Designing To Support Awareness: A Predictive, Composite Model

Rachid Hourizi
University of Bath
Claverton Down, Bath, UK
r.hourizi@bath.ac.uk

Peter Johnson
University of Bath
Claverton Down, Bath, UK
p.johnson@bath.ac.uk

ABSTRACT

In this paper we propose an account of human/computer awareness for use in the (re)design of complex human/computer interaction, before empirically testing its utility. Specifically, having situated our work in the wider field of human/computer awareness research, we address the well-reported phenomenon of "situation awareness" breakdowns in the aviation domain. We assert the need for an explanatory and predictive model of the phenomenon if the frequency of such breakdowns is to be reduced and propose such a model. We then go on to investigate the utility of our model as a guide for design through the discussion of a recent experiment involving manipulations of an animated warning signal on a simulated cockpit control panel. Our results show initial support both for the model and for our assertion of its utility. We conclude that our composite view of awareness yields practical benefit in the design of human computer awareness support.

Author Keywords

Awareness, Predictive Model, Interaction Design,

ACM Classification Keywords

H.5.2 User Interfaces

INTRODUCTION

As computer systems have become embedded in complex, dynamic environments with increasing numbers of users, objectives and potential information sources, human computer interaction (HCI) designers have been forced to confront a number of challenges, whose effects were, in many cases more muted in the more restricted domain of desktop computing. Prominent amongst these challenges is that of human computer awareness (HCA) – a growing problem in a number of complex domains, where groups of people are often involved with multiple automated systems in the pursuit of multiple objectives (e.g. medical practice [1], naval operations [2] and office work [3],[4]).This breadth of domain and context in which awareness failures occur is reflected in a growing number of awareness related threads in a widespread research community (e.g.. the study of *peripheral* [5], [10] and *context* [3] awareness.

In this paper, we will focus on a specific sub-category of HCA - that required in the highly specialised domain of the commercial aircraft cockpit (often referred to as *situation* awareness). In this domain, as those described above, an awareness of computerised system state and activity appears particularly problematic. Numerous reports and studies spell out the ongoing concern of active pilots that they are frequently unaware of current and future system activity or the logic, which drives it [e.g. 8] and a large number of accident reports document the potential for serious consequences when pilots fail to understand these activities. (e.g. [6],[7]).

In sum, therefore, the phenomenon of system awareness in the cockpit exists within a wider (and growing) context of awareness *of*, *through* and perhaps even *by* automation. Across this wider context, a series of higher-level questions crop up time and again: What is awareness? Why does it break down? How can we design interaction such that the most severe forms of breakdown become less likely? This paper will address domain and context specific awareness failures occurring between pilots and autopilot in common commercial aircraft. Behind these specific questions and investigations, however, we will both draw from and, wherever possible, feed back to the wider community interested in the design of computer systems, which support the acquisition, maintenance and repair of user awareness in its various forms, and in the fast moving multi-agent, multi-task, environments mentioned above.

CURRENT LITERATURE AND PROPOSED EXTENSION

For many of the relatively new, strands of research described above, the current state of their art consists of detailed descriptions of both successful and failed interaction between human and computerized tool [e.g. 3]. Beyond these observational studies, a second body of research findings are provided by experimental prototypes [e.g. 11], designed with one eye on execution of a particular task or activity and a second on broadening the community's understanding of awareness and the nature of the support required if it is to be maintained.

Our own chosen sub-field of (aviation) situation awareness, however, benefits both from a long history of reporting accidents and, somewhat as a consequence, from a number of descriptive frameworks, within which the problem of HCA can be discussed. These frameworks are often quite detailed in their description both of the information, which must form part of the ultimate *state* of awareness, and of the *processes* by which it can be achieved. Sarter & Woods [12] for

CHI 2004, April 24–29, 2004, Vienna, Austria.

example, draw upon their own observational studies to assert the importance of notions including perception, attention and knowledge to the acquisition, maintenance and repair of awareness. Endsley [13] adds anticipation to the mix, citing the pilots' need to be "ahead of the plane" in the complex, rapidly changing environment of the cockpit. Meanwhile, Rushby et al [14] have even generated a sufficiently explanatory model of (cockpit software) "mode" awareness that they have, in some circumstances, been able to predict potential awareness failures in systems where key information is not immediately available to pilots.

This formal approach is not, however, (and never claimed to be) a universal solution to the wide range of situation awareness failures reported. We have demonstrated in previous work [9] for example, that the availability of pertinent information in an environment is not, in itself, sufficient to guarantee that people will become aware of it. Work remains then in the building of explanatory and, importantly, predictive models of awareness in the areas not caused by a lack of information i.e. those situations where information is available but overlooked.

With this in mind, we propose such a model for HCA supporting design, drawing on and extending the findings discussed above. Our starting point for this model is the very ambiguity inherent in the term "awareness" itself. In everyday speech, for example, the seemingly straightforward question "Were you aware that the kitchen light was on?" can vary, according to context. In the simplest case, this can mean simply "Did you see the kitchen light?" In a context where both questioner and questionee are looking at a scene in which the light is obviously visible, however, the meaning of the question mutates to become "Did you pick out the presence of the kitchen light from amongst the many elements of the scene that you undoubtedly saw?" – a question which is roughly analogous to the question "Did you attend to the fact that the kitchen light was on?" In yet another context, however, the question could carry implications for the recipient of the question in the form of a required action or response. In this context, then the question becomes equivalent to "Did you understand the implicational meaning of the fact that the light was on?"

Building from the threads of previous research, therefore, we believe a composite model is needed in which the information in the environment must pass through a number of cognitive processes before it can be considered to form part of a person's awareness - a notion also inspired by, if not directly built upon, the work of Barnard and May in their work on Interacting Cognitive Subsystems [16]. We propose, therefore a minimal model of (the process of) awareness, in which raw data from the environment has been (1) available (in line with Rushby's models), (2) perceived, (3) attended to in some manner and (4) subject to further, higher level cognitive processing.

We can immediately use this model to provide insight to the variations in our kitchen light example. The first question "Did you see that the kitchen light was on?" can be thought of as a question about perceptual or level (2) awareness. The second question "Did you pick out the presence of the kitchen light from amongst the many elements of the scene that you undoubtedly saw?" becomes a question of whether level (3) or attentional awareness has been achieved and the last question "Did you understand the implicational meaning of the fact that the light was on?" refers to the higher level cognitive (often semantic) processing inherent in level (4) awareness.

Armed with these distinctions, we are in a position to describe cockpit awareness breakdowns not only in terms of the final awareness desired, but also in terms of the particular cognitive sub-process involved in the failure. We could, for example, imagine separate breakdowns in which information was available but not seen (a level 2 failure), picked up visually but overlooked because the pilots attention was elsewhere (a level 3 failure) or available, seen and attended to but not understood in terms of its meaning or implication (a level 4 failure). In other words, in the place of the single, rather clumsy question "Why was the pilot not aware of that issue?", we are now able to ask *multiple* focused questions each aimed at a separate potential failure point in the acquisition, maintenance and repair of awareness. By extension, we can use similar questions in the early stages of design in order to predict potential problems as Rushby [14] was able to do at the level of information availability – e.g. Is the users' attention likely to be drawn to the relevant information source? Is the information presented in such a way that the receiver can easily process its implicational meaning?

Whilst we could imagine that such an approach, if shown to be practical, could be of benefit across a range of environments and contexts, we must start by establishing a proof of concept both for the validity of the model and for its asserted utility. With this in mind, we will use the rest of this paper to investigate two questions; 1) Can we provide empirical evidence to support our intuitive description of awareness and 2) If so, can we manipulate design elements in an authentic interface to demonstrate its utility?

EXPERIMENT

Our exploration of these interwoven questions involved the investigation of the well-documented interaction between pilots and autopilot in the Airbus A320, a large commercial passenger aircraft. This interaction (described in more detail below) was chosen exactly because it has been implicated in a number of high profile HCA breakdowns, such as the crash of an Air Inter aircraft near Strasbourg, France in 1992. Investigation of this area, therefore, gave us both a context in which to test our model and an authentic arena in which to make a contribution to a real world problem. Specifically, we chose to investigate a descent scenario from a starting altitude of 10,000 feet (a context similar to the one in which the original Strasbourg HCA breakdown occurred). In the course of this scenario, our participants were asked to execute a series of instructions to effect this descent, all the while ensuring that the aircraft was travelling towards an airport, which was below them (!) and slightly to their right. In the course of this scenario, the automation would (unbeknown to the participants) make alterations to the course of the flight, such that the aircraft would start to move *away from* the fictional airport. We will refer to these independent actions as *interventions* on the part of the autopilot.

Next, we added components to the display (described in more detail below) warning the participants that such an intervention was taking place. We also varied the nature of the signal in different conditions in order to examine the different reactions occasioned by each warning type i.e. we used each condition to target a different level of awareness.

Finally, we measured the participants' awareness of the ongoing flight (i.e. "situation awareness"), recording both reported observations that "something unexpected was happening" and subsequent participant activity (if any) to correct the problem. In this way, we started to separate those interventions, which had been seen but not fully understood (indication only that our "perceptual" level of awareness had been achieved) from those, which had been further processed (indication that our "semantically processed" level of awareness had been reached).

This experiment, then, addressed the two high level questions posed in the previous section since 1) it was based upon a manipulation of the levels of awareness defined in our model, 2) it provided evidence of the utility (or otherwise) of a specific, principled design solution.

In order to move towards specific hypotheses, however, we needed to refine the detail of those interface manipulations, which we would chose to perform. Specifically, we needed to provide explicit support for our separate levels of awareness in separate experimental conditions.

We started with the notion that the existing interface provided information showing the autopilots activity in close to real time. The display (shown in figure 1 and described in more detail below) did, after all show the plane turning left and right, ascending and descending. In the context of our experiment, therefore, we could use this extant display as our base (control) condition (C1) and assert that it provided support for pilot awareness only at the first and second levels of our model – our target information was both available and perceivable but no more.

In our second condition (C2), however, we needed to provide awareness support at a higher level (level 3 in our model) i.e. we needed to increase the chances that our participants attention would be drawn to the autopilot activity. With this in mind, we added a warning signal designed to draw attention in this way, but containing very little semantic relevance to the autopilot action being signalled. After rejecting a series of candidate signals, which we feared would be either too strong (and therefore distracting), we decided upon a "throbbing" movement for the central aircraft icon on our simulated display i.e. a warning signal in which a prominent icon grew larger and then shrank repeatedly over a short period of time, a motion which turned out to draw our participants focus, without providing any information as to the specific action being performed (climb, descend, bank right or left). Figure 3 shows a series of key frames (or snapshots) drawn from the resulting animated warning, which, when viewed rapidly in sequence, combine to make up the warning animation described.

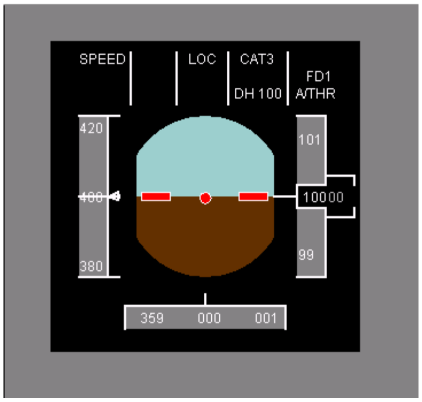

Fig 1: Airbus A320 Primary Flight Display (PFD)

Lastly, we added a third condition in which the warning signal provided was both attention grabbing and semantically relevant to the underlying autopilot action. In order to do this, we use a technique developed by the Disney animation studios in the mid 1900's. This signal involved a small "anticipatory" or "predictive" movement on the central display. In other words, shortly before the autopilot turned the plane to the right, the icon representing the aircraft (see figure 4) would draw to the left, before jumping back in the intended direction. Similarly, a left turn would be preceded by a shift to the right and an ascent by a small downward movement. This "anticipation" can be most clearly understood in terms of a runner taking a step backwards before running (forwards) down a path.

In this third condition (C3), we believed that the semantic relevance of the signal to the underlying autopilot activity would reduce the cognitive load on the viewer by reducing the amount of mental work required to map from the incoming signal to its underlying meaning – a notion explored by Johnson, Johnson & Hamilton [11] in the field of task performance, but often overlooked in HCA support. We believed that this reduction in cognitive load would lead to the higher levels of awareness being achieved with greater regularity and, ultimately, reduce the number of breakdowns observed.

We reasoned, however, that the utility of our two warning signals could vary both with the perceptual strength of the signal used (i.e. its size, range of movement, brightness etc) as well as its semantic relevance. In order to avoid a potential confound, therefore, we made sure that the warning signal involved in C3 contained an icon of similar size and brightness to that in C2 but actually involved a smaller range of motion. We could, therefore, argue that the warning signal in condition C2 was of high signal strength and low semantic relevance, whilst the one in C3 offered the reverse.

If we were to gain support for our model, then, we would need to show that support provided in each condition (i.e. the support targeted at different levels of awareness) would provide tangibly different results, ultimately affecting the extent to which people saw, attended to and/or understood the developing path of the flight. With this in mind, our first hypothesis was that the provision of any warning signal (i.e. any attempt to draw our participants attention towards the interventions) would increase the likelihood that these interventions would be reported.

More clearly stated then, this first hypothesis (H1) became:

H1: An explicit signal indicating autopilot activity would increase the number of reported observations that such activity had occurred i.e. a significantly greater number of such reports will occur where a warning signal is given (i.e. C2, C3) than in the condition where it is not, (C1).

Beyond this, however, we were able to make predictions about the likelihood of our participants moving from simply noticing that something had occurred to understanding what it was. Again we phrased this second hypothesis (H2) more clearly in terms of our experiment:

H2: The inclusion of a specific semantic link between the warning signal given and the underlying autopilot activity will increase the participants' understanding (cognitively processed awareness) of such activity, leading to a significantly higher rate of correction of those undesirable interventions reported. i.e. the ratio of interventions corrected to interventions reported will be significantly higher in the condition where explicit semantic support is included (C3) than in those where it is not (C1, C2).

We also included a third, weaker hypothesis that would hold only if the signal strength in C2 and C3 turned out to be similar, rather than skewed in the way that we had intended. In this case, if no significant difference existed in the number of interventions reported, we would expect to see not only a rise in the ratio of corrections to observations, but also a significant higher number of corrections per intervention in C3 (vs C2). In other words, if the only important difference between the characteristics of two interaction designs is that one better supports the cognitive processing of information which is available, perceived and attended to, then that design will result in a higher incidence of true awareness (understanding) than its competitors. In terms of our experiment, this third hypothesis (H3) could be phrased as follows:

H3: If the number of reported interventions is similar in the two conditions involving warning signals (C2, C3), then the absolute number of corrections observed in C3 will be significantly higher than in C2.

With these hypotheses in mind, we asked thirty postgraduate students to participate in our between-subjects experiments (separated into three groups of ten, one for each of our three conditions). Clearly, the use of non-professional participants reduced the ecological validity of our experiment, but the resource of commercial pilots' time is extremely limited and we felt that our interface literate replacements would be sufficient for this initial empirical study.

Having recruited our participants, we set up a simple working simulation of the panel and displays in question on a Pentium-4 PC with a 19" screen. We then programmed our control and extended interfaces using the Java programming language, relying heavily on the swing graphical interface packages to produce the simulated interfaces described below.

First, we constructed an input interface, a faithful replica of the Flight Control Unit (FCU) used in the A320. The FCU, shown in figure 2, consists of four dials, six buttons and three switches. Most importantly in the context of this experiment, the dials allow targets for speed (SPD), lateral heading (HDG), altitude (ALT) and vertical speed (VS) to be given to the autopilot. For those interested in a more complete description of the FCU or of the other A320 panels described here, one can be found in our previous work on the subject [15] or in the official accident report of the Strasbourg crash [6].

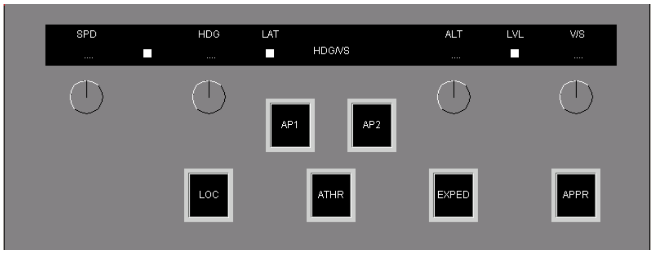

Fig 2: Airbus A320 Flight Control Unit (FCU)

When not entering a set of target parameters to the autopilot, the participants were asked to monitor the flight path via the display, which provided the only indication that the flight was/was not following its intended trajectory. The display concerned, appropriately named the Primary Flight Display, or PFD, showed an animated representation of the aircraft's flight relative to an artificial horizon, along with a series of moving bars indicating current speed, altitude and heading.

Having simulated both input panel and display, we added two further software modules, the first of which played a pre-recorded audio track containing a series of instructions from a fictional air traffic controller, whilst the second recorded the participants' input to the FCU. The resulting automated recording procedure was supplemented by manual recording on a carefully standardised form both during each the experiment and in a tightly controlled post-experiment debrief.

Having completed this preparation, our next task was the implementation of our experiment: Each participant was first given training in the use of controls and displays provided in the simulation and required to successfully complete a practice run before undertaking the final, recorded flight scenario. Importantly, they were given the context within which the test scenario would take place (a descent to an airport which was below them and to their right). They were also told that the automated systems might "assist" in their flight, that this might involve independent interventions, and that an ongoing awareness both of the interventions and the flight path itself would be crucial to a successful completion of the mission. They were not, however, given any further details about the interventions themselves.

Our next task was the implementation of our experiment: each participant was first given training in the use of controls and displays provided in the simulation and required to successfully complete a practice run before undertaking the final, recorded flight scenario. Importantly, they were given the context within which the test scenario would take place (a descent to an airport which was below them and to their right) and briefed on the vocabulary to be used throughout. They were also told that the automated systems might "assist" in their flight, that this might involve independent interventions, and that an ongoing awareness both of the interventions and the flight path itself would be crucial to a successful completion of the mission. They were not, however, given any further details about the interventions themselves.

The participants were then asked to complete a portion of descent from 10,000 feet altitude to a point shortly before the final landing sequence. ATC instructions and interaction with a fictional co-pilot were recorded on tape and our subjects were asked to complete them. Before commencing the flight, each participant was fully briefed on the flight context (goals, objectives and priorities) and given skeletal information about other environmental factors.

They were then asked to enter appropriate settings to the input panel, such that the autopilot would execute the instructions given by ATC e.g. in response to the ATC request "go to 5000 feet for now and await further instruction", the participant would instruct the autopilot to execute a descent to 5000 feet and watch the display to ensure that this was carried out. The final recorded scenario was based upon flight transcripts of an authentic A320 descent, with instructions, vocabulary and timing being almost identical to those reported.

Next, distractions and diversions were added through the use of checklists and verbal confirmations, based upon those observed during flight. Examples of these distractions included requests to confirm flight parameters, check weather reports and set speed bugs. In all, every attempt was made to recreate a complex, busy environment in which sufficiently competent participants could attempt authentic scenarios armed only with the electronic flight instruments described above.

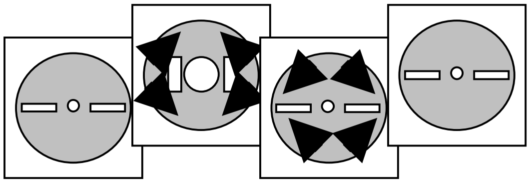

Figure 3: Keyframe Sequence From Condition C2

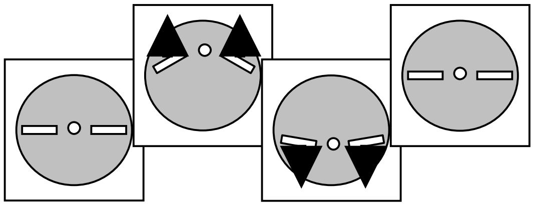

Figure 4: Keyframe Sequence From Condition C3 (Descent)

Given the difficulty of proving that participants had or had not noticed an intervention, we made three kinds of observations during the flight scenarios. Firstly, each interaction between pilot and FCU was recorded with dial (or button / switch), timestamp and resulting parameter target all noted. This allowed us to extract corrections entered by the participants when undesirable interventions were noticed. Secondly, we asked the participants to verbally inform the experimenters in the case that they saw either changes in flight path or any other unexpected behaviour on the part of the aircraft during the scenario (concurrent protocol). Finally we debriefed the participants at the end of each run, asking them again whether the aircraft and controls had behaved as they expected throughout (post-hoc protocol).

RESULTS

The results in each of our three conditions are reported in Figures 5, 6 and 7, below. The tables are divided into two sections. The first (left hand) side describes the reaction of each participant (numbered P1 to P10), in each condition to each individual intervention (described, in order of occurrence as Int1, Int2, Int3 and Int4). Each reaction is recorded as follows:

No Entry: Participant neither reported nor corrected the undesirable intervention.

R: Participant reported that "something was wrong" after an intervention, but was not sufficiently aware of the details of the intervention to be able to correct it.

C (or Time in seconds):Participant corrected the undesirable flight path (i.e. moved the aircraft back onto an appropriate downward and right turning trajectory). The time taken to effect this correction (in seconds) is also given.

Following this coding scheme, then, we can see from figure 5 that the third participant in the first condition missed the first intervention (Int1), noticed but was not able to correct the second intervention (Int2), noticed, understood and corrected the third intervention (Int3), taking 25 seconds to do so and then failed to notice the fourth intervention (Int4). These figures sum to give two interventions either reported or corrected of which only one was, in fact, corrected.

The right hand side of each table then summarizes the results for each participant, showing (from left to right), the total number of interventions and/or corrections, the total number of reported (but not corrected) interventions and the total number of corrections.

At this early stage in our results, a number of interesting trends were apparent: Our first hypothesis (H1) involved the frequency with which our participants would notice (report) the undesirable autopilot interventions. Specifically, we believed that the inclusion of some anticipatory warning signal would significantly improve the likelihood that interventions would be reported. The baseline for comparison was the control condition in which only twelve of the 40 interventions were reported (i.e. 12/40 = 30% of interventions were recognized using only the animation of the animated display). This result is particularly interesting since it is very close to the results obtained by Johnson and Pritchett in their reconstruction of the Strasbourg accident, suggesting that we had some success calibrating the complexity of experimental task to a realistic level of difficulty. The results of our second condition, C2 (i.e. the condition in which the interventions were indicated by additional animation of high signal strength but low semantic relevance) saw observations rise to 62% (i.e. 25 of a possible 40 interventions) and the third, C3 (in which we used a smaller but semantically more relevant signal) returned 47% (19 of a possible 40).

Here at least, we were able to perform one-way ANOVAs on the number of reported interventions (regardless of ultimate corrections). In individual tests, we found that the results in C1 were significantly different from the much higher numbers found in C2 ($p < 0.05$) and C3 ($p<0.05$), supporting H1 and, with it, our belief that an understanding of independent autopilot activity was important to the participants understanding of the flight. Importantly, we also found no significant differences between C2 and C3, leading us to conclude that we had assured approximate parity between the signal strengths in each condition.

As we described earlier however, alerting our participants that an intervention was taking place was only one part of our objective in this experiment. We were also interested in the frequency with which they would go beyond the level of awareness at which they *knew that something was happening* to the level at which they *understood the intervention* sufficiently to correct its consequences. Here again, our results were encouraging with only 2/40 (5%) of interventions being corrected in the control condition (C1), rising to 3/40 (10%) in C2 and 9/40 (22%) in C3, the condition offering most semantic connection between the signal chosen and the underlying autopilot action.

Intervention:	Int1	Int2	Int3	Int4	Total (R +/or C)	Total (R)	Total (C)
Participant							
P1	R			R	2	2	0
P2	R		R	**(11s)**	3	3	1
P3		R	**(25s)**		2	2	1
P4					0	0	0
P5			R	R	2	2	0
P6					0	0	0
P7			R	1	1	0	
P8		R			1	1	0
P9			R	1	1	0	
P10					0	0	0
Total	2	2	5	3	12	12	2
Mean					1.2	1.2	0.2

Fig 5: Results For Control Condition (C1) – No Warning Signal on Autopilot Intervention

Intervention:	Int1	Int2	Int3	Int4	Total (R +/or C)	Total (R)	Total (C)
Participant							
P11		R	**(18s)**	**(12s)**	3	2	2
P12	R			R	2	2	0
P13			R	R	2	2	0
P14	R			R	2	2	0
P15	R	R			2	2	0
P16	R	R	**(6s)**	R	4	3	1
P17	R	R			2	2	0
P18				R	1	1	0
P19	R		R	R	3	3	0
P20	R	R	R	R	4	4	0
Total	7	5	5	8	25	23	3
Mean					2.5	2.3	0.3

Fig 6: Results For Second Condition (C2) - High Signal Strength, Low Semantic Salience

Intervention:	Int1	Int2	Int3	Int4	Total (R +/or C)	Total (R)	Total (C)
Participant							
P21			R	**(4s)**	2	2	1
P22	R		R	R	3	3	0
P23					0	0	0
P24	**(8s)**	**(5s)**	**(5s)**	**(10s)**	4	3	4
P25				R	1	1	0
P26			R	1	1	0	
P27			R	**(5s)**	2	2	1
P28	**(4s)**	**(8s)**	**(6s)**		3	3	3
P29		R			1	1	0
P30	R	R			2	2	0
Total	4	4	6	5	19	18	9
Mean					1.9	1.8	0.9

Fig 7: Results For Third Condition (C3) - Low Signal Strength, High Semantic Salience

In this case, our parametric tests (one way ANOVA) showed no significant difference between either C1 and C3 or C2 and C3, although as we have shown above (and will discuss further below) there is a clear trend towards improvement in C3. One explanation could be that our relatively small number of participants contributed to these non-significant results, but we must also consider the notion that individual differences between participants played a role. Nonetheless, we still draw some encouragement from these results. A glance at Figure 8, for example suggests that a strong trend emerged both in the number of reports actually leading to corrections and the absolute number of corrections in C3 were dramatically different from the corresponding number in C2, providing support, at least for the spirit of hypotheses H2 and H3.

Equally interesting was the fact that these results were more widely spread between participants in C3 (in which 4 participants registered at least one correction) than in either C1 or C2 (2 participants registered at least one correction in each condition). Our results could not, therefore, be explained simply by a higher ability to follow complex data on the part of the successful participants in condition C3.

Despite these pockets of success, however, our experiment is not without flaws. Since these flaws also constitute a form of feedback to the community, we will discuss them here, before summarizing our findings and drawing up plans for the future. Our first problem was that some priming seems to have crept into our results, despite our efforts to the contrary. If we revisit the tables of results in Figures 5,6 and 7, it becomes clear that the frequency with which participants reported interventions rose dramatically in the second half of each scenario. Intuitively, at least, it seems that each participant who had noticed at least one intervention paid closer attention to subsequent developments in the scenario i.e. noticing a first intervention primed our subjects to notice the next one. As we dig deeper into this phenomenon, however, we find that this effect is noticeably weaker in our high signal strength condition (C2) than in (C1) or (C3). The high signal strength in that condition, combined with a relatively low workload at the beginning of our scenario may go some way to explaining the unusually high reporting frequency for Int1, but we do not feel we have fully understood this phenomenon and will return to it in future research.

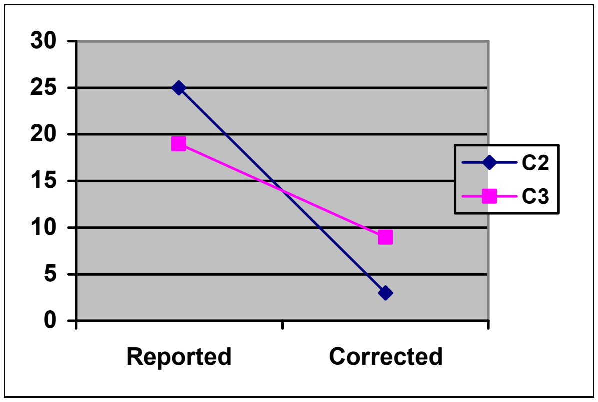

Figure 8: Total Number of 1) Reported and 2) Corrected Interventions Across All Participants, Grouped By Condition

If we return to our priming concern, however, it seems likely that we would still find support for our hypotheses, even if we had cut the length of our scenario considerably (i.e. reduced the opportunity for priming). If we consider only those results obtained in the first and second interventions (i.e. we ignore the second half of each scenario), in order to remove this priming effect, we can still see our expected trend towards higher reports in C2 (C1=4, C2=12, C3=8) and towards higher corrections in C3 (C1=0, C2=0, C3=4). Whilst our original diagnosis of a priming effect stands, therefore, we find it not to have been fatal to the underlying themes of the paper.

More aggressively still, we could take the opposite approach to these unexpected results and, in the place of excluding them from our findings, treat them as a pointer towards further useful research in the area. In this vein, we could assert this priming effect as a largely *positive* (though still unanticipated) outcome of this study, which if found to be repeatable and controllable, such an effect could lead us to an even more efficient support of high level awareness through the use of regular animated updates to pilots – sadly such findings are far beyond the scope of this current report and must be left for the future.

Returning to our immediate concerns, we could also find fault in the granularity of measurement, used to determine a “correction” in this experiment. Previous researchers [18] have shown in previous experiments that professional pilots are just as likely to pick up unintended autopilot activity from an unexpected flight path or aircraft position as from the primary flight display. In our experiment, then, we are prone to results, caused not by participant awareness of our warning signals, but rather by a comparison of subsequent deviation from the expected flight path (a result which lies outside the scope of our current hypotheses). The discrepancy in the mean delay between corrections being achieved in C2 and C3 (C2=12 seconds, C3=6 seconds), could, therefore, in part, be explained by this phenomenon.

Once again, however, whilst we concede the fact that future research may be needed to fully explain this phenomenon, it seems unlikely that the thrust of our argument need be greatly altered. There is no reason to believe that a significantly greater number of such “subsequent corrections” would have occurred in our semantically relevant condition (C3) than in its strong signal counterpart (C2). If any bias exists, it seems likely that C2, with its extended delay times would be its most likely location. By extension, therefore, the risk to our results is that the ratio of corrections in C3 against C2 would actually have risen, strengthening our claims. Once again, the minimum we can assert is that no damage was done to our central thesis.

DISCUSSION & CONCLUSION

So what, then, have we learned from this exercise? At the level of our specific hypotheses, we can report encouraging results. We were able to successfully demonstrate an ability to affect the number of undesirable events observed by our participants (i.e. hypothesis H1 returned a significant result) and a noticeable trend emerged in the number of observations, which led to *corrections*. Whilst our ability to statistically support these findings is currently limited, a glance at the distribution of our findings strongly suggest a relationship between our design manipulations and the final results in our experiment.

If we move to the level of the particular domain and context chosen, then, we can claim a limited advance in our search for specific design solution, which deals with the problems observed in the Airbus A320. On one hand, at our chosen level of authenticity (moderate), we seem to have found an interface design, which increases the likelihood that people (at least in the age range provided by our participant population) notice unexpected automation activity. In this sense, we have demonstrated a proof of concept for a specific enhancement, which can be fed back into our chosen domain for further field-testing.

Both the non-professional participants and the practical limitations in our simulation, however, keep us from making the assertion that we have a ready made solution, which could be used “As is” in a full scale commercial cockpit. It is, therefore, at the theoretical level, with support from some encouraging empirical results, that we make a contribution back to the wider community of HCA researchers and practitioners. For those involved in the development and prototyping of awareness supporting systems, for example, we have proposed that an approach based on a deeper, explanatory account of the processes by which awareness is achieved brings tangible utility to the design process in at least one authentic and problematic area. We have also provided initial evidence of a category of breakdowns, which cannot be solved simply through an increase in the signal strength of appropriate alarms or alerts. We believe this theoretical model can extend to imply further challenges for those interested in peripheral awareness. Could we, for example, manipulate the semantic relevance of our information sources rather than their raw signal strength to produce the kinds of interaction and awareness required by peripheral displays?

In conclusion, we believe that that this work does make a direct contribution to the wider field of awareness research. We have demonstrated, for example, that our model, involving availability, perception, attention and semantic processing is important not only in description of the space as a whole but also in the identification of local design solutions which allow us to identify individual, if overlapping elements, which must be included in any comprehensive support for higher level awareness. As yet, we can say little about the difference between particular design elements which draw attention to pertinent events and those which facilitate the further (semantic) processing of the information attended to, but we have clearly shown that the raw perceptual strength of a given signal is far from the end of the story in the creation and support of awareness.

ACKNOWLEDGMENTS

This work is funded by the EPSRC (grant number GR/R40739/01) and supported by QinetiQ and GKN Westland Helicopters.

REFERENCES

1. Gaba DM, Howard SK, Small SD. *Situation Awareness In Anaesthesiology*. Human Factors 37 (1995) pp20-31.
2. Hutchins, E. (1995).*Cognition In The Wild*. MIT Press.
3. Mankoff J, Dey AK, Hsieh G, Lederer J, Ames S, M.. *Heuristic Evaluation Of Ambient Displays.* In Proc. CHI 2003.
4. Tan, D.S., Czerwinski, M, & Robertson, GG. *Women Go With The (Optical) Flow.* In Proc. CHI 2003. pp. 209-215,
5. Grudin, J, *Partitioning Digital Worlds: Focal And Peripheral Awareness In Multiple Monitor Use*. In Proc. CHI 2001
6. Bureau Enquetes Accidents. (1992). *Rapport De La Commission D'Enquête Sur L'accident Survenu Le 20 Janvier 1992 Près Du Mont Sainte-Odile (Bas Rhin).*
7. ATSB. (2001). *Investigation Report 1999045538. Boeing 747-438, VH-OJH, Bangkok, Thailand, 23 Sep 1999.*
8. Wiener, E. (1989) *Human Factors Of Advanced Technology ("Glass Cockpit") Transport Aircraft. NASA Contractor Report 177528*
9. Hourizi R, And Johnson, P. (2001). *Beyond Mode Error: Supporting Strategic Knowledge Structures To Enhance Cockpit Safety.* In: Joint Proc. HCI 2001 And ICM 2001.
10. Cadiz, J., Gupta, A., Jancke, G., Venolia, G (2001) *Sideshow: Providing Peripheral Awareness Of Important Information.* Microsoft Research Tech Report Msr-Tr-200183.
11. Johnson, P, Johnson, H, Hamilton. (2000). *Getting the Knowledge into HCI - Theoretical + Practical aspects of Task Knowledge Structures.* Schraagen, JM, Chirman, SF and Shaun, VL (eds) Cognitive Task Analysis. pp 201-214
12. Sarter, N.B. And Woods, D.D. *How In The World Did We Ever Get Into That Mode? Mode Error And Awareness.In Supervisory Control.* Human Factors, 37. pp 5-19.
13. Endsley, M. R. (1999). *Situational Awareness In Aviation Systems.* In: Garland, Wise, & Hopkin. (eds): Handbook Of Aviation Human Factors.
14. Rushby, J. (1999). *Using Model Checking To Help Discover Mode Confusions And Other Surprises.* In Proc. 3rd Workshop On Human Error, Safety And System Development.
15. Hourizi, R., & Johnson, P. (2001). *Unmasking Mode Error: A New Application Of Task Knowledge Principles To The Knowledge Gaps* In Cockpit Design. Proc. Of Interact 2001.
16. Barnard P.J. And May, J. *Towards A Theory Based Form Of Cognitive Task Analysis Of Broad Scope And Applicability* (2000) In J.M.C. Schraagen And S.F. Chipman And V.L. Shalin (Eds) Cognitive Task Analysis pp147-163
17. Thomas F, Johnson O. *The Illusion of life: Disney Animation.* New York, NY, USA: Little, Brown and Company; 1981.
18. Johnson, A, Pritchett, A.R, (1995). *Experimental Study Of Vertical Flight Path Mode Awareness.* Presented At 6th IFAC/IFIP/IFORS/IEA Symposium On Analysis, Design And Evaluation Of Man-Machine System.