Workload benefits of colour coded head-up flight symbology during high workload flight

Blundell, J., Scott, S., Harris, D., Huddlestone, J. & Richards, D.

Original citation & hyperlink:
Blundell, J, Scott, S, Harris, D, Huddlestone, J & Richards, D 2020, 'Workload benefits of colour coded head-up flight symbology during high workload flight', Displays, vol. 65, 101973.
https://dx.doi.org/10.1016/j.displa.2020.101973

DOI 10.1016/j.displa.2020.101973
ISSN 0141-9382

Publisher: Elsevier

Workload Benefits of Colour Coded Head-Up Flight Symbology during High Workload Flight.

Authors:

Corresponding Author: James Blundell[1] (ac7045@coventry.ac.uk),

Co-Authors: Steve Scott[1] (ab8575@coventry.ac.uk), Don Harris[1] (ab3693@coventry.ac.uk), John Huddlestone[1] (ab4919@coventry.ac.uk), Dale Richards[1] (ab3204@coventry.ac.uk)

[1] Institute for Future Transport and Cities, Coventry University, Coventry, UK

Abstract

The manipulation of colour in display symbology design has been recognised as a method to improve operator experience and performance. An earlier paper by the authors demonstrated that redundantly colour coding head-up flight symbology supported the manual flying performance of both professional and non-professional pilots during low-workload flying scenarios. In this study the workload and performance of 12 professional airline pilots was evaluated in high workload conditions whilst they flew manoeuvres and an instrument landing system (ILS) approach with and without the presence of colour feedback on a head up display (HUD). Workload was manipulated by presenting pilots with a concurrent auditory n-back task. Colour coded flight symbology reduced the subjective workload of the pilots during high workload conditions. In contrast, manual flying performance during high workload was not improved by the presence of colour coded feedback.

Keywords:

Head up symbology, Colour, Aviation, Workload, Performance

Funding: This work was supported by Innovate UK, the UK's Innovation Agency, with support from the UK Aerospace Technology Institute (grant reference number 113108)

1. Introduction

In complex operational situations it is vital that the pilot is able to acquire sufficient levels of situational awareness to successfully meet the goals and objectives of a specific task or function. Endsley (1988) formally defines situational awareness (SA) as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status into the near future" [1]. Acquiring situational awareness is built upon a foundation of perceiving, "noticing", data and information inside and outside the flight deck that are relevant to a task. For instance, the safety of a flight is predicated on the pilot's ability to perceive and understand data and information related to their aircraft's height, speed, attitude and track, as well as the location of other potential hazards, such as other aircraft, weather and terrain.

What we perceive in a complex visual scene, and thus what data and information we attend to, is determined by top-down and bottom-up attentional mechanisms. Top-down mechanisms endogenously direct attentional resources towards data and information in the visual scene that the individual's mental model deems to be task-relevant to present and future states of their current situation [2]. In contrast, bottom-up attentional mechanisms involve the pre-attentive exogenous influence of intrinsic visual scene features (e.g. luminance, motion, size, orientation). Based on a strong body of neurophysiological and psychological research it is assumed that the salience characteristics of individual visual scene elements are summed, weighted, and spatially organised within salience maps [3]. In turn, locations with relatively high activity on a salience map correlate with visual scene areas containing the highest salience and, consequentially, correspond to areas of the visual scene that the individual is more likely to attend to. A pre-attentive feature channel that has been shown to robustly elicit strong activity within salience maps is colour [4–9]. This is particularly the case when colour coding complements top-down attentional mechanisms that are in line with the individual's mental model. For instance, where distinct colour coding is used to discriminate target relevant and non-critical information within a semantically viable context [10,11].

Colour as a pre-attentive attentional guide has long been recognised as a feature of head down cockpit display design that can be manipulated to improve operator experience, workload capability, and performance [12–16]. In contrast, the human factors considerations of colour use within head up displays have been largely ignored. This is echoed in the absence of head up symbology specific colour guidance from the Federal Aviation Administration (FAA Advisory circular: 25-11B [17]). In part, this is due to the practical difficulties of achieving sufficient luminance contrasts for a colour symbology set that is presented against the variable luminance profile of the outside scene. However, research by the United Force Airforce (USAF) has addressed several of these relevant luminance contrast ratio issues, at least with head mounted display (HMD) technologies [18,19]. Furthermore, recent advancements in HUD and HMD technology in the form of waveguides means that colour head up symbology could become an ubiquitous feature of advanced flight decks in the near future [20].

There have been mixed findings from the limited number of studies that have evaluated the performance benefits afforded by the use of colour in HMDs and HUDs in the military domain. Post, Geiselman and Goodyear [21] found that using colour to group task-related information supported visual search within a cluttered HMD, which enabled military pilots to reduce missile release time without sacrificing probability of kill. Dudfield [22] found that colour-coded error feedback provided on HUD flight symbology did not support pilot flying accuracy during straight and level flight, but did benefit subjective workload. Findings from our own recent study [23] demonstrated that a similar colour-coded head up symbology set granted minor and major flying accuracy enhancements to professional airline pilots and novices, respectively. It is likely that the performance benefits found in our study arose from the more varied range of manoeuvres, such as climbing turns, that required pilots to attend to both dynamic and static flight parameters. On the other hand, the performance benefits were observed in absence of corresponding subjective workload improvements. Such dissociations between performance and workload are common [24]. It is often observed that different pilots can produce similar levels of performance on a task but at a very different cognitive “costs”, assessed by their residual information processing capacity. The converse is also true: similar levels of workload may result in differing levels of performance [25]. Furthermore, in Dudfield’s study [22] a secondary task was employed to increased workload and divert attentional resources away from the primary flying task. However, whether Dudfield’s reported workload benefits were a consequence of colour supporting workload more generally or were more specific to high workload scenarios was unclear, since no colour-by-workload interaction analyses were reported. It is possible that the full benefit of colour coded head up symbology may only be realised in high workload flight scenarios, where spare mental capacity is more limited.

The aim of this study was to further investigate professional pilot flying performance, and the associated subjective workload and situational awareness benefits, during a manual flying task using either a colour or monochrome HUD. Similar to our previous study, HUD colour coding cues were based on economy to notify pilots when they flew outside pre-determined boundaries. We expanded on our own and Dudfield’s research design by evaluating HUD colour across several flight manoeuvre types in the presence or absence of an auditory working memory secondary task. In addition, both Dudfield and Blundell et al concluded that there was a need to evaluate the benefits of colour coding during flight conditions that place emphasis on visual attention and planning. Therefore, in addition to basic flying manoeuvres, we include an evaluation of the colour coding effects on flying accuracy and subjective workload and situational awareness during a precision instrumented landing system (ILS) approach and landing. Subjective measures of workload and situational awareness were measured via the NASA Task Load Index (TLX) and Situational Awareness Rating Technique (SART), respectively.

2. Materials and methods:

2.1 Participants

Twelve professional airline pilots, all holders of an Airline Transport Pilots Licence (ATPL), participated in the study. The rank of pilot participants included 5 First Officers and 7 Captains. Pilots' average flying experience was 4,950 hours (SD = 3,926). The majority of pilots (7/12) had predominantly experienced flying Airbus airliner types (e.g. A320-60). The experiment was approved by the Coventry University Ethics Committee and was in line with ethics guidelines as per the British Psychological Society.

2.2 Apparatus

A low-fidelity desktop flying simulator running X-Plane 9.71 (Laminar Research) was used for the flying task. Pilots were seated at a viewing distance of 75 cm from a 55-by-40 cm display, producing a total field of view of 40.3 deg. Pilots controlled a single-engine jet aircraft (the Cirrus Vision SF50) via a commercially available sidestick and throttle (Thrustmaster® Stick T flight Hotas, http://www.thrustmaster.com/en_UK/products/t-flight-hotas-4-ace-combat-7-limited-edition).

2.3 Scenario and Procedure

The current study involved a simulation scenario requiring participants to fly a light, single-engine jet aircraft in a 20-minute manual flying task involving two flight phases: 1) basic flight manoeuvres (which consisted of 4 distinct segments: flat turn, a straight and level, a climbing turn, and a descent) and; 2) an Instrument Landing System (ILS) approach (consisting of 3 segments: ILS intercept, approach, and final approach).

The experiment took approximately 2.5 - 3 hours to complete depending on the time pilots required to familiarise themselves with both the aircraft's handling and the feedback behaviour of the redundantly colour-coded HUD symbology (see section 2.4). Scenario instructions were relayed to pilots via pre-recorded air traffic control verbal messages at specific points of the trials in order to inform pilots which manoeuvres to conduct (e.g. "Turn left onto heading 230. Maintain 4000ft").

Pilots repeated the flight scenario four times. Each scenario trial varied depending on the presence/absence of HUD colour feedback and the auditory working memory task (2-back task – see section 2.5.3). The order pilots flew the four trials was divided into two HUD colour feedback dependent blocks. Where within each block the low workload scenarios (i.e. without the secondary task) were completed first and then the higher workload scenario (i.e. with the secondary task). The order of blocks, thereby the exposure to HUD colour feedback, was counterbalanced between participants. While a true counterbalanced approach would have been desirable, the decision to maintain the within block order of workload scenarios was taken to maximise the time pilots had to familiarise themselves with the primary task (flying the aircraft) during the limited time window of the study. Thus reducing the likelihood of superficially inflating overall task difficulty during potential initial trials that included the secondary task.

2.4 HUD flight symbology

The HUD symbology employed in the current study was adapted from the generic X-Plane HUD. A total of four HUD symbology items were colour coded (indicated airspeed tape, bank angle indicator, altitude tape, and the ILS localiser and glideslope) to provide real time flight profile feedback to pilots. Colour coding conventions were aligned with the FAA's guidance on colour coding for electronic flight displays (Advisory circular: 25-11B [17]). Figure 1 depicts the HUD symbology used alongside examples to illustrate the functionality of the colour coding implementation. For instance, if the pilot veered off course, flew too low or high, too fast or slow, or deviated from the ILS localiser or glide scope, the relevant HUD symbology item would transition from green to amber and ultimately to red, depending upon the size of the deviation from a set of flight profile criteria.

In the current study the criteria for the flight profile limits were established via discussion with two subject matter experts. This human-centred design approach allowed us to incorporate the end users perspectives in order to achieve a usable system [26], whilst also taking the context of use into consideration [27]. In particular, in regards to air speed, more narrow deviation allowances were permitted for under speeding compared to over speeding, with the intention of protecting pilots from putting the aircraft into a stall condition. For example, the indicated airspeed tape would turn to amber if a participant breached the instructed airspeed limit by -3 / +10 kts (e.g. Figure 1, panel B). Similarly, whilst flying an ILS, if the participant deviated from the localiser between 1-2 dots the localiser HUD item would change from green to amber (e.g. Figure 1, panel D).

The application of colour feedback to certain HUD symbology items was dependent on the current scenario phase segment pilots were flying. Specifically, colour feedback was only allocated to a HUD symbology item when the item's respective flight parameter required maintaining; during the straight and level manoeuvre segment feedback was applied to the airspeed tape, bank angle indicator, altitude tape, whereas during the climbing turn segment the colour feedback on the altitude tape was inactive. Similarly, during the ILS intercept and approach segments colour was applied solely to the ILS localiser and glideslope symbology. Colour was, however, reintroduced to the airspeed tape during the final approach segments (once pilots were stabilised 1000ft above the runway) in order provide VREF feedback. Details of how and when colour feedback was presented on each of the HUD symbology items are shown in Table 1.

HUD colour was manipulated in real time using Matlab (version R2018b) with the NASA X-plane Communications Toolbox (https://github.com/nasa/XPlaneConnect). The toolbox enables a user datagram protocol (UDP) to send relevant flight parameter data (e.g. current altitude, speed etc) from X-plane to Matlab. In turn, a custom Matlab function inspected the retrieved flight parameter data to monitor deviations from the set of flight profile criteria. In the event that a flight profile boundary was breached Matlab signalled to X-plane (also via UDP) the necessity for colour feedback to be presented on respective HUD symbology items.

Figure 1 around here

Table 1 around here

2.5 Outcome Measures

2.5.1 Subjective Measurements

Participants completed 2 questionnaires after each trial: 1) the NASA Task Load Index (TLX)[28] and 2) the Situational Awareness Rating Technique (SART) [29]. The TLX is a long-established scale designed to capture subjective workload ratings across six workload dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration. Each workload dimension is measured on a scale from 0 – 20 [30], where higher ratings represent higher subjective workload. In the current study, raw workload scores from the six TLX dimensions were analysed instead of using the alternative dimension weighting method. This abridged approach was chosen due to the inconclusive evidence that dimension weighting improves the TLX's sensitivity [30]. The SART is a 10 item questionnaire, on a seven point rating scale (1 = low, 7 = high). The 10 items map onto 3 distinct situational awareness (SA) dimensions: 1) demands on attention resources, 2) supply of attention resources, and 3) understanding of the situation. Composite SA scores were calculated and analysed by combining SART dimension responses using the formula: SA = (Understanding - (Demand – Supply)).

Furthermore, at the end of experiment, participants were invited to provide responses to a set of four open-ended post-trial questions related to the usability of the symbology (Appendix 1).

2.5.2 Primary Task Performance

For the basic manoeuvre phase, flight performance was measured as the root mean square error (RMSE) of a pilot's indicated airspeed (IAS), bank angle (BNK) and altitude (ALT) from each respective flight manoeuvre's limits. Table 1 shows the target IAS, BNK, ALT parameters that pilots were instructed to follow for the four different basic manoeuvre segments. RMSE performance measures were calculated based on a 2-minute sample window for each of the four manoeuvre segments.

Task performance during the ILS flight phase was based on pilots' ability to track the ILS localiser and glideslope. ILS tracking performance was simplified by analysing the RMSE of summed localiser and glideslope dot deviations (DOT). A single dot represents about 0.5 degrees of angular deflection. For each trial DOT RMSE measurements were taken from two 2-minute ILS segments: 1) ILS Approach - between when participants first intercepted and was established on the ILS (are within 0.5 localiser dots displacement) to 1000ft above the runway, and 2); ILS Final - the final 1000ft above the runway prior to landing. In addition, ILS RMSE was examined for the latter ILS segment where HUD colour coding was reintroduced to the indicated airspeed tape and pilots were instructed to maintain the aircraft's VREF (90 kts).

2.5.3 Secondary Task Performance

Scenario task load was systematically manipulated by presence or absence of a secondary working memory task, an auditory 2-back N-back task. Stimuli were created using Psychophysics Toolbox extensions within Matlab [31–33]. During trials where the secondary task was present participants were required to monitor a string of verbally presented letter stimuli (e.g. "A", "H", "F") and indicate whether the most recently presented stimulus matched the stimulus presented two trials previously. The inter-stimuli-interval was 250 milliseconds. Participants responded between stimuli presentations with either a "yes" or "no" response using the trigger buttons accessible by the index finger and thumb on the side stick. Errors were defined as incorrect responses (e.g. responding "yes" when the correct response was "no") and missed responses (e.g. the participant failed to respond to the most recent stimuli within the 250-millisecond response window). The percentage of errors was calculated for each of the six 2-minute segments (i.e. the four basic flight manoeuvres and two ILS segments).

2.6 Experimental Design and Analysis

Primary and secondary task performance, and subjective workload and situational awareness, data were analysed with a series of general linear mixed effects models (GLMMs) using the Matlab Statistical Toolbox. GLMMs are a powerful statistical method that allows the analysis of multiple observations from each participant without violating the critical statistical assumption of independence. For repeated measures designs, the use of common ANOVA methods which average across individual participant observations, are discouraged in favour of these more robust GLMM methods [34–37].

In the current analysis, we used maximal random effects structures that included random participant intercepts and slopes for all fixed effects that were included in the models. In this paper, we checked the significance (alpha = 0.05) of fixed effects by calculating p-values obtained by likelihood ratio tests. Visual inspection of residual plots was used to ensure no obvious deviations from homoscedasticity or normality.

3. Results

3.1 Subjective Measurements

To determine how HUD colour feedback affected pilot subjective workload a 3-way factorial GLMM analysis was performed on the post-trial NASA TLX workload ratings. Fixed effects were included for HUD symbology (*Colour*: colour vs. monochrome), the presence of the secondary N-back task (*TaskLoad*; Low vs. High) and the six different TLX dimensions (*Dimension*). For random effects, we included intercepts for participants, as well as by-participant random slopes for the three above fixed effects.

Pilot ratings on the six NASA TLX workload dimensions are presented in Table 2 below. The GLMM analysis revealed a significant one-way interaction effect between *Colour* and *TaskLoad* (F (1,263) = 21.71, $p < 0.001$). There was a main effect for *TaskLoad* (F (1,263) = 42.01, $p < 0.001$) but not for *Colour* (F (1,262) = 1.03, $p = 0.31$). Model coefficient estimates highlighted that the presence of the N-back task increased overall pilot subjective workload ratings by 7.74 ± 1.19 (SE) TLX points compared to when the N-back task was absent. The *Colour* and *TaskLoad* interaction reflected an overall reduction in TLX scores by 2.07 ± 0.44 (SE) points when HUD colour feedback was present during high task load scenarios. This is clearly depicted in Figure 2. HUD colour feedback had no effect on workload during low task load scenarios. There was neither a two- ($p = 0.28$) or one-way interaction ($p = 0.96$) between *Colour* and *Dimension*.

Post-trial SA responses on the three SART dimensions (demand on resources, supply of resources, understanding) and the calculated composite SART scores (Understanding - (Demand – Supply)) are shown in Table 3. Composite SART ratings were slightly higher when the HUD colour feedback was present during both low and high task load scenarios, which was attributed to lower subjective ratings on the Demand subscale. However, the GLMM analysis on SART composite scores revealed there was neither a main effect of *Colour* (p = 0.51) or an interaction between *Colour* and *Taskload* (p = 0.89).

Finally, findings from the post-study usability survey found that majority of pilots highlighted the usability value of colour in supporting attention during instances where the integrity of their scan pattern deteriorated.

Table 2 Mean inter-item NASA-TLX scores (range: 0 – 20) grouped by presence of colour coding feedback and scenario task load. NASA-TLX score standard deviations shown in parentheses.

Task Load	Low				High					
HUD Symbology	Mono		Colour		Mono		Colour		**Total**	
1 - Mental	6.33	(3.58)	7.50	(5.18)	17.05	(1.80)	14.64	(3.23)	11.38	(3.45)
2 – Physical	4.67	(4.62)	4.58	(4.80)	8.00	(5.62)	5.86	(4.12)	5.78	(4.79)
3 – Temporal	3.75	(1.66)	4.92	(3.58)	12.73	(4.20)	10.05	(4.29)	7.86	(3.43)
4 – Performance	5.50	(4.19)	5.50	(3.92)	11.68	(3.93)	11.05	(4.95)	8.43	(4.25)
5 – Effort	8.42	(4.06)	8.38	(4.47)	16.59	(1.77)	16.00	(2.54)	12.35	(3.21)
6 - Frustration	5.25	(3.74)	5.83	(5.24)	14.41	(3.51)	12.41	(4.65)	9.48	(4.29)
Total	5.65	(3.64)	6.12	(4.53)	13.41	(3.47)	11.67	(3.97)	9.21	(3.90)

Figure 2 around here

Table 3 Mean SART subscale scores (range: 1-7) grouped by presence of colour coding feedback and scenario task load. SART score standard deviations shown in parentheses.

Task Load	Low				High			
HUD Symbology	Mono		Colour		Mono		Colour	
1 - Demand	2.69	(1.32)	2.39	(0.74)	4.92	(1.46)	4.45	(1.75)
2 – Supply	4.56	(1.17)	4.19	(1.23)	5.43	(1.74)	5.40	(1.70)
3 – Understanding	5.08	(1.74)	5.46	(1.28)	4.36	(1.47)	4.30	(1.38)
Composite	6.94	(2.60)	7.26	(1.38)	4.88	(2.04)	5.24	(2.23)

3.2 Performance

3.2.1 Primary Task Performance

Primary task performance (RMSE) descriptive results for the different scenario segments of the basic flight manoeuvres and ILS flight phases are presented in Table 4. Separate GLMMs were conducted to analyse the four primary task performance outcomes (IAS, BNK, ALT and DOT). Each analysis included fixed effects for *Colour* and *TaskLoad*. Flight performance variations associated with the different scenario segments (e.g. descent vs climbing turn vs ILS Final) were not the focus of the current study, however, these parameters were included as random effects with the GLMM analysis in order to account for their influence on performace. Thus, random effects consisted of both intercepts for participants (N = 12) and flight segment (N = 6), as well as by-participant and by-flight segment random slopes for the above fixed effects. All RMSE data was log-transformed to account for floor effects that are typical of error data.

Overall, IAS RMSE differences were minimal across the four flight manoeuvre phase segments (S&L, Flat Turn, Descend, CT). Compared to the monochrome HUD, during the ILS Final scenario segment the colour HUD symbology reduced IAS RMSE by 5.5 kts across the low and high task load conditions on average. However, the GLMM analysis revealed no significant main effect for *Colour* ($p = 0.56$) or *Taskload* ($p = 0.07$), or an interaction between *Colour* and *Taskload* on IAS ($p = 0.12$*).* Similarly, ALT RMSE appeared to be lower with colour HUD symbology, particularly during the high task load conditions, however, the GLMM analysis here revealed neither a main effect of *Colour* or interaction between *Colour* and *Taskload*. The between pilot IAS and ALT variations were relatively wide during the ILS Final and S&L segments, respectively, particularly when flying with the monochrome HUD. Potentially contributing to the lack of significant *Colour* main effects and interactions.

In contrast, the analysis of BNK RMSE found a significant interaction between *Colour* and *Taskload* (F (1,82) = 6.15, $p < 0.05$). However, this reflected an increase in BNK error that was specific to when flying with the monochrome HUD in the low task load condition. Figure 3 clearly demonstrates this pattern of results in BNK RMSE performance.

The analysis of pilots' ILS localiser and glideslope tracking performance (DOT) showed no significant main effects or interactions ($p > 0.05$). Pilots' ability to maintain a prescribed airspeed, altitude, or to follow an ILS, were not affected by the presence of colour feedback in either of the task load conditions.

Table 4 around here

Figure 3 around here

3.1.2 Secondary Task Performance

Secondary N-back task performance, percentage or errors and response time, was analysed with the fixed effects *Colour* and *Phase* (basic flight manoeuvres vs. ILS phase). Descriptive results are shown in Table 4. Random intercepts were included for participant and the six different scenario segments over the basic manoeuvre and ILS phases, as well as by-participant and by-segment random slopes for the above fixed effects. For the percentage of N-back errors there was a significant main effect for *Phase* ($F(1,120) = 20.85$, $p < 0.001$). Pilots committed 11.4 ± 2.5% more errors when flying the ILS compared to flying manoeuvres. There was no significant main effect for *HUD* ($p = 0.80$) or an interaction between *HUD* and *Phase* ($p = 0.98$).

4. Discussion

The primary finding from the study was that colour-coded feedback reduced the workload experienced by pilots, and that this perceived benefit was exclusive to flying conditions where pilots were under higher task load demands imposed by the presence of a secondary task. The absence of a reported workload benefit under low task load conditions is in line with the findings from our previous study [23]. In addition, the current findings clarifies the nature in which workload was likely reduced in Dudfield's 1991 study [22], where the specificity of the colour-coded related workload benefit was not reported. Variations in pilot flying performance, in contrast, were independent of the colour-coded symbology. A single exception was found for bank angle deviation, where the colour-coded symbology appeared to support pilot flying performance in low task load situations. In fact, with the monochrome display pilot bank angle deviations were higher during the low task load condition compared to the high task load condition. One possibility is that the low task load condition induced a reduced arousal state, leading to a resultant drop in vigilance. Similar trends have been recently reported in military pilots flying combat tasks of varying complexity by Mansikka, Virtanen, & Harris [24] and other researchers [25,38,39], whereby lower performance can arise in the event of diminished task load and situational awareness.

Whilst performance appeared to be relatively unaffected by the presence of the colour symbology set in the high task load condition, subjective workload rating scores reported via the NASA-TLX were not. Hence, in a dual task scenario where spare mental capacity is more limited, there is the potential that the required workload "cost" for pilots to meet a given flight performance requirement is significantly reduced by the presentation of colour coded feedback. This explanation is in line with past research where colour has been used as a cue (or alerting signal) in situations where the presentation of information is more complex [12,15,25]. From a situational awareness standpoint, it can be assumed that the workload support granted by the colour symbology set may stem from two complimentary cognitive mechanisms. Firstly, the introduction of redundantly colour coded feedback provides a means of directly presenting higher level performance deviation information to the pilot. Presenting information in this manner has been shown to decrease the demands placed upon working memory and attention by bypassing, or supporting, the necessity to calculate performance deviation information based on lower level data sources [40]. Secondly, the colour symbology bestows a pre-attentive advantage to neglected flight information whenever their associated error margins are breached. Essentially, neglected information “pops-out” of the display[41–45]. Evidence for the latter pre-attentive benefit is offered by the pilots’ post-trial usability reports. Specifically, pilots found that the feedback supported them to reallocate attentional resources in instances where the integrity of their scan pattern had deteriorated. In this regard, such usability reports align with the functional benefits of colour that have been reported widely in the visual attention literature [41–45]. Thus, it is possible that the specific colour benefits found during the low task condition served as a situational awareness aid that potentially mitigated performance degradation during reduced arousal states.

Unlike the pilot usability reports, evidence that the colour symbology supplied a situational awareness benefit was unclear based upon pilot post-trial SART responses; Situational awareness ratings were marginally higher with the colour symbology but the difference was not significant. A possibility for this may lie with the choice of a post-trial situational awareness measurement approach. While the non-intrusive nature of the SART has clear merit, it does introduce the opportunity that post-trial ratings could selectively correspond to pilot performance [46]. For example, pilots may only rate their situational awareness high if they are also performing well, and they may forget instances where they had poor situational awareness [47]. The latter issue could have been further exacerbated by the 20-minute duration of trials which may have challenged post-trial recall of past situational awareness. This problem could have been further compounded by the large variation in flying experience that existed within the current pilot cohort. For instance, the use of colour in display design is known to facilitate learning [48], hence it is possible the situational awareness and performance benefits of colour may be more pronounced within a more inexperienced subset of the current participants. Nevertheless, overall these findings underline the importance of considering mental workload, performance, and task awareness together when evaluating operator performance in any complex human–machine systems.

A finer understanding of the workload and situational awareness benefits could be realised through the application of a more objective neuroimaging techniques. Recording various neurocognitive and physiological signals using a variety of these techniques (including Electrocardiography (ECG), electroencephalography (EEG), functional near-infrared spectroscopy (fNIR) and eye tracking) have been employed to across a range of settings, including aviation, to objectively measure mental workload and situational awareness [49]. In particular, EEG and ECG have been used successfully in air traffic [50] and military fast-jet settings [51] to differentiate between different levels of mental workload experienced by operators. Furthermore, EEG has been proven that it can be applied in an exploratory way to compare different display systems, such as HUD and HMD [52]. In the context of the current study, it is likely that the addition of neuroimaging techniques could have provided a clearer picture of *when* workload and situational benefits associated with the implementation of colour-coded head-up symbology occurred. Which, admittedly, is difficult to discern using a subjective post-trial approach.

Several important future implementation considerations of head up colour symbology remain to be addressed in order to appropriately review future FAA design standards that incorporate advanced head up symbology capability guidelines (e.g. FAA Advisory circular: 25-11B [17]). This includes investigating the systematic workload, situational awareness and flight performance benefits of colour head up symbology when viewed against different background scenes. Research by the United Force Airforce (USAF) has already addressed several perceptual processing issues in this respect, such as the appropriate luminance contrasts ratios of required to perceive coloured head up imagery against different background combinations [18,19]. However, a clear understanding of the relevant cognitive factors is elusive. For example, how might colour and other symbology features (e.g. shape, size etc) interact with one another to influence performance, workload and situational awareness? Indeed, McFadden, Kaufmann, and Janzen [53] suggested that shifts in colour appearance (as a function of surrounding colours) impair the accurate interpretation of information and suggested the use of limited colour combinations. Further research using classical conjunction visual search paradigm types with head up symbology stimuli may provide valuable insights in this regard. Other issues related to HUD and HMD visual processing, should also be considered in future higher fidelity research; such as depth of field, focal distance, visual acuity, eye box and off-axis viewing. Therefore, there is a need for higher fidelity experiments using collimated head up symbology imagery that can directly evaluate colour influence on human visual perception of symbols and information presentation [54].

5. Conclusion and Recommendations

The main finding of this study was that head up symbology colour coding reduced the workload experienced by professional commercial pilots during a high task load manual flying task. Furthermore, we built upon our previous published findings related to the performance benefits of colour coded head up symbology. Specifically, that these manual flying performance benefits may be restricted to just low task load scenarios. In this context, such a finding has not been previously reported. The results of this study provide an encouraging basis for the future review of FAA HUD design guidelines regarding the design and development of colour HUD implementation (FAA Advisory circular: 25-11B [17]). Pertinent questions still remain regarding where and how HUD colour coding would be best utilised on the flight deck. In particular, the benefit of HUD colour coding needs to be examined during flight conditions that place emphasis on divided visual attention and planning, for example, detecting traffic and/or weather hazards.

Acknowledgements

This article is based on work performed in the programme: Open Flight Deck, which has received funding from Innovate UK programme, the UK's Innovation Agency, with support from the UK Aerospace Technology Institute (grant reference number 113108). The views and opinions expressed in this paper are those of the authors and are not intended to represent the position or opinions of the Open Flight Deck consortium or any of the individual partner organisations.

References

[1] M.R. Endsley, Design and Evaluation for Situation Awareness Enhancement, Proc. Hum. Factors Soc. Annu. Meet. 32 (1988) 97–101. https://doi.org/10.1177/154193128803200221.

[2] H. Frey, P. König, What ' s color got to do with it ? The in fl uence of color on visual attention in different categories, 8 (2018) 1–17. https://doi.org/10.1167/8.14.6.Introduction.

[3] L. Itti, C. Koch, COMPUTATIONAL MODELLING OF VISUAL ATTENTION, 2001. www.nature.com/reviews/neuro (accessed November 13, 2019).

[4] C. Bundesen, L.F. Pedersen, Color segregation and visual search, Percept. Psychophys. 33 (1983) 487–493. https://doi.org/10.3758/BF03202901.

[5] R.C. Carter, Visual search with color., J. Exp. Psychol. Hum. Percept. Perform. 8 (1982) 127–136. https://doi.org/10.1037/0096-1523.8.1.127.

[6] E.W. Farmer, R.M. Taylor, Visual search through color displays: Effects of target-background similarity and background uniformity, Percept. Psychophys. 27 (1980) 267–272. https://doi.org/10.3758/BF03204265.

[7] B.F. Green, L.K. Anderson, Color coding in a visual search task, J. Exp. Psychol. 51 (1956) 19–24. https://doi.org/10.1037/h0047484.

[8] S.L. Smith, Color coding and visual separability in information displays, J. Appl. Psychol. 47 (1963) 358–364. https://doi.org/10.1037/h0041203.

[9] A. Treisman, Perceptual grouping and attention in visual search for features and for objects, J. Exp. Psychol. Hum. Percept. Perform. 8 (1982) 194–214. https://doi.org/10.1037/0096-1523.8.2.194.

[10] A.L. Nagy, R.R. Sanchez, T.C. Hughes, Visual search for color differences with foveal and peripheral vision, J. Opt. Soc. Am. A. 7 (1990) 1995. https://doi.org/10.1364/josaa.7.001995.

[11] A.L. Nagy, R.R. Sanchez, Critical color differences determined with a visual search task, J. Opt. Soc. Am. A. 7 (1990) 1209. https://doi.org/10.1364/josaa.7.001209.

[12] R.W. Backs, L.C. Walrath, Ocular measures of redundancy gain during visual search of colour symbolic displays, Ergonomics. 38 (1995) 1831–1840. https://doi.org/10.1080/00140139508925230.

[13] N. Dahlstrom, S. Dekker, R. Van Winsen, J. Nyce, Fidelity and validity of simulator training, Theor. Issues Ergon. Sci. 10 (2009) 305–314.

[14] K.F. van Orden, J. Divita, M.J. Shim, Redundant Use of Luminance and Flashing with Shape and Color as Highlighting Codes in Symbolic Displays, Hum. Factors. 35 (1993) 195–204. https://doi.org/10.1177/001872089303500201.

[15] G. Derefeldt, Ö. Skinnars, J. Alfredson, L. Eriksson, P. Andersson, J. Westlund, U. Berggrund, J. Holmberg, R. Santesson, Improvement of tactical situation awareness with colour-coded horizontal-situation displays in combat aircraft, Displays. 20 (1999) 171–184. https://doi.org/https://doi.org/10.1016/S0141-9382(99)00022-0.

[16] S. DeMars, Human factors considerations for the use of color in display systems, Sci. Tech. Aerosp. Reports. (1975) 36.

[17] F.A. Administration, Advisory Circular AC 25-11B Transport Category Airplane Electronic Flight

Displays., Washington, DC US Dep. Transp. (2014).

[18] G.L. Martinsen, P.R. Havig, Luminance Contrast Requirements for Colored Symbols in Helmet-Mounted Displays, Proc. SPIE. 4711 (2002).

[19] P.R. Havig, S.S. Grigsby, E.L. Heft, J.M. LaCreta, D.L. Post, Luminance contrast and color recognition in helmet-mounted displays, 4361 (2001) 198–204. https://doi.org/10.1117/12.2274736.

[20] C. Bigler, P. Blanche, K. Sarma, Holographic waveguide heads-up display for longitudinal image magnification and pupil expansion, Appl. Opt. 57 (2018).

[21] D.L. Post, E.E. Geiselman, C.D. Goodyear, Benefits of Color Coding Weapons Symbology for an Airborne Helmet-Mounted Display, Hum. Factors J. Hum. Factors Ergon. Soc. 41 (1999) 515–523. https://doi.org/10.1518/001872099779656798.

[22] H.J. Dudfield, Colour head-up displays: help or hindrance ?, (1991) 146–150.

[23] J. Blundell, S. Scott, D. Harris, J. Huddlestone, D. Richards, With flying colours: Pilot performance with colour-coded head-up flight symbology, Displays. 61 (2020) 101932. https://doi.org/10.1016/j.displa.2019.101932.

[24] H. Mansikka, K. Virtanen, D. Harris, Dissociation between mental workload, performance, and task awareness in pilots of high performance aircraft, IEEE Trans. Human-Machine Syst. 49 (2019) 1–9. https://doi.org/10.1109/THMS.2018.2874186.

[25] Y.-Y. Yeh, C.D. Wickens, Dissociation of Performance and Subjective Measures of Workload, Hum. Factors J. Hum. Factors Ergon. Soc. 30 (1988) 111–120. https://doi.org/10.1177/001872088803000110.

[26] L. Damodaran, User involvement in the systems design process-a practical guide for users, Behav. Inf. Technol. 15 (1996) 363–377. https://doi.org/10.1080/014492996120049.

[27] M. Maguire, Context of Use within usability activities, Int. J. Human-Computer Stud. 55 (2001) 453–483. https://doi.org/10.1006/ijhc.2001.0486.

[28] S.G. Hart, L.E. Staveland, Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research Sandra, Adv. Psychol. 52 (1988) 139–183.

[29] R.M. Taylor, Situational Awareness Rating Technique (SART): The Development of a Tool for Aircrew Systems Design., in: Situational Aware. Aerosp. Oper., 1990: pp. 111–127.

[30] S.G. Hart, NASA-TASK LOAD INDEX (NASA-TLX); 20 YEARS LATER, (2006) 904–908.

[31] D.H. Brainard, The Psychophysics Toolbox, Spat. Vis. 10 (1997) 433–436. https://doi.org/10.1163/156856897X00357.

[32] D.G. Pelli, The VideoToolbox software for visual psychophysics: transforming numbers into movies., Spat. Vis. 10 (1997) 437–42. http://www.ncbi.nlm.nih.gov/pubmed/9176953 (accessed September 17, 2019).

[33] M. Kleiner, D. Brainard, D. Pelli, A. Ingling, R. Murray, C. Broussard, Perception., [Pion Ltd.], 2007. https://nyuscholars.nyu.edu/en/publications/whats-new-in-psychtoolbox-3 (accessed September 17, 2019).

[34] D. Graziotin, X. Wang, P. Abrahamsson, Do feelings matter? On the correlation of affects and the self-assessed productivity in software engineering, J. Softw. Evol. Process. 27 (2015) 467–487. https://doi.org/10.1002/smr.1673.

[35] N.M. Laird, J.H. Ware, Random-Effects Models for Longitudinal Data, Biometrics. 38 (1982) 963–974. https://doi.org/10.2307/2529876.

[36] R. Gueorguieva, J.H. Krystal, Move over ANOVA: progress in analyzing repeated-measures data and its reflection in papers published in the Archives of General Psychiatry., Arch. Gen. Psychiatry. 61 (2004) 310–317. https://doi.org/10.1001/archpsyc.61.3.310.

[37] B. Winter, S. Grawunder, The phonetic profile of Korean formal and informal speech registers, J. Phon. 40 (2012) 808–815. https://doi.org/10.1016/j.wocn.2010.11.010.

[38] M.R. Endsley, D.B. Kaber, Level of automation effects on performance, situation awareness and workload in a dynamic control task., Ergonomics. 42 (1999) 462–492. https://doi.org/10.1080/001401399185595.

[39] M.R. Endsley, E.O. Kiris, The Out-of-the-Loop Performance Problem and Level of Control in Automation, Hum. Factors J. Hum. Factors Ergon. Soc. 37 (1995) 381–394. https://doi.org/10.1518/001872095779064555.

[40] M. Endsley, D. Jones, Designing for Situation Awareness: An Approach to User-Centered Design, Second Edi, Taylor & Francis Group, Boca Raton, 2016. https://doi.org/10.1080/10686967.2017.11918512.

[41] A.M. Treisman, G. Gelade, A feature-integration theory of attention., Cogn. Psychol. 12 (1980) 97–136. https://doi.org/10.1016/0010-0285(80)90005-5.

[42] U. Leonards, S. Sunaert, P. Van Hecke, G. a Orban, Attention mechanisms in visual search -- an fMRI study., J. Cogn. Neurosci. 12 Suppl 2 (2000) 61–75. https://doi.org/10.1162/089892900564073.

[43] J.M. Wolfe, Visual Attention, in: Seeing, CA: Academic Press., San Diego, 2000: pp. 335–386.

[44] J. Blundell, S. Frisson, A. Chakrapani, S. Kearney, S. Vijay, A. Macdonald, P. Gissen, C. Hendriksz, Markers of cognitive function in individuals with metabolic disease : Morquio syndrome and tyrosinemia type III, 3294 (2018). https://doi.org/10.1080/02643294.2018.1443913.

[45] K. Nakayama, G. Silverman, Serial and Parallel processing of visual feature conjunctions, Nature. 320 (1986).

[46] M.R. Endsley, Measurement of Situation Awareness in Dynamic Systems, Hum. Factors J. Hum. Factors Ergon. Soc. 37 (1995) 65–84. https://doi.org/10.1518/001872095779049499.

[47] T. Nguyen, C.P. Lim, N.D. Nguyen, L. Gordon-Brown, S. Nahavandi, A Review of Situation Awareness Assessment Approaches in Aviation Environments, IEEE Syst. J. 13 (2019) 3590–3603. https://doi.org/10.1109/jsyst.2019.2918283.

[48] L. Silverstein, Human factors for color display systems: concepts, methods, and research, in: H. Durrett (Ed.), Color Comput., Academic Press Professional, San Diego, 1987: pp. 27–61.

[49] R.L. Charles, J. Nixon, Measuring mental workload using physiological measures: A systematic review, Appl. Ergon. 74 (2019) 221–232. https://doi.org/10.1016/j.apergo.2018.08.028.

[50] G. Borghini, P. Aricò, G. DI Flumeri, G. Cartocci, A. Colosimo, S. Bonelli, A. Golfetti, J.P. Imbert, G. Granger, R. Benhacene, S. Pozzi, F. Babiloni, EEG-Based Cognitive Control Behaviour Assessment: An Ecological study with Professional Air Traffic Controllers, Sci. Rep. 7 (2017). https://doi.org/10.1038/s41598-017-00633-7.

[51] H. Mansikka, P. Simola, K. Virtanen, D. Harris, L. Oksama, Fighter pilots' heart rate, heart rate

variation and performance during instrument approaches, Ergonomics. 59 (2016) 1344–1352. https://doi.org/10.1080/00140139.2015.1136699.

[52] G. Borghini, P. Aricò, G. Di Flumeri, S. Salinari, A. Colosimo, S. Bonelli, L. Napoletano, A. Ferreira, F. Babiloni, Avionic technology testing by using a cognitive neurometric index: A study with professional helicopter pilots, in: Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS, Institute of Electrical and Electronics Engineers Inc., 2015: pp. 6182–6185. https://doi.org/10.1109/EMBC.2015.7319804.

[53] S. McFadden, R. Kaufmann, M. Janzen, Classification of colours presented against different coloured backgrounds, Displays. 15 (1994) 203–214. https://doi.org/https://doi.org/10.1016/0141-9382(94)90069-8.

[54] R. Patterson, M.D. Winterbottom, B.J. Pierce, Perceptual issues in the use of head-mounted visual displays, Hum. Factors. 48 (2006) 555–573. https://doi.org/10.1518/001872006778606877.

Figure 1. Examples of the HUD symbology format and colour coding implementation: A) Straight and level flight profile requirements within limits (HUD symbology all green); B) Minor indicated airspeed limit breach (amber IAS tape) during climbing; C) Excessive altitude limit breach (red altitude tape) during straight and level; D) Minor localiser deviation, 1 dot (amber ILS localiser), during ILS approach.

Figure 2: Mean overall NASA-TLX scores grouped by HUD colour coding and scenario task load condition. Standard deviations shown as error bars.

Figure 3: Mean bank angle (deg) root mean square error (RMSE) performance Low task load results. RMSE standard deviations represented as error bars.

Table 1: Details of flight profile colour feedback behaviours

Manoeuvre	Description	Red	Amber	Green	Target	Green	Amber	Red
IAS (KNTS)								
S&L	Straight and level	< 165	165 - 167	> 167	170	< 180	181 – 190	> 190
Descend	Straight decent for 1000ft	< 165	165 - 167	> 167	170	< 180	181 – 190	> 190
Flat Turn	Flat left 130° turn	< 165	165 - 167	> 167	170	< 180	181 – 190	> 190
Climbing Turn	Climbing right 190° turn for 2000ft	< 165	165 - 167	> 167	170	< 180	181 – 190	> 190
Final ILS	Stabilised approach from 1000ft	< 85	85 - 87	> 87	90	< 100	101 – 110	> 110
BNK (Deg)								
S&L	Straight and level	< 10	10 - 5	> 5	0	< 5	5 – 10	> 10
Descend	Straight decent for 1000ft	< 10	10 - 5	> 5	0	< 5	5 – 10	> 10
Flat Turn	Flat left 130° turn	< 30	25 – 30	> 15	20	< 25	25 – 30	> 30
Climbing Turn	Climbing right 190° turn for 2000ft	< 30	25 – 30	> 15	20	< 25	25 – 30	> 30
ALT (ft)								
S&L	Straight and level	< 3940	3940 - 3950	> 3950	4000	< 4050	4050 – 4060	> 4060
Flat Turn	Flat left 130° turn	< 3940	3940 - 3950	> 3950	4000	< 4050	4050 – 4060	> 4060
ILS (Dots)								
ILS Approach	ILS intercept to 1000ft above runway	< -2	-2 - -1	> -1	0	< +1	+ 1 – 2	> +2

Flight Profile Limts (columns Red–Red)

Final ILS	Stabilised approach from 1000ft	< -2	-2 - -1	> -1	0	< +1	+ 1 – 2	> +2

Table 4. Primary and secondary task performance descriptive results. Means presented along with standard deviations within parentheses.

	IAS (kts)		BNK (deg)		ALT (ft)		ILS (dots)		N-back Error (%)	
High Task Load	*Mono*	*Colour*	*Mono*	*Colour*	*Mono*	*Colour*	*Mono*	*Colour*	*Mono*	*Colour*
S&L	3.8 (1.4)	3.8 (1.9)	1.1 (0.6)	1.3 (0.7)	59.8 (38.8)	32.9 (16.6)	-	-	22.0 (14.8)	23.9 (13.6)
Flat Turn	3.5 (3.1)	2.4 (1.4)	2.5 (1.1)	2.9 (1.2)	38.2 (23.4)	37.2 (15.2)	-	-	27.7 (21.0)	22.6 (11.8)
Descend	5.1 (3.8)	4.3 (3.0)	0.5 (0.4)	1.1 (0.6)	-	-	-	-	22.5 (9.3)	23.9 (18.2)
CT	9.1 (6.9)	6.5 (9.2)	3.5 (1.5)	3.1 (1.4)	-	-	-	-	24.3 (13.9)	27.1 (10.3)
ILS Approach	-	-	-	-	-	-	0.6 (0.2)	0.6 (0.2)	38.5 (21.3)	37.6 (22.2)
ILS Final	15.7(13.8)	10.4 (7.7)	-	-	-	-	0.6 (0.2)	0.6 (0.2)	29.5 (16.0)	26.7 (6.3)
Low Task Load	*Mono*	*Colour*	*Mono*	*Colour*	*Mono*	*Colour*	*Mono*	*Colour*	*Mono*	*Colour*
S&L	2.8 (1.6)	2.8 (2.0)	0.8 (0.5)	0.9 (0.3)	33.3 (21.5)	19.3 (9.5)	-	-	-	-
Flat Turn	2.5 (2.1)	2.6 (2.7)	3.8 (1.9)	2.6 (1.1)	25.8 (13.4)	31.1 (18.0)	-	-	-	-
Descend	3.3 (1.8)	4.1 (2.6)	0.7 (0.3)	0.8 (0.7)	-	-	-	-	-	-
CT	4.2 (2.6)	3.2 (1.6)	4.4 (1.9)	2.8 (0.9)	-	-	-	-	-	
ILS Approach	-	-	-	-	-	-	0.5 (0.2)	0.5 (0.2)	-	-
ILS Final	16.3 (9.4)	10.7 (4.2)	-	-	-	-	0.5 (0.2)	0.6 (0.3)	-	-

Note: Straight and level (S&L) ; Climbing Turn (CL); N/A data (-)

Appendices

Appendix 1: Symbology usability survey

1. How did you find using the head-up symbology (useful / difficult)?

(Manoeuvre / element / Events / Other / Changes)

2. How did you find the use of colour as a cue in the symbology?

(Manoeuvre / element / Events / Other / Changes)

3. How representative did you find the flight you flew?

(Manoeuvre / element / Events / Other / Changes)

4. How did difficult did you find the flight scenario you flew?

(Manoeuvre / element / Events / Other / Changes)