Aviation Research Lab
Institute of Aviation

University of Illinois
at Urbana-Champaign
1 Airport Road
Savoy, Illinois 61874

Pilot Task Management: Testing an Attentional Expected Value Model of Visual Scanning

Christopher D. Wickens,
John Helleberg, Juliana Goh,
Xidong Xu, & William J. Horrey

Technical Report
ARL-01-14/NASA-01-7

November 2001

Prepared for

NASA Ames Research Center
Moffett Field, CA

Contract NASA NAG 2-1120

ABSTRACT

Two models of information acquisition in visual scanning are described. A descriptive model identifies the role of event salience, effort, expectancy and value in influencing where and when people look at different channels to sample information in dynamic environments. An optimal prescriptive model accounts for the role of expectancy and value, as these characterize the properties of channels necessary to serve tasks that may differ in their importance. The prescriptive model is validated against the data from four experiments in which skilled pilots flew a high fidelity visual flight simulator, while engaged in traffic detection with (Experiments 1, 3, and 4) and without (Experiment 2) a cockpit traffic display, and with different forms of data link displays (Experiment 4). All four experiments provided a good fit between model predictions and the percentage of time that pilots spent viewing different areas of interest in the airplane environment. Implications of the model to training optimal strategies, and to possible refinement of training policies are described.

INTRODUCTION

Models of visual information acquisition can generally be placed into one or two categories of two sorts. On the one hand, psychologists have extensively modeled the process of visual search (e.g., Brogan, Gale, & Carr, 1993; Wolfe, 1994; Neisser, Novick, & Lazar, 1964; Teichner & Mocharnuk, 1979). In such endeavors they have contrasted serial and parallel search models, and created hybrid models in between (Bundensen & Pedersen, 1983), and such models have been useful in both basic laboratory paradigms as well as more applied domains such as those involved in searching graphs (Gillan & Lewis, 1994; Lohse, 1993), maps (Wickens, Kroft, & Yeh, 2000), menus (Fisher & Tan, 1989) or roadway environments (Theeuwes, 1994). A key facet of such models is the emphasis on **locating** a single target, and an emphasis on search **time** as the critical dependent variable (although see Drury, 1975 for the importance of accuracy in search models).

In contrast to search models, another class of models has focused on supervisory control/sampling, using, primarily visual sampling or scanning as a dependent variable (e.g., Moray, 1986; Senders, 1964, 1983; Carbonell, 1966; Carbonnell, Ward, & Senders, 1968; Ellis & Stark, 1986; Sheridan, 1970). Such models have typically been more engineering based, and focus very much on the eye (as measured by visual scanning) as a “single server queue”. Four key features distinguish these from the visual search models above. (1) The operator is not looking for a static target, but is rather supervising a series of dynamic processes, such as temperature gauges, or aircraft movements. (2) The primary focus of the models is on noticing critical **events** at relatively consistent spatial locations, rather than finding critical **targets** at uncertain locations. (3) The key dependent variable is not target detection RT, but is instead the proportion of visual attention distributed to various “areas of interest” (AOIs) as a function of the quantitative properties of those AOIs. (4) The process of defining specified AOIs means that the challenge in visual attention is not so much knowing **where to look** (e.g., to find a target), but in knowing **when to look where** to assure that the dynamic processes are under control, and that the necessary information to understand those processes is retrieved in a timely manner.

A characteristic that permeates many of the visual sampling models is that of **optimal strategies** of attention allocation, designed to maximize or minimize some benefit or cost function, given the scarce resources of the “single server queue” (visual attention). As an example, Senders’ (1964) original search model focused on the optimum sampling of different dynamic AOIs’ as a function of the bandwidth (event rate) of signals located there, employing optimal sampling theory. His model was subsequently elaborated by others (Sheridan, 1970; Sheridan & Rouse, 1971; Carbonell, 1966; Carbonnell, Ward, & Senders, 1968; Tulga & Sheridan, 1980; see Moray, 1986 for a good review) to account for value, in addition to bandwidth, and to dictate optimal scanning strategies, given the **expected value** of perceiving information at different AOIs (or the expected cost of missing critical events at those AOIs). Such models, echoing normative expected value models of decision making (e.g., Edwards, 1961), penalize attention allocation performance to the extent that channels that contain high probability and important (valuable) events are undersampled. That is, attention allocation should be directly proportional to the product of probability and value.

The expected value model may be considered **optimal** or prescriptive in the sense that only two properties should drive the allocation of attention: expectancy and value. An operator who possesses a well calibrated mental model (Smallwood, 1967), that captures the objective levels of these two parameters will minimize the chance of missing important information. Of course expectancy is driven not only by the bandwidth or frequency of events occurring along a channel, but also by any contextual cueing that may signal the appearance of information along an otherwise low bandwidth channel. This might, for example, characterize the role of an alarm, signaling the operator to look at the visual display of the indicator variable that triggered the alarm. Under normal circumstances the indicator changes rarely (low bandwidth) but now will be sampled at a time other than that dictated by its low bandwidth. One general finding that comes from the earlier research on sampling, is that people tend to sample low bandwidth channels somewhat more frequently than the optimal models predict, a characteristic attributed, in part, to people’s limited working memory of the exact state of a channel when it was last sampled (Sheridan, 1970; Moray, 1986).

In addition to expectancy and value, there are two other important factors that also influence the frequency of visual sampling. First, Kvalseth (1977) and Sheridan (1970) have both identified the inhibiting role of information access **effort** required to sample information. Eye movements are “cheap” but not “free”, and in some environments when a head movement is also required to sample information, the cost of such samples can be quite high (while wearing cumbersome head gear or, for the pilot, making head movements while engaged in vertical or lateral maneuvering can cause vestibular disorientation). In addition to the effort required by attention movement, the effort required of concurrent cognitive or perceptual tasks can also inhibit the control of visual scanning (Liu & Wickens, 1992) or information access. The second additional factor is the **salience** or conspicuity of an event that occurs on a channel (or within an AOI), a factor that can capture attention. While this property received little attention from the engineering models of supervisory sampling, it has been a cornerstone of the psychological models of visual search, with explicit focus on the concept of “attentional capture” and its causes (Yantis, 1993; Wolfe, 1994; Folk, Remington, & Johnston, 1992; Pashler, Johnston, & Ruthruff, 2001).

Given the above discussion, it should then be feasible to combine the influence of the four factors driving visual attention into a **descriptive** model of scanning, characterizing the distribution of visual attention across areas of interest, (or the probability that a given area will be attended)

(1) $P(A) = sS - efEF + (exEX)(vV)$.

We may refer to this as the SEEV model, as derived from its terms. Each term in capital letters is a characteristic of a particular environment that is determined by (1) the physical properties of events (Salience = S), (2) the physical distance between a previously fixated and a current AOI, or the demands of concurrent tasks (Effort = EF), (3) an information-related measure of event expectancy (e.g., bandwidth, event rate; Expectancy = EX), and (4) an objective measure of the value (=V) of processing information at the AOI in question (or the cost of failing to attend there). The coefficients, s, ef, ex, and v, represent the relative influence of these four factors on human scanning.

It is apparent from the model and the previous discussion, that the two added components (S and EF) that distinguish the optimal prescriptive model, from its descriptive counterpart may exert a greater or lesser influence on scanning to the extent that designers have adhered to good human factors practice in display layout, by correlating EF with EX, and S with V. In particular, to the extent that high expectancy, high bandwidth sources of information are close together (low EF), this will attenuate the inhibitory role of effort in seeking information. Such a correlation represents the **frequency of use** principle of display layout (Wickens, Vincow, Schopper, & Lincoln, 1997). To the extent that valuable information is made salient when it occurs, this will assure the capture of attention when important events occur: for example, the role of alarms in good human factors practice (Stanton, 1994). Correspondingly, making less valuable events less salient will inhibit undesirable failures to focus attention appropriately.

It is also apparent that effective training, creating a well calibrated mental model (Smallwood, 1967) of expectancy and value, should be used to overcome any potentially negative influences of salience and effort, when these are not explicitly correlated with value and effort respectively. Indeed Moray (1986) observes that those empirical evaluations of the information sampling models that have used more highly trained operators (such as skilled pilots; Carbonnell, Ward, & Senders, 1968) tend to show fewer departures from optimal prescriptions, than those employing less skilled subjects (e.g., Sheridan & Rouse, 1971).

As noted above, some researchers have evaluated components of the model, although they have typically done so either in relatively “abstract” context free environments (e.g., Senders, 1964; Kvalseth, 1977; Sheridan & Rouse, 1971), or, when evaluated in more realistic settings, with more skilled participants (e.g., Carbonnell et al., 1968; Moray, Richards, & Low, 1980), have done so with only a small sample of subjects. Furthermore, the model validations that were carried out, were typically accomplished on only a single set of participants. Lacking was any cross validation, whereby the model fits obtained on one sample were validated on another sample. The goal of the research we report here is to provide validation and cross validation of a version of the expected value model of information seeking and attention allocation, with a larger sample of well trained pilots.

The Optimal Prescriptive Model

Our model also extends previous models of information sampling in an important respect. While previous models have defined properties of each AOI or channel, purely in terms of the bandwidth and value of events **along that channel**, we explicitly consider the many—to one mappings of tasks to channels, and channels to tasks in complex environments that makes this parameter assignment more difficult and requires the integration of models of information seeking with those of **task management** (Chou, Madhavan, & Funk, 1996; Dismukes, 2001; Raby & Wickens, 1994). This integration is shown schematically in Figure 1. At the top, the simpler models such as those of Senders (1964) and Carbonell (1966) assign tasks, values and bandwidths on a 1-1 mapping to AOIs. At the bottom, we see the structure of the current expected value model, as explicitly used in the framework of a pilot scanning three areas of interest (the instrument panel, the outside world, and a cockpit display of traffic information, or CDTI) to support two different tasks: the task of **aviating**, which involves maintaining accurate control over the "inner loop" aircraft flight parameters of pitch, roll and yaw, in order to keep the aircraft from stalling, and the task of **navigating** which, in our paradigm, requires the pilot to be aware of, and (if necessary) navigate around, any other air traffic in the forward path.

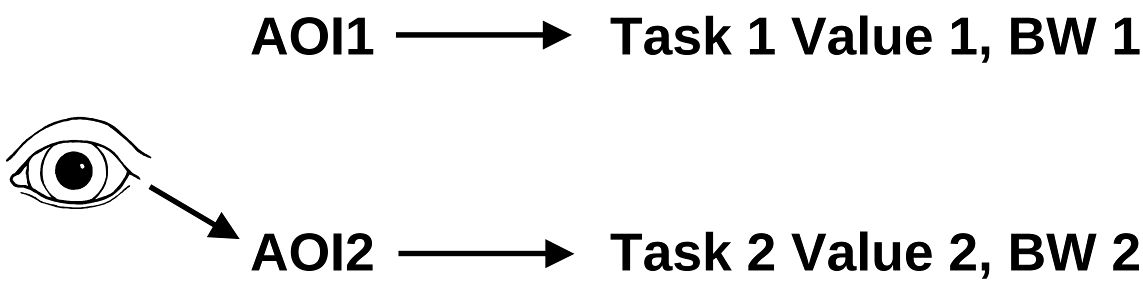
Within aviation, there is a clearly established task priority hierarchy which sets aviating to be a more important task than navigating (although both are more important than communicating; Schutte & Trujillo, 1996). After all, a plane cannot navigate if its lift is not sufficient to keep it in the sky (a failure to aviate), whereas a plane can aviate (fly) even if its heading and altitude are unknown (failure to navigate). As shown in Figure 1, this task hierarchy defines the importance or **value (V)** of AOIs that serve the respective tasks. However aircraft piloting also is such that, for example, the task of aviating (knowing the attitude of the aircraft relative to the horizon) can be supported both by the attitude indicator on the instrument panel (IP), and by the view of the true horizon in the outside world (OW). Correspondingly, the detection of traffic to be avoided (supporting navigation with heading or altitude changes) can also be supported by both the direct view of that traffic out the window (OW) and the cockpit traffic display (CDTI). These relations are also shown in Figure 1.

Accordingly, as shown at the bottom of Figure 1, an expected value model would predict that the probability of attending to a particular AOI is related to the sum, across all tasks supported by that AOI, of the value of those tasks, multiplied by the degree of **relevance** (importance) of the AOI for the task in question, and by the bandwidth of their informational "events". Expressed computationally, this is then represented by the **prescriptive**, or optimal form of the model (in contrast to the descriptive Equation 1) as:

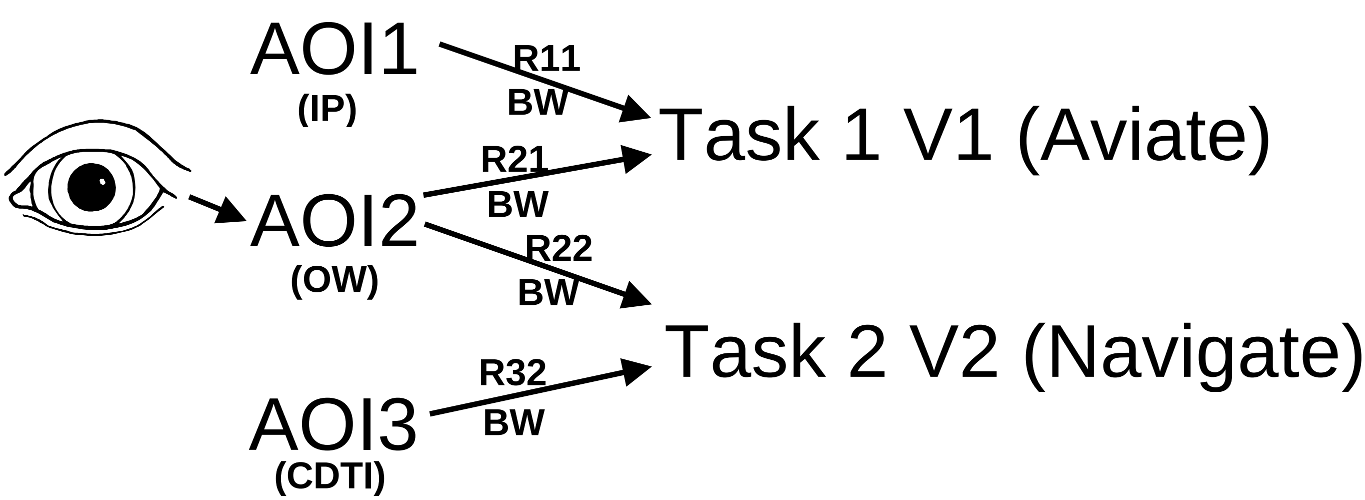
$$(2) \qquad P(AOI_j) = \sum_{t=1}^{n} (bB_t)(rR_t)(V_t)$$

where t = tasks, numbering 1-N.

SIMPLE SCAN MODEL (Senders)



AVIATION SCAN MODEL



Visual Attention (Scan) to AOI =

$$\sum_{\text{TASKS}} \text{[(BW x relevance(value) of AOI to task x task priority]}$$

Figure 1. Representation of different scanning models.

Notice that Expectancy is incorporated by bandwidth, while Value is incorporated by the product of relevance and the value (rank) of the task in the priority hierarchy. Whether "bandwidth" is assumed to be a general property of an AOI (and therefore can sit outside the summation), or is assumed to be a specific property of each task serviced by the AOI (and therefore must sit inside the summation, as shown in Equation (2)), may be determined by the particular application. Equation (2) is obviously different from the original SEEV descriptive model (Equation 1), in that the former does not contain the non-optimal salience and effort factors. Our goal is to assess how accurately the simpler optimal model can account for the performance of well trained pilots. After we report the results of this exercise, we will discuss the possible influences of salience and effort.

In providing numerical predictions for such a model, two general steps are required. First, as shown in Figure 2, matrices must be set up, to accommodate the different parameter estimates of bandwidth, relevance and value, that will be imposed in the varying conditions whose visual scanning (attention allocation) measures can be employed to validate the model. Second (and more challenging), one must assign the numerical coefficients to the different cells of the matrix, across the varying conditions, in order to compute model predictions.

<u>Bandwidth</u>

AOI

A	B	C

<u>Relevance</u>

Task		A	B	C
	X			
Task	Y			
	Z			

	Priority
X	
Y	
Z	

Figure 2. Generic matrices used to compute scanning predictions from the model shown in equation (2). Three AOI's (A, B, C) are depicted. The label "task" includes both separate tasks (such as aviate, navigate) as well as the same task under different conditions (such as aviating while maneuvering, or while flying straight and level). Thus there will be as many data points to predict, as there are cells in the "Relevance" matrix.

Rather than trying to estimate absolute values of some of these parameters, the approach we take in coefficient assignment is based on a "lowest ordinal algorithm". That is, across the range of conditions and AOIs employed (i.e., the rows and columns of the matrices in Figure 2), we simply order the values from highest to lowest (e.g., aviating priority is higher than navigating priority). Second, we assign the lowest values of these parameters that can still preserve all ordinal relations within the rows and columns of the matrices. As an example we shall encounter below, we may, in condition I, initially assign aviating a priority value of 2, higher than navigating (lower priority = 1); but once we include another condition, II, in which aviating is more important than it was in condition I, then aviating will need to take on **two** lower priority values (1 and 2), and so the priority value of aviating must be elevated to become 3. Generally we try to apply integer values. This lowest ordinal algorithm has the advantage that coefficients can be set based on simple relationships that multiple model users should agree upon.

MODEL VALIDATION

We report below the data from four aviation experiments used to generate data to validate (Experiments 1 and 2) and cross validate (Experiments 3 and 4) the optimum expected value model. The procedural details of Experiments 1, 2 and 3 are reported elsewhere (Wickens, Helleberg, & Xu, in press; Helleberg & Wickens, in press). However these articles do not report the efforts to model the scanning data which we report here.

All four experiments have the following common features.

Participants. All experiments employed well trained pilots who were flight instructors at the University of Illinois Institute of Aviation. Flight hours ranged from X to Y. All participants were paid for their participation.

Tasks/AOIs. Pilots always flew a general aviation flight simulator, with full outside visual projection (Figure 3) on a series of flight legs, defined by target altitude, heading and airspeed parameters (Figure 4). On most legs, one or more traffic aircraft was potentially visible. The pilots primary task was always to **aviate**. That is, to maintain stability of the aircraft by appropriate control of attitude (pitch and roll). As in conventional non-automated flying, pitch and roll also controlled altitude and heading respectively. These components of altitude and heading were fundamental to the second task, **navigating**, which involved flying the aircraft either in a direction specified by instructions, or in such a way as to avoid other traffic aircraft in the nearby airspace. Finally, in Experiment 4, a communications task was added, requiring the pilot to either listen to or look at strings of communications data.

These tasks were performed across various conditions, to be described in more detail below (e.g., while flying straight or maneuvering); conditions that were assumed to change coefficients of the model. During some trials, eye movements were measured on an Applied Life Sciences eye movement recorder, coupled with a head tracker, in order to establish the proportion of time that the visual gaze spent within each of four AOIs: the Outside World (OW), the Instrument Panel (IP), a Cockpit Display of Traffic Information (CDTI) and a source of data-linked communications information (DL). Only the OW and the IP were present across all experiments and experimental conditions. In all four experiments we also measured different aspects of performance, which were used to assess the pilots allocation of resources to tasks (e.g., flight path error, traffic detection, communication readbacks). These data are reported elsewhere (Helleberg & Wickens, in press; Wickens et al., in press).

Figure 3. The simulation environment, showing the CDTI to the left and the instrument panel (IP) to the right.

General Procedures. Each pilot participated for between 5 and 10 hours in one of the experiments. Pilots were instructed to assume that they were in visual meteorological conditions, in which other traffic was potentially visible, and it was thereby important for them to call out "traffic in sight" once they had spotted any traffic aircraft in the outside world. In most conditions they were also to assume direct responsibility for maneuvering the aircraft, in any manner they chose, in order to avoid creating a conflict or "loss of separation" with any traffic aircraft. This conflict event was normally defined to occur when a traffic aircraft flew both within 15 miles of the pilots' aircraft and within 1000 vertical feet of altitude. Except when maneuvering, pilots were instructed to come to, or stay on, the flight parameters of heading altitude and airspeed that were instructed at the beginning of each leg. Following each maneuver, pilots were instructed to return to the original target parameters. Each flight leg lasted between four and seven minutes, and a given flight scenario consisted of 10 or 11 consecutive legs.

Analyses. In all experiments reported below, we analyze the mean scanning data (percentage dwell time) across all pilots in the experiment, as a function of condition and AOI, and correlate these data with the model predictions for the AOI/condition in question. We do not here report individual pilot data (but see Wickens, Helleberg, Kroft, Talleur, & Xu, 2001). In the following, we described the specific characteristics of each experiment, and its model fitting in turn.

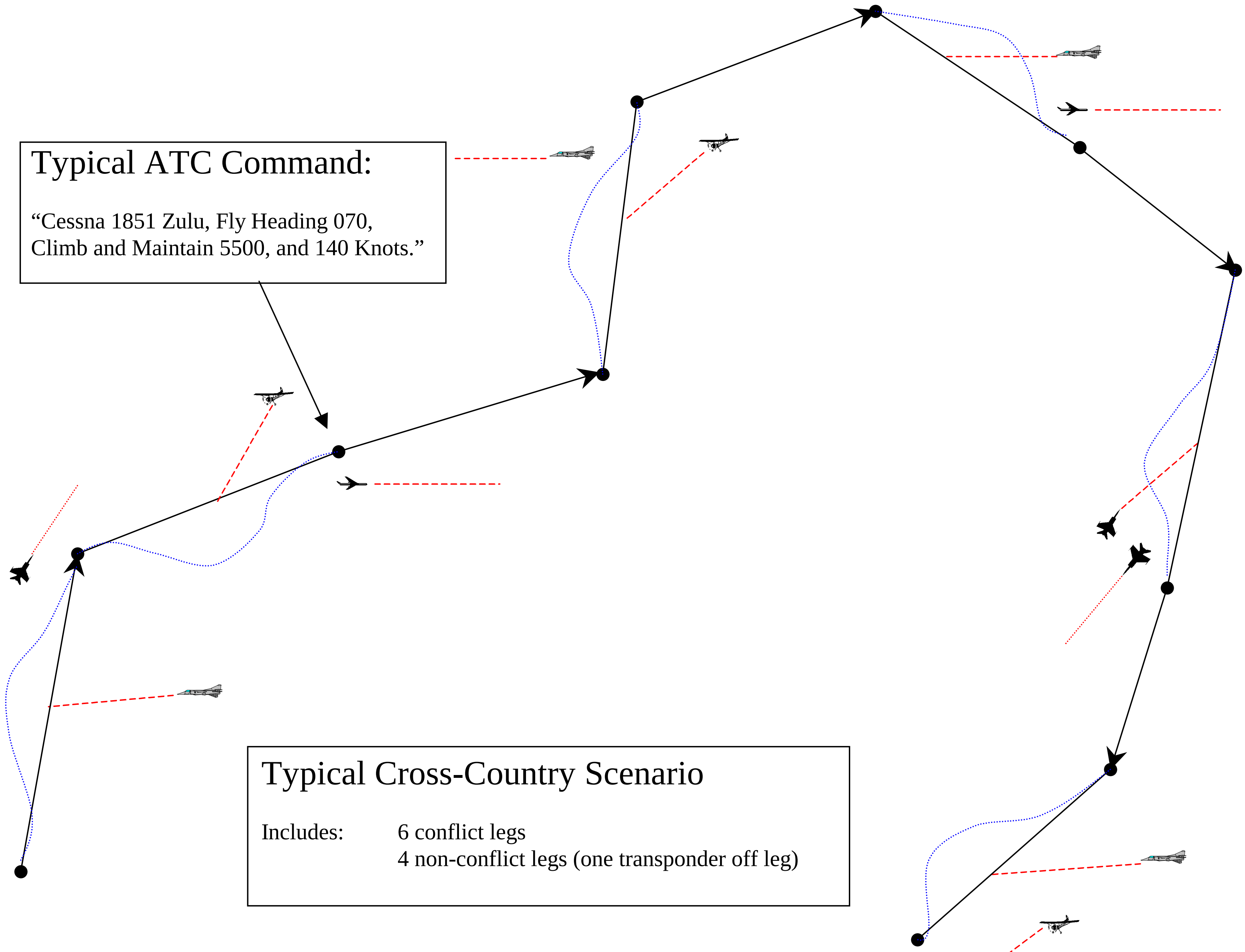

Typical ATC Command:
“Cessna 1851 Zulu, Fly Heading 070,
Climb and Maintain 5500, and 140 Knots.”
Typical Cross-Country Scenario
Includes:
6 conflict legs
4 non-conflict legs (one transponder off leg)

Figure 4.

Experiments 1 and 2. Free flight/baseline. In these two experiments, a total of 17 pilots had their eye tracking recorded as they engaged in a series of flight simulations designed to examine the impact of free flight responsibilities on visual scanning. Seven pilots provided scanning data in the free flight/CDTI experiment, in which they were required to maneuver in order to avoid a loss of separation with a "conflict aircraft". The CDTI was a coplanar display presenting both a map view and rear-view vertical situation display. Predictor lines were presented on both ownship and traffic aircraft (see Wickens, Gempler, & Morphew, 2000 for details). Conflict aircraft were present on 60% of the flight legs (defined by a given heading, altitude and airspeed requirement). The remaining 40% (non-conflict legs) contained traffic aircraft, ultimately visible in the outside world and the CDTI, but these aircraft were not on a conflict path.

In order to provide a non-free flight baseline, against which to compare the scanning behavior of the seven free flight pilots, a second group of 10 pilots (for which good scanning measures were available) flew a simulation experiment (2: baseline) after the free flight experiment had been completed. In Experiment 2, these pilots flew the same set of conflict avoidance profiles (as did the free flight pilots), but these maneuver profiles were now instructed by a simulated air traffic controller. As a consequence, the nature of the flight profile (and hence, of the information seen in the instrument panel and outside world) was essentially equivalent between the two experiments. The primary difference between the experiments was the presence of the CDTI, and the acceptance of **sole responsibility for traffic avoidance navigation** by pilots in the free flight experiment. For pilots in the baseline experiment, this responsibility was shared with ATC.

For the purposes of model fitting then, there were four trial types, defined by crossing experiment (baseline vs. free flight) with leg (conflict maneuver vs. nonconflict straight and level). When these four trial types are crossed with the three (free flight) or two (baseline) AOIs, a total of 10 model predictions can be made, as represented by the two matrices in the top of Table 1. Parameter values for the bandwidth matrix in Table 1 were estimated as follows: The highest value was assumed for the instrument panel, as it contains six instruments, and one of these, the attitude indicator or artificial horizon, is that which represents the highest bandwidth (state changing) aspect of aircraft dynamics (Harris & Christhilf, 1980; Bellenkes, Wickens, & Kramer, 1997; Wickens, in press). The next lower ordinal value was assumed for the outside world, because it too contains the high bandwidth horizon (the true horizon), but does not contain the additional four dynamic instruments contained by the instrument panel. The lowest value was assumed for the CDTI, because this instrument does **not** contain the high BW artificial horizon, although it does represent lower bandwidth aspects of heading, altitude and vertical speed. This ordinal relation of bandwidth across AOI was preserved for both conflict and non-conflict legs. However on the latter, the coefficients were halved, because the absence of lateral or vertical maneuvering on these legs is assumed to reduce the frequency of changes in the instruments. Because of the yoking of behavior of pilots in the baseline condition to that of pilots in the free flight condition, the identical set of inputs was assumed to be visible for both sets of pilots. Hence the bandwidth coefficient values of the IP and OW in the baseline condition match those in free flight (when the CDTI was removed).

Table 1. Parameter values assumed for Experiment 1 (free flight) and Experiment 2 (baseline).

		AOI		
Parameter		IP	OW	CDTI
Bandwidth (B)	Freeflight (Conf)	3	2	1
	Freeflight (Nconf)	2	1	0.5
	Baseline (Conf)	3	2	
	Baseline (Nconf)	2	1	
Relevance (R)	Aviate (FF)	2	1	0
	Navigate (FF)	1	2	2
	Aviate (Base)	2	1	
	Navigate (Base)	1	2	

Priority (V)	Aviate (FF)	3
	Navigate (FF)	2
	Aviate (Base)	3
	Navigate (Base)	1

The relevance coefficients in the second matrix of Table 1 were determined initially by the logic shown in Figure 1. Thus it was assumed that the relevance of the instrument panel for aviating was greater than that of the outside world. This assumption was made since the instrument panel contained four channels of aviating information (pitch, bank, airspeed and vertical speed), always visible, whereas the OW contained only 2 (pitch and bank of the true horizon), which could sometimes be obscured. The CDTI contained no direct representation of attitude or airspeed. For the navigation task, we assumed that the instrument panel took on less relevance since it had no representation of traffic, the primary issue in navigation. Relevance was not however set to 0, since the IP did contain altitude and heading information necessary for complying with navigational maneuvering plans. For navigating (traffic avoidance), the relevance of both the OW and the CDTI was set to be greater than their value for aviating, as well as greater than the relevance value of the IP.

Finally, the priority coefficients were assigned on the basis of the “aviate-navigate” task hierarchy underlying aviation as described above (Schutte & Trujillo, 1996). However the importance or priority of navigating was assigned a higher value in free flight than in the baseline, since in the free flight condition the pilot was the only person responsible for this task, whereas in the baseline condition navigation (traffic avoidance) was primarily the responsibility of ATC. That is, the simulated controller provided instructions of heading altitude and airspeed that were guaranteed to avoid the loss of separation.

In analyzing the visual scan data, it should be noted that the differences in percentage dwell time in the three AOI’s, across the various conditions were statistically significant (Wickens et al., in press). Hence we have confidence in the generalizability of our findings. Figure 5 shows the data for the scan measures (percentage of dwell time within each AOI) regressed against the model predictions for the 10 different conditions. The product moment correlation between model predictions and scanning parameters was $r = 0.89$, indicating that 79% of the variance of scanning was accounted for by model predictions. Examination of Figure 5 suggests that the major source of variance accounted for by the model is the difference between the three AOI’s, shown by the three clusters of points. However, even within the clusters, there is a tendency for the points to align themselves with the regression line, particularly within the IP (the cluster of four points on the upper right). Further examination of the regression plot reveals that pilots scanned the OW less, in the free flight condition, then the model predicts. Nevertheless, the fit of the model by the data is quite satisfactory. There are of course justifiable ways that the model coefficients could be changed which could improve the fit still further. As one example, it could be assumed that the CDTI does have **some** relevance to aviating, given that the vertical profile predictor display on the CDTI shows a very salient vertical trend indicator. Replacing the 0 in this cell of Table 1 with a 1 actually increases the model fit to r = 0.90.

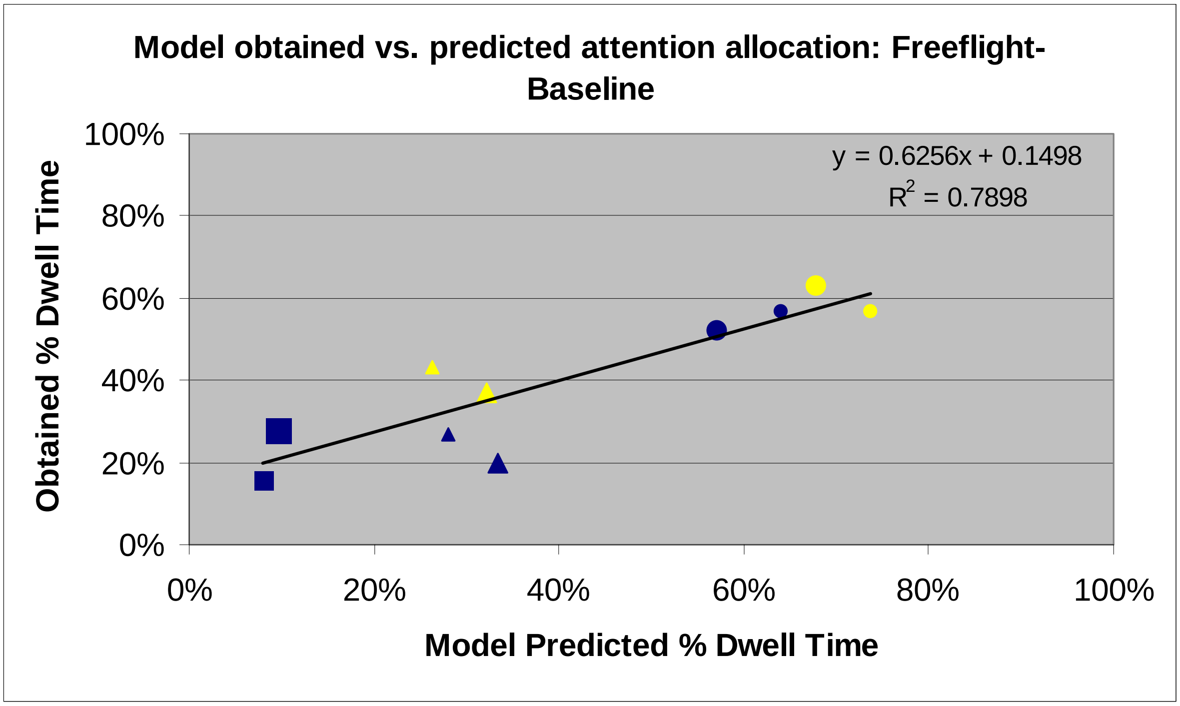


Figure 5. Model prediction versus scan (percentage dwell time) performance for Experiments 1 and 2. CDTI: squares; OW: triangles; IP: circles; Conflict: small; Nonconflict: large; Free flight: dark shading; Baseline: lighter shading.

Experiment 3: Modality and traffic density. Rather than pursuing parameter adjustment of Experiment 1 data to optimize the current model fit, we chose instead to apply the parameters of the model used in Experiment 1, to a new set of data, collected in a different experiment, with some different manipulations. In Experiment 3, a different set of pilots (N=12) engaged in the same general flight simulation (using the same equipment) as in Experiments 1 and 2. However, for the purposes of model prediction, the following important changes were made. (1) All trials were "free flight" in that pilots were required to judge traffic conflicts on their own, and to maneuver if necessary to avoid traffic. (2) Ninety percent of the trials were non-conflict (straight and level), and only those non-conflict trial data will be reported (although maneuvers were not required on the trials, this discrimination was challenging for the pilots). (3) On 2/3 of the trials, pilots received information about traffic, as in Experiment 1, on a CDTI. On the remaining third of the trials, pilots received the same essential information, delivered auditorally (e.g., "traffic

2:00 low, 4 miles”). The CDTI was not present as an AOI on these trials. (On half of the CDTI trials, pilots received traffic information redundantly on the auditory channel, a distinction that will not be relevant to the current modeling effort). (4) On half of the flight legs, there was only one traffic aircraft, whereas on the other legs, four aircraft were encountered, hence greatly elevating the bandwidth of any AOI that represented traffic.

Table 2 presents the coefficients used for the model in Experiment 3. In essence, we have borrowed the coefficients from the non-conflict free flight trials from Experiment 1 (see Table 1). These values are represented across the top rows of the three panels within Table 2. In the second row of the bandwidth panel, given the increase in traffic, we have increased the bandwidth parameters for the two AOIs that present traffic information, the OW and the CDTI. We provide a greater increase (x4) for the CDTI than for the OW (x2) because all four traffic aircraft are typically visible on the CDTI, whereas the narrower scope of the OW view, presents roughly two traffic aircraft at any given time. These values are replicated on the two rows below for the two relevant columns, as was done in Experiments 1 and 2 (e.g., without the CDTI).

Table 2. Parameter values for Experiment 3: traffic density and modality.

Parameter		AOI: IP	AOI: OW	AOI: CDTI
Bandwidth (B)	Visual (1)	2	1	0.5
	Visual (4)	2	2	2
	Auditory (1)	2	1	
	Auditory (4)	2	2	
Relevance (R)	Aviate (V)	3	1	0
	Navigate (V)	1	2	2
	Aviate (A)	3	1	
	Navigate (A)	2	3	

Priority (V)		
	Aviate	3
	Navigate	2

The relevance parameters in Experiment 3 are the same as they were for the corresponding condition from Experiment 2 with the following important exception. Unlike Experiments 1 and 2, where the relevance values were the same across the two experimental conditions (baseline and free flight), here we increase the relevance of the OW for navigation in the auditory, relative to the visual condition (increase from 2 to 3). The rationale for this increase is that, in the auditory condition, the OW is the **only** source of information available for traffic detection and maneuvering. Hence, it must be viewed as more relevant than either the CDTI or the OW in the visual conditions (both relevance = 2), since in these visual conditions the two AOIs share a good bit of redundant information regarding where the traffic is located.

Priority parameters in this free flight condition were set to their respective values for aviating and navigating as they were in the free flight conditions (Experiment 1). Using these parameter values, the model predictions, plotted against the scanning data, are depicted in Figure

6. The model predictions are here extremely well fit by the obtained data, with r = 0.974, or 95% of the variance accounted for. Indeed there seems to be few ways in which the model fit could be improved.

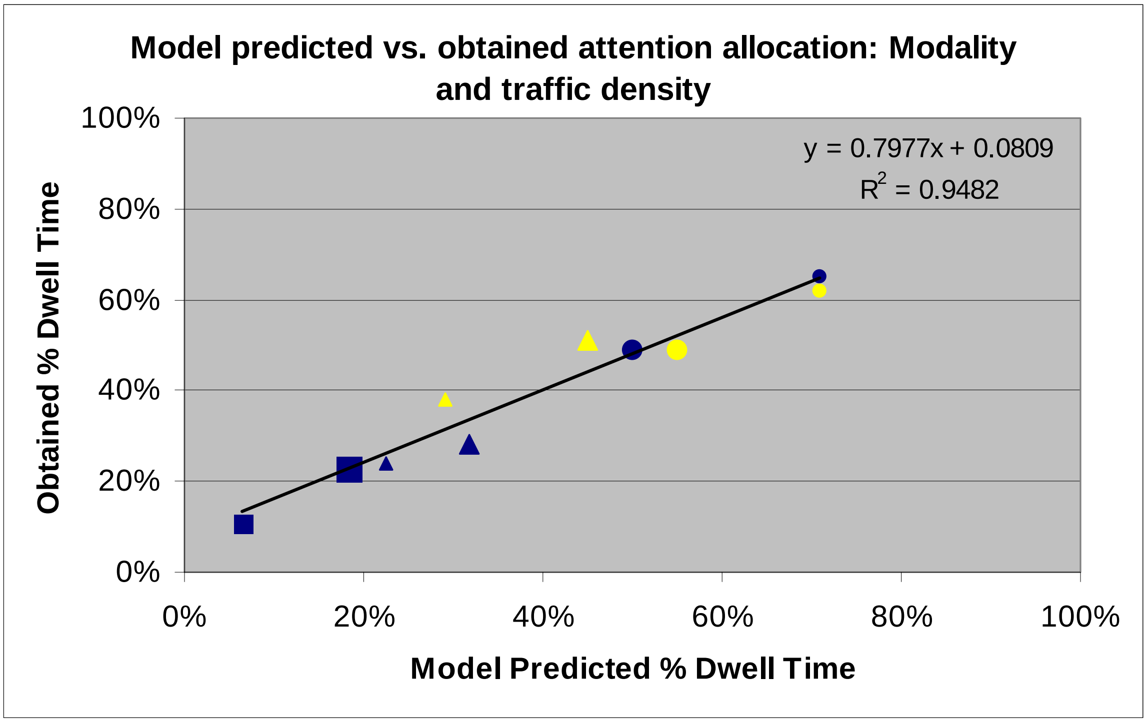


Figure 6. Model fit of Experiment 3. Squares=CDTI. Triangles=OW. Circles=IP. Small=1 traffic. Large=4 traffic. Dark=Visual CDTI. Light=Auditory.

Experiment 4: Communications. In this experiment, whose details are reported in Helleberg and Wickens (in press), the same general procedures were followed. Again, pilots (N = 17) participated in the same general scenario, although under baseline (ATC responsibility) rather than free flight conditions. As in Experiment 3, most trials did not involve maneuvering to avoid traffic, so our modeling is of the straight and level flights. In this experiment, as in Experiments 1 and 2, there was only one traffic aircraft. The primary difference between this experiment and the other three was the comparison between a visual data link display to present communications information regarding flight parameters, radio settings, etc., and a configuration in which the same communications information was presented auditorally. This change imposed two additional changes to the modeling effort. First, we now added a third task, communications, which logically falls at the bottom of the priority hierarchy. Thus now there are three tasks with priorities assigned to aviate (=3), navigate (=2) and communicate (=1) (Schutte & Trujillo, 1996). Second, the data link task adds two AOIs. When communication information was presented visually, the AOI is a data link display, located in the same position as the CDTI display in Experiments 2 and 3 (see Figure 2). When the datalink communications information was presented auditorally, pilots were forced to write down the communications messages, so as not to forget them (some of these messages were quite lengthy). This requirement defined an

AOI of the clipboard, located on the pilot's lap just below the yoke. Between legs, we varied the length of the ATC communications from two to six "chunks" of relevant information (i.e., heading, altitude, airspeed, communications frequency/tower name, transponder code, altimeter setting); hence this varied the effective bandwidth of information "delivered along" (or inherent in) the communications AOI (datalink display or clipboard).

The matrix for Experiment 4 is shown in Table 3. The format corresponds relatively well with that in Tables 1 and 2, with the following exceptions. First, the bandwidth of the added communications AOI is set to ordinal values, based upon message length. These are somewhat arbitrarily set to increment in values of 0.5. Second, the highest value of the communications channel (2.5) is set below the bandwidth value of the OW (=3). This ordering is justified because even at the highest communications load, the total number of "chunks" of communications information (6) is considerably less than the number of events – state changes in pitch and roll of the far horizon – that characterize the OW. (Note that this forces us to increase the OW bandwidth value from 1.0, its value in corresponding conditions of Experiments 1 and 2, to 3.0.) As a consequence we must increase the IP bandwidth to 4, in order to preserve the ordinal relationship across AOIs). Third, we define different AOIs for the visual (data link) condition than for the auditory (clipboard) condition, since these are located in spatially different regions. However the parameter sets for these two AOI locations do not differ. Fourth, given the added task (communications) we now add this task to the priority matrix. Fifth, we assign these communications channels some (but minimal) relevance to both aviating and navigating, given that they deliver target values that influence the flight controls of pitch and bank (aviating) as well as the navigational parameters (navigating).

Table 3. Parameter values for Experiment 4: communications experiment.

Parameter		AOI: IP	AOI: OW	AOI: COM DL	AOI: COM Clip
Bandwidth (B) Visual/Red	Com Load 2	4	2	0.5	0
	Com Load 3	4	2	1	0
	Com Load 4	4	2	1.5	0
	Com Load 5	4	2	2	0
	Com Load 6	4	2	2.5	0
Bandwidth (B) Auditory	Com Load 2	4	2	0	0.5
	Com Load 3	4	2	0	1
	Com Load 4	4	2	0	1.5
	Com Load 5	4	2	0	2
	Com Load 6	4	2	0	2.5
Relevance (R)	Aviate	3	1	1	1
	Navigate	1	2	1	1
	Communicate	0	0	2	2

Priority (V)		
	Aviate	3
	Navigate	2
	Communicate	1

Figure 7 presents the results of the model fitting exercise. As with Experiments 1 and 2, statistical analyses reveal that the conditions differ from each other (Helleberg & Wickens, in press), so that the variance in data points that are modeled is “meaningful”. The regression analysis of obtained on predicted scanning percentages revealed a very high degree of fit, with r = 0.97, or 95% of the variance accounted for. While the figure reveals that a large proportion of the shared variance can be accounted for by the scanning differences between the three AOIs (and particular, between the IP in the upper right, and the other two AOIs to the lower left), it is also noteworthy that in each of the three AOI clusters, the variance in data points lies along, or relatively parallel to the regression line, indicating that variance in task characteristics, captured by the model, is also reflected in the data.

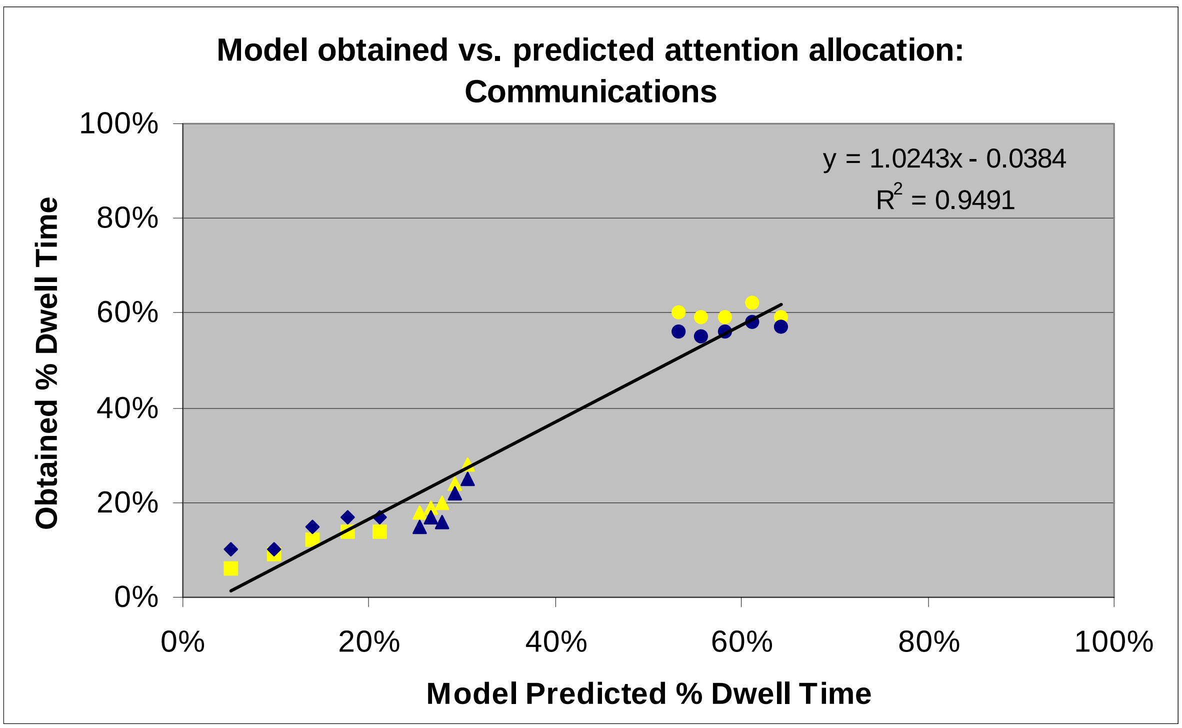


Figure 7. Model fit of Experiment 4. Dark=auditory. Light=visual. The six communicative load points are not separately labeled, but fall in monotonic order.

DISCUSSION

We have tested an expected value model of information acquisition (assessed operationally by visual scanning), against the data generated by well trained pilots, in three operationally meaningful simulations carried out on a high fidelity flight simulator. The model is relatively simple, involving only two parameters, of expectancy and value, and a fairly simple heuristic for assigning those parameter values to particular conditions, according to the lowest set of values that preserve ordinal relations across tasks, conditions, and AOIs.

In spite of this simplicity, which we believe reflects the valued attribute of model parsimony, the model has done a reasonably good job in predicting between-condition variance in visual scanning. In the initial model development experiment, nearly 80% of the variance was accounted for. More impressively, in the two cross validation experiments, the model fits increased, accounting for 95% of the variance in both cases. It is noteworthy that this increase in variance accounted for with cross validation, is in contrast to the decrease that is typically found as regression models are cross validated. The difference here of course is that we did not use the least-squared-deviation regression line itself to define the weights of parameters of our model (thereby increasing the risk of capitalizing on chance), as is typically done in regression model fitting. Instead, our model weights were defined a-priori, on the basis of what we believe are cognitively plausible (and therefore defendable) hypotheses regarding pilot goals (priority and relevance) and environmental characteristics (bandwidth).

It is appropriate to ask how the architecture of the current model would fare with simpler assumptions. There are several ways of testing this. One way is to replace the coefficients in the three matrices with the same values that are now reassigned randomly to the cells. We tried such an approach, and in all three experiments, the correlation between prediction and data was near 0. A second way would be to systematically “freeze” the parameters of each term in turn to a constant value, in order to find out how much (or little) the predictions suffer, without the contributions of that particular component.

The results of our modeling effort suggests that well trained pilots are indeed quite optimal in allocating attention, a conclusion agreeing with that offered by Moray (1986) in his comparison of model fitting results applied to novice subjects (substantial departures from optimality) and more trained experts (fewer departures). As a consequence, we feel confident that the model can serve as somewhat of a “gold standard” for training attention allocation skills in different complex environments; or for remedial training of particular pilots, who show severe departures from the optimal prescriptions. We believe that our data also have at least three other implications or ramifications.

First, both the optimal prescriptions **and** our observed performance revealed that pilots scanned the instrument panel much more than the outside world, at a ratio that approximated 2:1 across experiments. This ratio stands in marked contrast to a ratio of 1:3 which is typically argued to be the appropriate scanning ratio for pilots in visual conditions, according to the Airman’s Information Manual (AOPA Air Safety Foundation, 1993). The empirical basis for developing this 1:3 standard is not entirely clear, but it is noteworthy that it would appear to contradict, to some degree, the aviate-navigate-communicate hierarchy that underlies the current model.

The second important aspect of our data relates to the first, and can be stated in terms of a question: does this lower than recommended scan of the outside world (both predicted and obtained) place the pilot “at risk” for detecting traffic? In all four experiments we did record the latency and accuracy to call out “traffic in sight”, and an analysis of individual differences **between pilots** indeed revealed that those who scanned the OW relatively more, did detect the traffic relatively sooner (Wickens, 2001). However this conclusion does not necessarily mean that those **conditions** that induced more OW scanning necessarily supported better detection, and indeed consideration of our experimental results (Wickens et al., in press) revealed only a weak between-condition relationship between OW scanning and detection performance. In particular, during a small number of trials in Experiment 1 (free flight), we would present an aircraft in the OW which was **not present** on the CDTI, to assess whether the reduced OW scanning resulting as attention was allocated to the CDTI, would harm this traffic detection. Such harm was observed slightly, but only on the first encounter with such an unannounced traffic aircraft (there were four to five additional encounters during the experiment), and then only when the aircraft was itself of poor visibility/conspicuity. Thus, it does not appear that this lower value of OW scanning places the pilot much at risk.

The third issue is in turn related to the second, and this concerns the potential role of conspicuity or salience, as well as effort in directing scanning; that is, the two components of the descriptive SEEV model that were not a part of the optimal EV model. With regard to Salience – the attention-capturing properties of an event – is would appear that the statistical averaging technique we used here (average PDT across a trial) would not be the best way of capturing a salience effect. The role of salience, defined as a property of an **event**, rather than as a channel or AOI, would be to shorten the latency of attention allocation following an event, rather than to alter the overall distribution of attention to that AOI relative to others. Stated in more intuitive terms, you do not necessarily look more frequently at channels where salient events occur; and indeed, it might be optimal to look at these channels less frequently, since you know that your attention is more likely to be captured by those more salient events in your peripheral vision.

In contrast to salience, it is likely that effort did play some role in influencing scanning, although this role was not extensive (given the high variance accounted for, by the prescriptive model that excluded a salience influence). The inhibitory role of effort (inhibiting longer distance scans) was suggested indirectly by the analysis of dwell duration in Experiments 1 and 2. In both experiments we observed relatively long dwells on the instrument panel; periods averaging as long as 6 seconds (baseline) or 3.5 seconds (free flight), during which the eye stayed on the instrument panel, scanning different instruments, but not venturing to examine the outside world or (in Experiment 1), the CDTI. We label this the “in the neighborhood” heuristic of scanning, whereby the operator chooses to make several short saccades to coupled AOIs (in the neighborhood), rather than making repeated long distance excursions out (to the OW) and back (to the IP). Such a choice thereby reduces the effort of information access. An important observation is that these long ‘in the neighborhood” dwells were reduced considerably in their duration in free flight, compared to the baseline condition, suggesting that the added responsibilities of traffic monitoring availed in freeflight, by increasing the importance of the outside world, thereby suppressed the inhibitory role of effort. Stated in terms of the coefficients of equation (1), the weighting of **v** is considerably greater than the weighting of **ef**. This then is another statement of support for the optimality of pilot scan.

It is also likely that effort played some role in modulating the scan patterns of Experiment 4, given that the communications AOI on the clipboard was located farther from the OW and the IP, than was the communications AOI on the datalink display. Accordingly, following the same application of the “in the neighborhood” heuristic described in the context of Experiments 1 and 2, we would predict longer dwells on the clipboard AOI than on the Datalink display, a prediction consistent with the data which revealed a mean dwell duration of 2 and 1 seconds respectively (Helleberg & Wickens, in press). (These data are consistent with the role of effort, but certainly do not confirm its role, since other features of the two AOIs differed besides their distance from the IP and OW.)

As the two foregoing examples suggest, while it is possible to identify the role of effort in inhibiting attention allocation, it is probably more challenging to ascertain a physical metric of effort that would predict differences in scanning. A simple metric would be to assume that effort is a linear function of the visual angle of separation, but such an assumption does not account for differences between eye movements and head movements. The role of effort, and its reflection in dwell durations as well as percent dwell time on AOIs will require considerable added complexity of the model. However the ability to incorporate an effort component will provide a valuable means of predicting more or less optimal display layouts, when "optimal" is defined in terms of the location of instruments (Wickens, Vincow, Schopper, & Lincoln, 1997)

Conclusions

In conclusion, we have shown how a relatively simple model of the top down, knowledge-based forces that should drive attention accounts for a high degree of variance in pilot scanning, across a range of conditions and cockpit configurations. Naturally there is more to visual attention than foveal vision (inferred by scanning), and more to attention than visual attention, so that the model has limitations as a general attention model. Nevertheless, we believe that it can serve as a useful component of more general models of human performance, and serves to supply an important link between the two research domains of supervisory sampling and task management as well as providing or valuable reminder of the extent to which well trained operators are optional.

ACKNOWLEDGEMENTS

Support provided by a grant from NASA Ames #NASA NAG 2-1120. Dave Foyle was the scientific technical monitor. Data generated for Experiments 1, 2, and 4 was supported by a grant from Federal Aviation Administration #DTFA 98-G-022. Dennis Beringer was the scientific technical monitor. Authors wish to acknowledge invaluable contributions of Roger Marsh and Ron Carbonari for simulation development, and of Don Talleur for professional advice. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of NASA or the FAA.

REFERENCES

AOPA Air Safety Foundation (1993). How to avoid a midair collision. January 16.

Bellenkes, A. H., Wickens, C. D., & Kramer, A. F. (1997). Visual scanning and pilot expertise: The role of attentional flexibility and mental model development. Aviation, Space, and Environmental Medicine, 68(7), 569-579.

Brogan, D., Gale, A., & Carr, K. (1993). Visual search 2. Bristol, PA: Taylor & Francis.

Bundensen, C. & Pedersen, L. F. (1983). Color segregation and visual search. Perception and Psychophysics, 33(5), 487-493.

Carbonell, J. R. (1966). A queueing model for many-instrument visual sampling. IEEE Transactions on Human Factors in Electronics, HFE-7, 157-164.

Carbonnell, J. F., Ward, J. L., & Senders, J. W. (1968). A queuing model of visual sampling: Experimental validation. IEEE Transactions on Man Machine Systems, MMS-9, 82-87.

Chou, C., Madhavan, D., & Funk, K. (1996). Studies of cockpit task management errors. The International Journal of Aviation Psychology, 6(4), 307-320.

Dismukes, K. (2001). The challenge of managing interruptions, distractions, and deferred tasks. Proceedings of the 11th International Symposium on Aviation Psychology. Columbus, OH: The Ohio State University.

Drury, C. (1975). Inspection of sheet metal: Model and data. Human Factors, 17, 257-265.

Edwards, W. (1961). Behavioral decision theory. Annual Review of Psychology, 12, 473-489.

Ellis, S. R., & Stark, L. (1986). Statistical dependency in visual scanning. Human Factors, 28(4), 421-438.

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

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. Journal of Experimental Psychology: Human Perception and Performance, 18, 1030-1044.

Gillan, D. J., & Lewis, R. (1994). A componential model of human interaction with graphs: I. Linear regression modeling. Human Factors, 36, 419-440.

Harris, R. L., & Christhilf, D. M. (1980). What do pilots see in displays? In G. Corrick, E. Hazeltine, & R. Durst (Eds.), Proceedings of the 24th Annual Meeting of the Human Factors Society. Santa Monica, CA: Human Factors Society.

Helleberg, J., & Wickens, C. D. (2001, in press). Effects of data link modality and display redundancy on pilot performance: An attentional perspective. International Journal of Aviation Psychology.

Kvalseth, T. (1977). Human information processing in visual sampling. Ergonomics, 21, 439-454.

Liu, Y. & Wickens, C. D. (1992). Visual scanning with or without spatial uncertainty and divided and selective attention. Acta Psychologica, 79, 131-153.

Lohse, G. L. (1993). A cognitive model for understanding graphical perception. Human-Computer Interaction, 8, 353-388.

Moray, N. (1986). Monitoring behavior and supervisory control. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), Handbook of perception and performance, Vol II (pp. 40-1-40-51). New York: Wiley & Sons.

Moray, N., Richards, M., & Low, J. (1980). The behaviour of fighter controllers (Technical Report). London: Ministry of Defense.

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

Pashler, H., Johnston, J. C., & Ruthruff, E. (2001). Attention and performance. Annual Review of Psychology, 52, 629-651.

Raby, M., & Wickens, C.D. (1994). Strategic workload management and decision biases in aviation. International Journal of Aviation Psychology, 4(3), 211-240.

Schutte, P. C., & Trujillo, A. C. (1996). Flight crew task management in non-normal situations. Proceedings of the 40th Annual Meeting of the Human Factors and Ergonomics Society (pp. 244-248). Santa Monica, CA: Human Factors and Ergonomics Society.

Senders, J. (1964). The human operator as a monitor and controller of multidegree of freedom systems. IEEE Transactions on Human Factors in Electronics, HFE-5, 2-6.

Senders, J. W. (1983). Visual scanning processes. Netherlands: University of Tilburg Press.

Sheridan, T. (1970). On how often the supervisor should sample. IEEE Transactions on Systems Science and Cybernetics, SSC-6(2), 140-145.

Sheridan, T. B., & Rouse, W. B. (1971). Supervisory sampling and control: Sources of suboptimality in a prediction task. Seventh NASA Annual Conference on Manual Control. University of Southern California.

Smallwood, R. (1967). Internal models and the human instrument monitor. IEEE Transactions on Human Factors in Electronics, HFE-8, 181-187.

Stanton, N. (Ed.). (1994). Human factors in alarm design. Bristol, PA: Taylor & Francis.

Teichner, W. H. & Mocharnuk, J. B. (1979). Visual search for complex targets. Human Factors, 21(3), 259-275.

Theeuwes, J. (1994). Visual attention and driving behavior. In C. Santos (Ed.), Human factors in road traffic. Lisbon Portugal: Escher.

Tulga, M. K., & Sheridan, T. B. (1980). Dynamic decisions and workload in multitask supervisory control. IEEE Transactions on Systems, Man and Cybernetics, SMC-10, 217-232.

Wickens, C. D. (2001). A model of attention allocation for aviation. Association for Aviation Psychology Newsletter, 23(1), 4-9.

Wickens, C. D. (2002, in press). Aviation displays. In P. Tsang & M. Vidulich (Eds.), Principles and practice of aviation psychology. Mahwah, NJ: Lawrence Erlbaum Publishers.

Wickens, C. D., Gempler, K., & Morphew, M. E. (2000). Workload and reliability of predictor displays in aircraft traffic avoidance. Transportation Human Factors Journal, 2(2), 99-126.

Wickens, C. D., Helleberg, J., Kroft, P., Talleur, D. A., & Xu, X. (2001). Mid air target detection: What makes it difficult? Application of attention and situation awareness model (Technical Report ARL-01-9/NASA-01-5). Savoy, IL: University of Illinois, Aviation Research Lab.

Wickens, C. D., Helleberg, J., & Xu, X. (in press). Pilot maneuver choice and workload in free flight. Human Factors.

Wickens, C.D., Kroft, P., & Yeh, M. (2000). Data base overlay in electronic map design: Testing a computational model. Proceedings of the IEA 2000/HFES 2000 Congress (pp. 3-451-3-454). Santa Monica, CA: Human Factors & Ergonomics Society.

Wickens, C. D., Vincow, M. A., Schopper, A. W., & Lincoln, J. E. (1997). Computational models of human performance in the design and layout of controls and displays (CSERIAC SOAR Report 97-22). Wright Patterson AFB, OH: Crew System Ergonomics Information Analysis Center.

Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. Psychonomic Bulletin and Review, 1, 202-238.

Yantis, S. (1993). Stimulus-driven attentional capture. Current Directions in Psychological Science, 2, 156-161.