the same array) we can determine how long it takes the subject to switch attention from one to the other class of stimuli. We tested one subject extensively on the reversal task with intervals from the letter to the number target of ±480, ±240, and 0 msec. The results indicate that this subject could switch from one task to the other in from 0.24–0.48 sec.

In a quite different paradigm, Adam Reeves and the senior author (1976) had been able to measure the reaction time distribution for an attention shift, and found it to be similar to the distribution of reaction times for motor responses. The mean attentional RT is typically between 0.3–0.5 sec for the discrimination of one of a set of three target letters from a background of letters. These direct measurements of attention-switching latency are consistent with the indirect inferences from the search task.

In summary, we found that our various instructions to subjects profoundly influenced *what kind* of targets were detected, and *where* they were located. Shaw, Kohn, and Nemeth (1976) recently demonstrated that "the probability of a target being at a spatial location has a dramatic effect on how subjects distribute their attention or processing capacity over the visual field." Previously, Sperling (1960) observed that when attention was directed by a poststimulus cue to one row of a briefly-flashed three-row display, the cue exerted a profound effect on which letters were processed for recall – provided the cue occurred while there were still more letters visually available to the subject than he could ultimately report. The physically unobservable distribution of attention during (and immediately after) a brief exposure largely determines what will be detected and what will be recalled.[2]

[2]For a provocative, molecular analysis of some factors that might limit performance see Norman and Bobrow (1975).

## III. THE ATTENTION OPERATING CHARACTERISTIC (AOC)

The best way we know of quantitatively describing the effects of attention is the Attention Operating Characteristic (AOC). The AOC is quite analogous to the Receiver Operating Characteristic (ROC) in discrimination tasks. Consider two density functions $f_1(t)$ and $f_2(t)$, and a criterion, $c$. In a discrimination task, $f_1$ and $f_2$ could represent the density function of the noise ($f_1$) and of the signal plus noise ($f_2$) on a sensory continuum. As the subject varies his criterion $c$, he defines his Receiver Operating Characteristic (ROC) or preferably, his Discrimination Operating Characteristic (DOC), as we believe it should be called. The graph of a DOC is a mirror image of the ROC graph. The DOC graph is preferable because it represents better performance on each task (correct identification of signal when it occurs; correct identification of noise when it occurs) by increasing coordinate values (upward and rightward). All of the representations, ROC (or DOC) and AOC, have the useful property that the mixture of two strategies is represented along a straight line joining the two strategies, and that *any* mixture of strategies is represented at the center of gravity of the probability-weighted component strategies.

In attention tasks, let $f_1$ and $f_2$ represent the attention demands of two tasks. A good example is that of a student who wished to attend two courses offered at overlapping times in different classrooms on the same day. At the end of the day, he is given two test questions, once chosen from course 1, the other from course 2. The probability that the information needed to answer the test question is offered at any instant of the class period is given by $f_1$ and $f_2$ respectively for the two classes (we assume there is no repetition or redundancy in the classroom lectures.) The student is allowed to switch classrooms just once, by quickly running from classroom 1 to classroom 2 at time $c$. For each switching criterion $c$, his performance on questions about course 1 and course 2 defines his *Attending* Operating Characteristic (Fig. 4). Two AOCs are shown in Fig. 4 – one in which it is assumed that if the student were present in Class $n$ at the critical moment, his probability $p_n$ of correctly answering the subsequent test questions is 1.0, and another AOC for which $p_n$ is assumed to be 0.7. Note the similarity of Fig. 4c to Fig. 3.

The classroom analogy leads to a quite general formulation. The coordinate axes (arguments) of $f_1$ and $f_2$ need not be interpreted as time. For example, arranging "feature detectors" on a continuum according to the ratio $t$ of their utilities for task 1 and task 2 would yield a feature-detector utility interpretation for $t$. This is a mathematically equivalent basis and, in fact, corresponds closely to the likelihood ratio of signal to noise, which is an appropriate axis for ROC tasks. Moreover, the AOC method of analyzing and interpreting data from visual attention is equally applicable to other attention tasks, such as the sharing – or switching – of auditory attention between two ears.

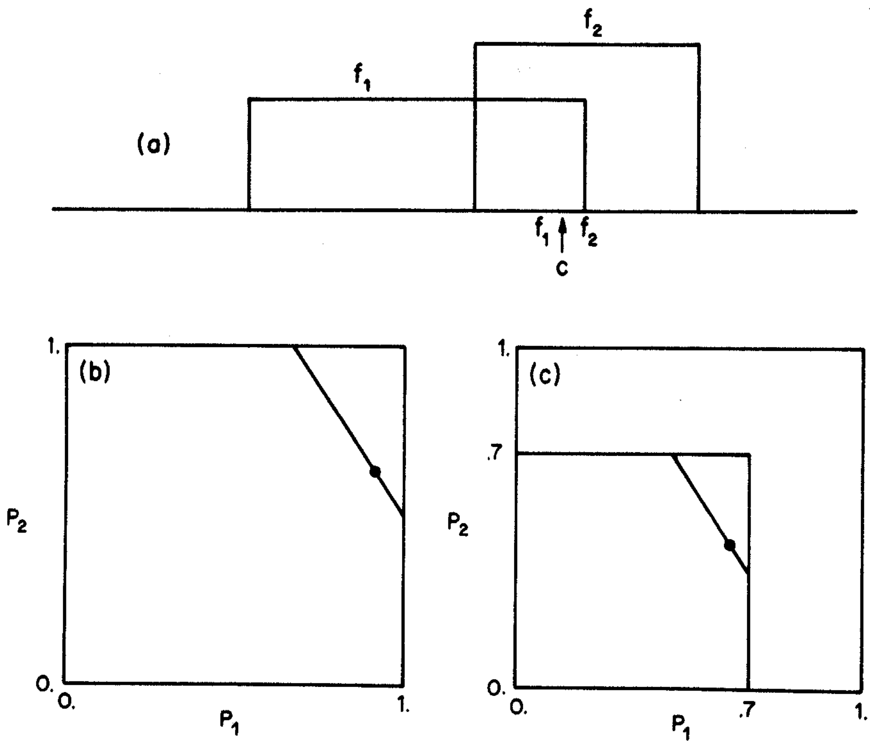


FIG. 4. Criterion Model of the Attention Operating Characteristic. (a) Density functions $f_1, f_2$ and the criterion. The abscissa represents time. In this example, $f_1$ and $f_2$ represent the density of information in two classrooms as a function of time, and $c$ represents the time of switching from classroom 1 to classroom 2. (b) Attending Operating Characteristic (AOC) for a student as a function of his switching time, $c$. The abscissa represents the probability $p_1$ of correctly answering a test question from classroom 1; the ordinate represents the probability $p_2$ of correctly answering a test question from classroom 2. The data point represents performance for the particular criterion $c$ shown in (a). (c) Same, but with the assumption that the student has only a probability $p$ ($p = .7$) of answering the test question correctly even if he was in the classroom during the instant the relevant information was presented. (See text.)

## A. Moving Along an AOC: Sharing or Switching?

In order to examine in more detail the mechanism by which the subject moves along the AOC curve, that is, the mechanism by which attention is shifted from one search task to the other, consider the 2 x 2 contingency table (Table 1) in which the joint occurrences of correct reports on the two tasks are tabulated for a single instructional condition. This table has three degrees of freedom; two of these, the marginals $P_1$ and $P_2$, are used to make the AOC. The third degree of

| TASK 2 | TASK 1: WRONG | TASK 1: RIGHT | |
|---|---|---|---|
| WRONG | $(1-P_1)(1-P_2)$ $+K$ | $P_1(1-P_2)$ $-K$ | $(1-P_2)$ |
| RIGHT | $(1-P_1)(P_2)$ $-K$ | $P_1 P_2$ $+K$ | $(P_2)$ |
| | $(1-P_1)$ | $(P_1)$ | |

TABLE 1. Contingency table for the joint probability of correct responses in a divided-attention experiment. When $k = 0$, the algebraic expressions represent the predictions of ideal shared attention (statistical independence of the two tasks); when $k < 0$, there is negative correlation representing some degree of attention-switching between different states, i.e., more than one criterion in the AOC model.

freedom (correlation) provides information about the mechanism of attention. We consider here two (of many) possible mechanisms: *sharing* and *switching.* The mechanisms are outlined below without formal derivations.

In sharing, attention is assumed to be divided between the two tasks in some fixed proportion, which does not vary from trial to trial within a single instructional condition. Insofar as there is less attention available for each task than in a control condition, performance suffers relative to the control. Here, shared attention corresponds to having a fixed criterion $c$ in Fig. 4a, and the probabilities of correct responses are directly proportional to the areas under $f_1$ and $f_2$ to the left and to the right, respectively, of $c$. The 2 x 2 contingency table is assumed to show statistical independence. By definition, in shared attention the probability of a correct response on one task is the same whether the response on the other task (for that trial) was correct or incorrect.

In two-state attention switching, two different component attention states $(S_1, S_2)$ are assumed to occur randomly, from trial to trial, in the search task. In the ideal case, the two component states $(S_1, S_2)$ are states of shared attention; each component state corresponds exactly to one of two criteria $c_1$, $c_2$ in the model. (However, the following analysis also applies to the case where the component states $S_1, S_2$ are themselves mixtures of yet other states.) To move along the AOC curve by switching attention, the subject is assumed to vary the proportion of trials he is in $S_1$ by switching between $c_1$ and $c_2$ from trial to trial. Two interesting properties of two-state attention switching are (1) mixtures of $S_1$ and $S_2$ produce a straight line AOC curve connecting $S_1$ and $S_2$; and (2) the contingency table for a mixed state is the mixture of the two separate contingency tables (from states $S_1$ and $S_2$).[3] From (2) it can be shown that, under the conditions of the experiment, we expect any contingency table produced by switching between states to have a negative correlation and to show nonindependence by the Chi-square test. By assuming the ideal case (i.e., two-state switching in which each component state is a state of shared attention), one can estimate the particular two states between which attention is being switched.

When this analysis is applied to the data described above, we discover that the major mechanism of altering attention is switching – i.e., altering the proportion of times the subject is in $S_1$ or $S_2$. On the other hand, we can also reject the hypothesis, at least for some of the data, that the subject manifests only two attentional states (i.e., only the two extreme states determined by the intersections of the AOC curve with the control condition lines). Thus, attention *sharing* is not merely the operative mechanism at the end points of the AOC, but is also a mechanism for moving along an AOC.[4]

## B. Conclusion

In a more general vein, the AOC is a useful way of studying attention, and particularly of describing the compatibility of two tasks. A pair of tasks to be performed simultaneously determines an AOC. To compare two pairs of tasks, one cannot use just one condition of attention for each pair, for this would be comparing one point from each of two curves and not comparing two curves. (An analogous problem occurs in signal detection theory with ROC curves.)

In conclusion, we see that subjects cannot simultaneously search two areas for a large and a small target as well as they can search for two different equal-sized targets. The instruction to search simultaneously with equal attention for a large and a small sized target causes the subject to switch his attention from trial to trial between searching primarily for large and searching primarily for small

[3]The *mixture* $S_3$ of two contingency tables $S_1$ and $S_2$ is defined as follows. Let the proportion of $S_1$ in the mixture be $a$, $0 \leqslant a \leqslant 1$. For any triple $p_1$, $p_2$, and $p_3$ of corresponding elements in $S_1$, $S_2$, and $S_3$, we define $p_3 = ap_1 + (1 - a)p_2$.

[4]In this simplified theory, when a negative contingency is observed at the endpoint of an empirically determined AOC, the negative contingency is assumed to be generated by switch-between two states, one of which lies beyond the observable endpoint, at the "true" endpoint of the AOC.