EXAMINING THE VIABILITY OF THE NEISSER SEARCH MODEL IN THE FLIGHT DOMAIN AND THE BENEFITS OF HIGHLIGHTING IN VISUAL SEARCH

Ashley Nunes, Christopher Wickens and Shanquin Yin
University of Illinois, Department of Psychology and Beckman Institute
Champaign, IL

We conducted four experiments to examine the visual search capability of pilots and examine how search times could be reduced, invoking the framework of Neisser's serial self terminating (SST) search model. In Experiment 1 which manipulated target presence and set size, our results show increases in response time as array size increases: an effect predicted by the SST model. In Experiment 2, we manipulated the background against which the visual search had to be performed. We found that the detrimental effects of array size on visual search time are maximal when the search is performed against maps, which have large quantities of terrain information. In Experiment 3, we sought to assess the degree to which high and low-lighting could be used to reduce the search times of pilots. We observed that although both high and lowlighting reduced the search times of pilots, the observed reductions were not predicted by the SST model. The advantage of highlighting elements of the target was shown in Experiment 4.

INTRODUCTION

The designer of advanced aviation displays is confronted with two conflicting pressures. On the one hand, there is a desire to provide the pilot with a great deal of integrated hazard information, in which traffic, weather, airspace information and terrain may be displayed on a single map (Kroft & Wickens, 2003). On the other hand, such designs present the enduring problems of clutter. Valid quantitative models will have the advantage of enabling precise predictions of "how much clutter is too much" (an important question to be answered for display certification), and of predicting the value of solutions to clutter problems through highlighting or decluttering (Remington et al., 2001, Wickens et al., 2004).

The serial self terminating search model of Neisser et al. (1964) provides a foundation for such a model, stating that the search time through an array of N elements is predicted by the formula: $ST = a + bN/2$ when a target is present, and $a + bN$ when the target is absent, where b is the time required to examine each item and determine that it is **not** the target, and a is the intercept term that characterizes the "readout" of the actual target. The division by two reflects the fact that, when a target is present, it is located, on the average, after searching half the array. This serial self terminating (SST) search model also predicts that, if there is a coherent search order (e.g., top→ bottom), the location of the target will contribute to the search time. Some aspects of the Neisser SST model have shown limited success in predicting search time both in air traffic control (ATC) displays (Remington et al., 2001), and in soldier-used electronic maps (Yeh & Wickens, 2001). The general goal of the present research was to establish the viability of the serial self-terminating model for pilots searching a traffic map over the background of varying types of ground map clutter. Three specific subgoals were:

(1) examining the viability of the model
(2) examining how, within the context of the model, maximum allowable clutter effects might be predicted
(3) exploring two alternative means of reducing clutter.

Three experiments are described in which (1) pilots search a schematic traffic map with no background clutter, as the number of elements is increased (2) map overlays are provided to induce a background clutter of varying levels and types, (3) two alternative ways of using intensity coding (highlighting/low lighting) to reduce the amount of information that must be searched (e.g., "decluttering" the display).

GENERAL METHOD

All experiments used similar methodology in which a target aircraft was identified (aircraft at a particular altitude), then a map was presented, and pilots searched the map until the target was located, positioning a cursor on the target and depressing a mouse button to indicate the target was located. On some trials (10% in Experiment 1), where no target was present, a separate button was pressed. Experiments 2 and 3 had no target-absent trials.

EXPERIMENT 1: BASELINE, NO-MAP BACKGROUND

Subjects

The 16 pilots had an age range from 18 to 23 years (mean = 19.3) and an experience range from 15 – 300 flight hours (mean = 105.6).

Display Design

The search array consisted of multiple aircraft being presented against a blank background (Figure 1). The pilot's own aircraft was always located in the center of the screen and was clearly denoted by the label "OWNSHIP.' In contrast, the other aircraft had data tags, which conveyed the aircraft's call-sign, altitude, speed and heading (the latter could also be inferred from the orientation of the aircraft icon).

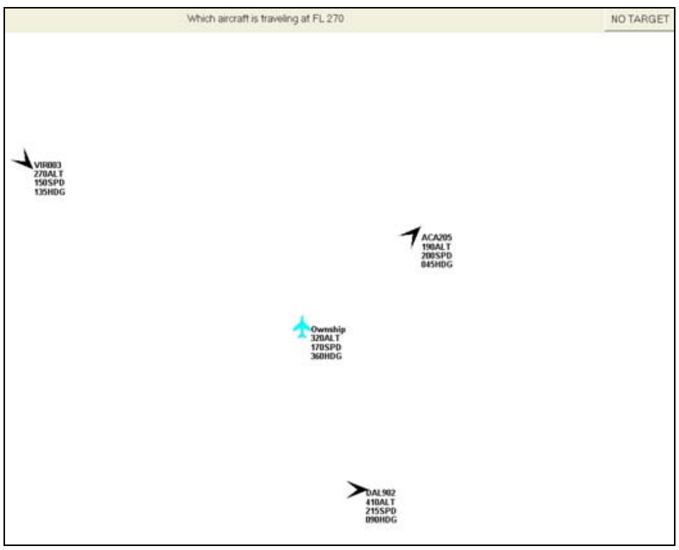


Figure 1: Screen shot from Experiment 1.

Design and Variables

A within-subjects design was used, with two independent variables being manipulated, namely target presence and aircraft array load. On 90% of trials (54 trials), the target was present whereas on remaining 10% of trials, the target was absent. There were either three, five or seven traffic aircraft present on any array. The order in which the trials were presented was randomized across participants to control for order effects.

Results

RT analysis (Figure 2) revealed a significant main effect for target presence ($F(1,15)=102.72$, $p<.01$), where responses to target present trials were 3 sec faster than responses to target absent trials. There was also a linear effect of aircraft array size ($F(2,30)=42.71$, $p<.01$), with increases in response time being evident as the array size increased from three to five to seven aircraft. The interaction between the target presence and array size variables was also significant ($F(2,30)=3.57, p<.05$), where the effects of increasing array size were more pronounced on target-absent compared to target-present trials, a pattern quite consistent with the serial self terminating model. Also supporting the model was the serial position effect whereby targets located in the upper left quadrant were located more rapidly than those to the right, and/or below. There was no effect on error rate.

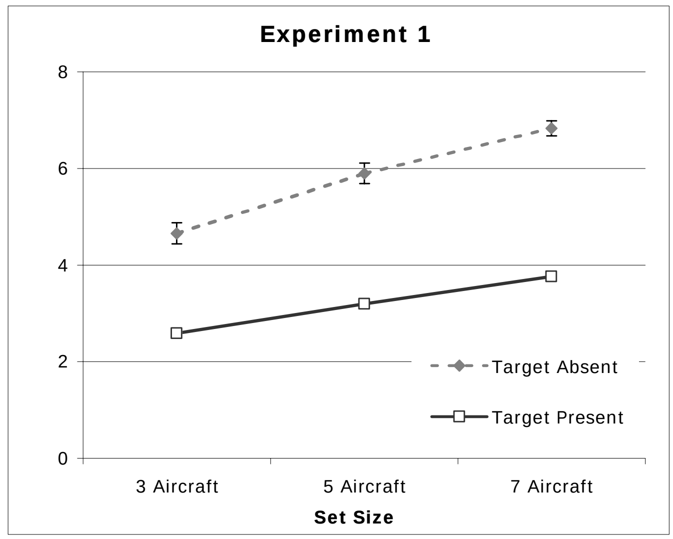


Figure 2: Response time data from Experiment 1.

Discussion

In Experiment 1, four properties of the data were consistent with the serial self terminating model: (1) search time is longer when the target is absent. (2) search time increased linearly with set size and (3) more so when the target is absent. (4) the serial position effect indicated a sequential search process.

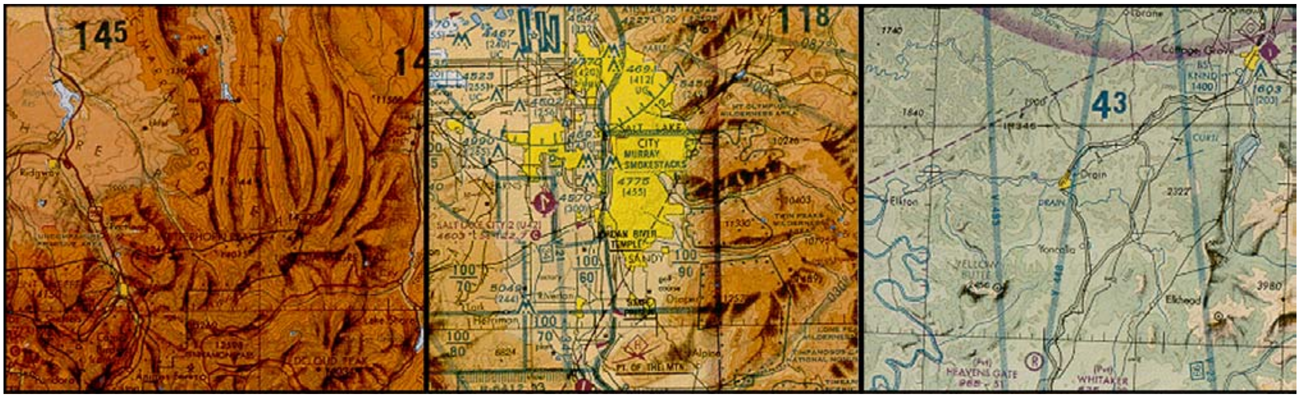
EXPERIMENT 2: SEARCH WITH BACKGROUND MAPS

Subjects

The 16 pilots ranged in age from 18 to 23 years (mean = 19.1) and in experience from 20 – 350 flight hours (mean = 78.9).

Task and Variables

The design and task in Experiment 2 was similar to that in Experiment 1. However, only target present trials were used and whereas the manipulation of aircraft array



Map 1: Terrain Clutter Map 2: Readout Clutter Map 3: Sparse Clutter

Figure 3: Maps used in Experiment 2.

size was preserved (three, five or seven aircraft) we also manipulated the background against which the array was presented. Three maps were tested with each map varying in the type of clutter it possessed (Figure 3). One map contained ‘terrain’ clutter (i.e., high density terrain features). A second map contained ‘readout’ clutter (i.e., features that would make it difficult aircraft to read the data-tags of aircraft in the array due to writing on the map itself). A third map contained relatively sparse features, designed to approximate the blank map of Experiment 1. A fully factorial within-subjects design was used. The procedure was identical to that used in Experiment 1.

Results

Analysis of response time data (Figure 4) revealed a significant main effect for map type ($F(2,30)=3.91$, $p<.05$) with the slowest RT for the terrain clutter map. Examination of Figure 4 revealed both set size and its interaction with map type, while both significant ($F(2,30)=4.95$, $p<.05$), and ($F(4,60)=4.25$, $p<.05$) respectively, were non monotonic effects across set size, and hence at odds with the Neisser model. Subsequent analysis revealed that when the features of the map were sparse (Map 3), there was no effect of array size ($F(2,30)=0.18$, $p>.83$). Contrastingly, the effects of array size showed a non-significant but monotonic trend in the expected direction when the search was conducted against the maps that contained rich terrain features (Map 1) ($F(2,30)=2.19$, $p<.12$) and was significant, although non-monotonic when it contained textual (Map 2) features ($F(2,30)=7.88$, $p<.05$). There was a position effect similar to that found in Experiment 1 (fastest search in top left quadrant with no effect on errors.

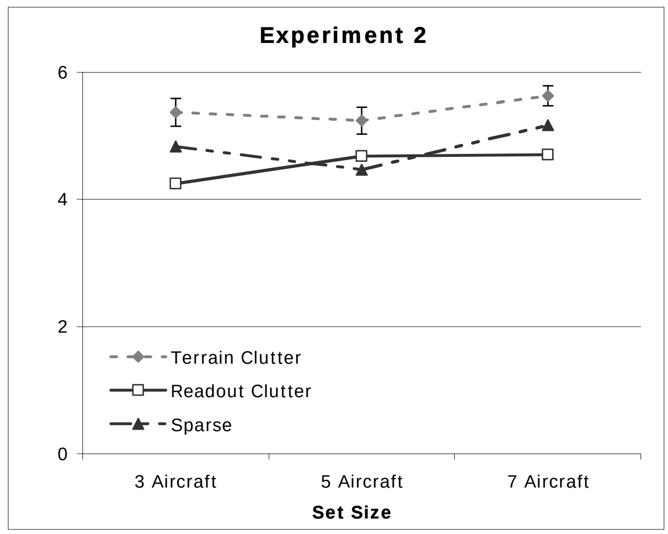


Figure 4: Response time data from Experiment 2.

Discussion

The results from Experiment 2 are not in concordance with those of the first, whose results were successfully predicted by the SST model. In Experiment 2, linearity (and even monotonicity) across set size, the key signature of the serial model, was absent. Most noteworthy was the profound cost of terrain clutter, relative to the sparse or text clutter (an effect replicated in a subsequent experiment, not reported here: see Wickens et al, 2005). We may attribute this cost to the low contrast.

In addition to eliminating the set size effect, the increases in search time observed here over maps with no background (seen in Experiment 1) may be attributed to the relative difficulty associated with being able to discriminate between textual information associated with the map and that associated with the target itself . Support for both these explanations comes from the 'readout' times (i.e., the '*a'* intercept of the SST model) observed for all three maps (approximately 5 seconds) which was higher that that observed in the baseline condition (which was approximately 2 seconds). In study three, we sought to determine and quantify a performance benefit in terms of search time associated with two kinds of highlighting the target. In both cases, benefits could be associated with specific features of the SST model: reducing N, and reducing b.

EXPERIMENT 3: SEARCH WITH HIGHLIGHTING

Subjects

Twenty-two pilots were assigned to one of two groups, namely the plane highlighting and the text highlighting group. Groups were matched for gender, age, years of education and amount of flight experience.

Task and Variables

The same search task was employed in the present study. However the search was always performed against a map with terrain clutter, there were always eight aircraft on the screen and only altitude questions were used. In the baseline condition, participants in both groups performed the search task unassisted. In the plane highlighting condition participants had half of their search array (i.e., 4 aircraft icons but not their data-tags) highlighted in blue, with the target being included as one of the highlighted aircraft. The goal here was to assess the degree to which search times could be reduced by requiring pilots to only search half the array (four versus eight aircraft). For pilots in the text highlighting group, none of the aircraft were highlighted. However, the heading and speed information presented in all the aircraft's data-tag was grayed out to reduce the amount of information the pilot had to scan through in order to find the altitude information. Experiment 3 employed a between-subjects design with the type of highlighting condition serving as the between-subjects variable.

Procedure

The procedure used was similar to that used in the two previous experiments with the order of conditions (baseline and experimental) being counterbalanced across participants.

Results

To accurately assess the benefits associated with highlighting, a series of t-tests were conducted between specified groups, given the unequal sample size that existed between the baseline condition (22 subjects) compared to the highlighting/data-tag condition (11 subjects subject per group). As shown in Figure 5, while the baseline RT was the same for both groups, reducing the number of planes to be inspected, from 8 to 4 reduced RT (by 1.1 sec) substantially more ($t_{(10)}$=4.34, p<.01) than did reducing the number of items in each data tag to be inspected (0.6 sec, ($t_{(10)}$=1.46,p>.17). There was no effect on errors, and in this experiment, no effect of target position.

Figure 5: Response time data from Experiment 3.

Discussion

The results clearly indicated that reducing the number of items to be searched (N), was more effective than reducing the amount of text that needed to be read/item (b). The former effect was what would be predicted from the Neisser model. However the resulting gain (for eliminating 4 items) of 1.1 sec was substantially less than what would be predicted from the data in Figure 2 (Experiment 1), Here the predicted reduction can be realized by the reduction from searching all 7 targets (target absent RT: 6.8 s.) to searching all 3 targets (target absent 4.6 s.), a predicted gain in Figure 2 of 2.2 sec.

In explaining why the 1.1 sec benefit of effective set size (N) reduction in Experiment 3 was greater than the 0.6 sec benefit of text reduction, we hypothesized that the smallness of the latter benefit resulted from the *consistency* of location of the necessary altitude information in the baseline condition. That is, the relevant altitude information was always located on the second line of the data tag (see Figure 1). Pilots, even in the baseline condition, could start and stop their examination of each item there, and so effectively "eliminating" (or lowlighting) the two items below (airspeed and heading) would have no benefit. To confirm this, we carried out a fourth experiment in which the requested information in the data tag varied from question to question. Under these circumstances we now observed a significant ($t_{(34)}=3.90, p<.01$)) benefit of 1.5 Seconds, one that is essentially (and statistically) equivalent to that of the plane highlighting.

CONCLUSION AND PRACTICAL IMPLICATIONS

The current results unambiguously show that clutter hurts, in at least three ways: (1) Increasing the number of items to be inspected (all three experiments), (2) Increasing the time for each item to be inspected, by presenting a cluttered background (Experiment 2), in which case terrain appears to be particularly "lethal" given its reduction in target contrast. Here however, even a sparse map with little on it (Experiment 2), substantially increases search time compared to a fully blank background (Experiment 1). (3) increasing the uncertainty of location of text to be inspected for each target.

In spite of these clear clutter effects, the simple Neisser SST model was somewhat illusive, only exhibiting the clearly predicted performance in Experiment 1. In particular, the straightforward prediction of longer search time with greater set size vanished when the search took place against any sort of cluttered map background. Nevertheless there were some tantalizing but inconsistent hints of the predicted SST process present here: the well validated serial position effect was observed in Experiment 2 (but not 3), and the search time in Experiment 3 was reduced by eliminating (graying) half of the items from the search field (but not as much as predicted).

In the absence of the predicted serial self terminating effects, the issue of "how much clutter is too much" becomes more difficult to answer, since in the absence of a strong linear effort of set size, (N), it is impossible to define a particular set size above which RT will exceed some designer designated criterion time (e.g., 5 seconds). Furthermore, the data did not reveal any "knee" of an accuracy curve (e.g., above search of N elements, errors begin to occur) which also might have helped to establish a clutter "red line".

Nevertheless, the clear costs of clutter through its different manifestations here, did support the effectiveness of both solutions to improve performance: reducing the number of items to be searched, as well as reducing the amount of information to be searched with each item, although the latter was only viable to the extent that, without such highlighting, the necessary information was inconsistently located.

The finding that terrain clutter was most disruptive suggests that heavy efforts should be made to maximize the contrast, by lowlighting background terrain relative to traffic information (see also Podczerwinski, Wickens & Alexander, 2001). Finally, the effectiveness of reducing the effective target set will, of course, depend on the extent to which the needed target is within that highlighted set; that is, the issue of highlighting validity (Fisher & Tan, 1989). We did not vary this reliability in the current experiment, but it is important to consider the extent to which computer automation (or pilot-selectable highlighting tools) may "know" in advance the restricted set that contains the target of interest.

ACKNOWLEDGMENTS

This work was supported by NASA Langley Research Center, with Lance Prinzel serving as the technical monitor. Special thank goes to Sharon Yeakel for her programming assistance. Ashley Nunes is currently a Senior Scientist in the Cognitive Systems Division at Charles River Analytics Inc, located in Cambridge, MA.

REFERENCES

Fisher, D. L., & Tan, K. C. (1989). Visual displays: The highlighting paradox. *Human Factors, 31*, 17-30.

Kroft, P. D., & Wickens, C. D. (2003). Displaying multi-domain graphical database information: An evaluation of scanning, clutter, display size, and user interactivity. *Information Design Journal, 11*(1), 44-52.

Neisser, U., Novick, R., & Lazar, R. (1964). Searching for novel targets. *Perceptual and Motor Skills, 19*, 427-432.

Podczerwinski, E. S., Wickens, C. D., & Alexander, A. L. (2001). *Exploring the "out of sight, out of mind" phenomenon in dynamic settings across electronic map displays* (ARL-01-8/NASA-01-4). Savoy, IL: University of Illinois, Aviation Research Laboratory.

Remington, R. W., Johnston, J. C., Ruthruff, E., Gold, M., & Romera, M. (2000). Visual search in complex displays: Factors affecting conflict detection by air traffic controllers. *Human Factors, 42*(3), 349-366.

Wickens, C. D., Nunes, A., Alexander, A. L., Steelman, K. (2005) 3D Navigation and Integrated Hazard Display in Advanced Avionics: Workload, Performance, and Situation Awareness Display Clutter Issues in SVS Design. *University of Illinois Human Factors Division Technical Report AHFD-05-19/NASA-05-7*

Yeh, M., & Wickens, C. D. (2001). Attentional filtering in the design of electronic map displays: A comparison of color-coding, intensity coding, and decluttering techniques. *Human Factors, 43*(4). 543-562.