

With Flying Colours: Pilot Performance with Colour-Coded Head-Up Flight Symbology

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

Author post-print (accepted) deposited by Coventry University's Repository

Original citation & hyperlink:

Blundell, J, Scott, S, Harris, D, Huddleston, J & Richards, D 2020, 'With Flying Colours: Pilot Performance with Colour-Coded Head-Up Flight Symbology' *Displays*, vol. 61, 101932.
<https://dx.doi.org/10.1016/j.displa.2019.101932>

DOI 10.1016/j.displa.2019.101932

ISSN 0141-9382

ESSN 1872-7387

Publisher: Elsevier

NOTICE: this is the author's version of a work that was accepted for publication in *Displays*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Displays*, 61, (2020) DOI: 10.1016/j.displa.2019.101932

© 2020, Elsevier. Licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Copyright © and Moral Rights are retained by the author(s) and/ or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This item cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder(s). The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

This document is the author's post-print version, incorporating any revisions agreed during the peer-review process. Some differences between the published version and this version may remain and you are advised to consult the published version if you wish to cite from it.

With Flying Colours: Pilot performance with Colour-Coded Head-Up Flight Symbolology

Authors:

Corresponding Author: James Blundell¹ (ac7045@coventry.ac.uk),

Co-Authors: Steve Scott¹ (ab8575@coventry.ac.uk), Don Harris¹ (ab3693@coventry.ac.uk), John Huddleston¹ (ab4919@coventry.ac.uk), Dale Richards¹ (ab3204@coventry.ac.uk)

¹ Centre for Mobility and Transport, Coventry University, Coventry, UK

Abstract

The manipulation of colour in display symbology design has long been recognised as a method to improve operator experience and performance. Recent developments in colour head-up display (HUD) and helmet-mounted display (HMD) technology underline the necessity to understand the human factors considerations of symbology colour coding against conventional monochrome symbology formats. In this low-fidelity desktop human-in-the-loop experiment, the colour of flight symbology on an overlaid symbology set was coded as a redundant cue to indicate the accuracy of professional and non-professional pilots' flight profile across a range of simulated flight manoeuvres. The main finding of this study was that colour coding flight symbology supported the manual flying performance of both professional and non-professional pilots. Notably, colour-coding of the bank indicator and airspeed tape minimised performance error during turning and altitude change manoeuvres, respectively. The usability of colour coded symbology was also rated higher than the monochrome symbology. We conclude that colour coded HUD/HMD symbology is preferred by the user and may improve performance during low workload manual flying tasks. A fuller understanding of performance and workload effects will require future studies to employ higher workload flying tasks and examine the utility of colour coding within higher fidelity environments.

Keywords:

Head-Up-Displays, Head-Mounted-Displays, Colour, Aviation, 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

The conscious attentional effort required to locate and scrutinise target information in a busy visual scene can be minimised by pre-attentive processing when the target stimuli differs from noncritical information on a single dimension [1–5]. When pre-attentive processing is successful in this manner the target should “pop-out” of the display [6]. An example would be the search of a single red item among a set of green distractor items. In this case, the red target pops out and summons attention with minimal interference from the green distractors. When symbology varies along multiple dimensions (e.g. colour and shape), pre-attentive processing may be able to isolate a group of likely target candidates based on one dimension (e.g. colour), but then explicit attentional resources are required to guide attention over the reduced symbology set (e.g. shapes of a specific colour). Evidently, performance will be much slower in this instance since explicit attention is required to scrutinise individual items within the reduced symbology set in serial fashion until the target is located [7–9]. Nevertheless, performance is still superior to the case where there is no colour coding. Because it is so effective, selection by colour is a common dimension of symbology that has been manipulated in the design of visual displays to improve operator experience and performance [10–12].

The above principles have been successfully applied to the design of cockpit displays to improve the communication of safety critical information for more than 75 years [13]. The Federal Aviation Administration (FAA) advises that in addition to utilising visually distinct colour sets, colours in electronic flight displays must be employed only as a redundant cue and be semantically standardised (FAA Advisory circular: 25-11B [14]). For example, the progression from green to red is commonly used to semantically convey increasing degrees of threat, a potential hazard, safety criticality, or the need for flight crew awareness and/or response. There is strong body evidence that has highlighted that colour can be used to support cognitive functions, improve pilot spatial orientation, enhance accuracy, decision time and workload [13,15–17]. In the military domain,

colour has been used to support the identification of targets, smoke, flags, signal and navigation lights, and terrain differences [18]. In the commercial world, colour coding has been employed within TCAS modes of head-down navigational displays (ND) to support discrimination between TCAS proximity cautions and alerts. Future avionic applications such as the Airborne Separation Assurance System (ASAS) will utilise colour within head-down Primary Flight Displays (PFD) and ND to support pilots maintain self-assured separation during free-routing operations [19,20]. Specifically, yellow or orange coloured “no-go” bands placed on the vertical speed tape and heading rose would represent potential conflicts between 3 and 6 minutes away, respectively.

Head-up displays (HUDs) and Helmet-Mounted Displays (HMDs) allow pilots to see key flight instrumentation on a transparent display whilst maintaining their view of the outside world. To achieve this both technologies optimally superimposed the symbology of the transparent display onto the user’s field of view. This collation of near and far flight information removes the need to look down at the flight instruments, resulting in increased situational awareness and greater precision in aircraft control [21]. HUD and HMD imagery is often restricted to monochrome (green) as a consequence of the single P-53 phosphor that is used to generate the imagery [22]. This results in the omission of information normally provided, or organised, by colour coding. However, recent advancements in waveguide optical technology means that the development of colour HUDs could be viable in the near future [23]. The display technology in HMDs is different and the development and design of colour displays has matured further in comparison to HUDs. For example, the United Force Airforce (USAF) has addressed several relevant HMD visual processing issues such as the appropriate luminance contrasts ratios for a colour HMD [24,25]. Nonetheless, due to their complexity and high cost, colour HUDs and HMDs have been late in development [26]. Consequently, the related human factors considerations of colour have been largely ignored. This is reflected in the absence of specific colour guidance from the FAA regarding the presentation of information on HUDs (FAA Advisory circular: 25-11B (Federal Aviation Administration, 2014)). However, these factors have not decreased their desirability to the user [16]. With the possibility to

develop more visually distinct and complex head-up and head-mounted display systems in the near future an understanding of relevant human factors has become more urgent.

Several studies have confirmed the positive impact of colour on flight performance and operations. DeMars (1975) concluded that, for certain applications, colour enhanced accuracy, decision time, and workload capability. In a study by Derefeldt et al (1999), an upgraded military colour coded head-down display was discovered to provide more target search and tracking advantages than the earlier monochrome display. Furthermore, the colour displays reduced reaction times and helped pilots to see the grouping of information on the display. Similarly, colour-coding weapon symbology of military pilot HMDs can reduce missile release time without sacrificing probability of kill [28]. Conversely, Dudfield (1991) found that the performance benefits of colour-coded flight symbology on a HUD far outweighed its perceived importance. However, the pilots in Dudfield's study noted that the difficulty of the employed task, maintenance of a straight-level profile, was not sufficiently challenging.

The intention of the current study is to evaluate the performance and workload benefits of a colour coded head-up flight symbology set. Rather than creating a physical HUD or HMD platform we decided to present an artificial overlay on a computer screen that would in essence create an "artificial HUD". This of course creates an offset in terms of visual acuity and human performance, but the primary focus of this study was to examine the cognitive effects of colour (and not assess the focal demands of the display). Manual flying performance and subjective workload of professional commercial pilots and non-professional pilots was examined in response to flying with a redundantly colour coded flight symbology set across a range of low-fidelity simulated flight trials. Similar to Dudfield (1991), symbology colour coding cues were based on economy so that colour was used only when participants flew outside pre-determined boundaries, e.g. flying off course, providing the subject with immediate feedback on the accuracy of their performance. However, we expanded on Dudfield's study by evaluating the use of colour feedback across several flight manoeuvre types

(ranging in complexity), not just straight-and-level flight. The use of colour in display design has been used frequently to facilitate learning [30], the inclusion of inexperienced participants in the current study was to determine whether the availability of the colour-coded redundant information served a cognitive purpose beyond facilitating the learning process in a novice group. In addition, subjective measures of workload and usability were measured via the NASA Task Load Index (TLX) and the Post-Study System Usability Questionnaire (PSSUQ), respectively.

2. Materials and methods:

2.1 Participants

Nine professional commercial airline pilots and eighteen non-professional pilots participated in the study. The rank of professional pilot participants included three first officers and six captains. Pilots' average flying experience was 5,550 hours (SD = 4,149). Non-pilots consisted of volunteer aviation / automotive engineering students and staff from the Faculty of Mobility and Transport at Coventry University. The experiment was approved by Coventry University Ethics and was in line with Ethical guidelines as per the British Psychological Society.

2.2 Experimental Design and Apparatus

The purpose of the study was to investigate systematic individual differences between professional (pilots) and non-professional pilots (non-pilots) in terms of flight performance and workload when flying a manual flying simulator task with either a colour or monochrome head-up symbology set. A desktop simulator running X-plane 9.71 (Laminar Research) was used for the manual flying task. Participants were seated at a viewing distance of 75 cm from a 55-by-40 cm display, produced a total field of view of 40.3 deg. The simulation required the participants to fly a very light single-engine jet aircraft (the Cirrus Vision SF50) in a 20-minute manual flying task involving a take-off, 4 separate manoeuvres (Table 1), an Instrument Landing System (ILS) interception and landing.

Table 1: Flight Manoeuvres Details

Manoeuvre	Description	Flight Parameters Evaluated			
		Alt (ft)	IAS (knts)	Hdg (deg)	Bnk (deg)
S&L	Straight and level	4000	170	260	0
Descend	Straight decent for 1000ft	-	170	130	0
Flat Turn	Flat left 130° turn	4000	170	-	20
Climbing Turn	Climbing right 190° turn for 2000ft	-	170	-	20

Figure 1 presents the symbology that was overlaid onto the simulation's visual scene, highlighting how the colour coding of the symbology was implemented. Task complexity was increased by omitting several information elements (i.e. the velocity vector, vertical airspeed), promoting synthesis of disparate information sources to accomplish specific tasks. For example, maintaining a steady vertical airspeed required attending to altitude and airspeed information simultaneously.

Colour was a changeable independent variable, with the symbology feature/item colour used to indicate the accuracy of the flight profile. The colour coding conventions were aligned with the FAA's guidance on colour coding within electronic flight displays (Advisory circular: 25-11B (Federal Aviation Administration, 2014)). Whereby, the progression from green to red represented the increasing need for flight crew awareness and response. If a participant was to veer off course, fly too low or high, too fast or slow, or with a high degree of roll, the relevant item would change from green to amber and ultimately to red, depending upon the amount of deviance from a set of flight profile criteria (Table 1). For example, if a subject flew too low or high then the altitude tape would turn to amber at +/-50ft and red at +/-60ft. The limits in which colour changes occurred were pre-defined. The symbology colour was manipulated in MATLAB® (version R2018b) in real time, using user datagram protocol (UDP) to send/retrieve relevant flight parameter data between X-plane and MATLAB®.

Figure 1 about here

The experimental design was a 2x4x2 mixed factorial design. Within subject factors included type of symbology (*Colour*: colour or monochrome) and type of flight manoeuvre (*Manoeuvre*; Table 1). The between subject factor was the experience of the participant group (*Group*: pilot or non-pilot). Performance was measured as the root mean square error (RMSE) data of the deviation from a given manoeuvre's required altitude, indicated air speed, heading and bank angle. The order participants experienced the two *Colour* conditions were counterbalanced across participants.

2.3 Subjective measurements

Subjective measures of workload and usability were measured after each 20-minute scenario via the NASA Task Load Index (TLX)[31] and the Post-Study System Usability Questionnaire (PSSUQ)[32], respectively. For the TLX, an overall workload score was calculated for each participant by averaging workload rating across the 6 TLX workload dimensions (scale 0 – 20). Higher TLX rating represented higher subjective workload. This approach was chosen, in contrast to the alternative dimension weighting method, due to the inconclusive evidence that dimension weighting improves the TLX's sensitivity [33]. To aid interpretation of the PSSUQ the valence of scores were reversed so that higher PSSUQ scores reflected higher rating of subjective usability. In addition, participants were invited to provide responses to a set of four open-ended questions related to the usability of the symbology at the end of the experiment (Appendix 1).

2.4 Data Analysis:

Flight performance data (Indicated Airspeed (IAS), Bank Angle (BNK), Altitude (ALT)) was fitted using linear mixed effects models (LME), using the MATLAB Statistical Toolbox) with fixed factors for *Manoeuvre*, *Group* and *Colour*. Random and fixed effects were selected using Akaike

Information Criterion (AIC) following the method described by Diggle (2013). AIC is preferred for model comparison because, unlike p-values, AIC balances fit and the number of model parameters when choosing models [35] (see Appendix 2 for further notes).

The set of models compared were similar to those found in traditional analyses (e.g., analysis of variance, ANOVA), whereby initial maximal, generating models consisting of the 2-way interaction *Group* × *Manoeuvre* × *Colour* (this model will also contain all lower level one-way interactions and main effects). This model is compared to a set of simpler models that have one or more effects removed. If a simpler model accounts for data as well as a complex model (e.g. AIC difference between models is smaller than 2), it can be assumed that the missing effects were not important. Hence, a final model should only include effects that are needed to account for the data. The highest order model for each analysis, the one reduced models are derived from, is listed as the generating model along with the model results that are provided in Table 2 and 3. Respective random effects structures for each analysis are also listed here. The list of model results includes the models closest to the minimum AIC model. Models not listed were worse and provided poor accounts of the data.

3. Results

3.1 Performance Results

3.1.1 Indicated Air Speed (IAS)

Figure 2 about here

Mean RMSE IAS performance, with standard deviations, is presented in Figure 2. For non-pilots, mean IAS RMSE clearly decreased when colour feedback was provided (RMSE mean difference: Straight and Level (S&L) = 0.14 ; Climbing Turn = 0.86 ; Flat Turn = 1.04 ; Descent = 1.74). Similarly, mean IAS RMSE was lower with colour feedback for pilots, but to a lesser extent, on S&L (RMSE mean difference: = 0.26), climbing turns (0.25) and descents (0.44). Comparison of AIC values of main effect and interaction models revealed the best model (AIC = 0) of IAS performance included separate main effects for *Colour*, *Manoeuvre* and *Group* (Table 2). Evidence for models without any 3 main effects was weak (omitted *Colour* Δ AIC = 5.3 ; *Manoeuvre* omitted Δ AIC = 9.8 ; *Group* omitted Δ AIC = 6.3). The main effect of *Group* represented a difference of 2.95 (CI: 1.16 - 4.74) IAS RMSE between pilots and non-pilots. The main effect of *Colour* reflected an improvement of 1.04 (CI: -0.29 - 1.78) in IAS RMSE when pilots and non-pilots flew with colour feedback across the four flying conditions.

Evaluation of the systematic differences within *Manoeuvre* revealed that the main effect was best explained by an increase in IAS RMSE by 2.73 (CI: 1.89- 3.58) during manoeuvres requiring a change in altitude – Climbing Turn and Descent. There was some evidence for the inclusion of an interaction between *Group* and *HUD* (Δ AIC = 0.6) and weak evidence of an interaction between *Colour* and *Manoeuvre* (Δ AIC = 4.4).

3.1.2 Bank Angle (BNK)

Figure 3 about here

Clear differences were observed in mean bank angle (BNK) RMSE between the four manoeuvres. In addition, the presence of colour feedback had a greater influence on BNK RMSE when a change of heading was required for both pilots and non-pilots (BNK means and standard deviations shown in Figure 3). The best model of BNK performance was the maximal 2-way interaction model (Table 2). However, equivalent evidence was found for a simpler model that included the interaction for *Colour* and *Manoeuvre*, and a main effect for *Group* ($\Delta AIC = 1.1$). The main effect of *Group* revealed that, compared to non-pilots, pilot bank angle deviations were lower overall by 1.6 degrees RMSE (CI: 0.91-2.30).

Model reduction procedures (post-hoc) revealed that the interaction model could be considerably improved ($\Delta AIC = 3.1$) by collapsing *Manoeuvre* into a 2-level factor, according to whether a manoeuvre required a heading change. The final model highlighted that manoeuvres with a heading change increased BNK RMSE by 6.71 deg (CI: 6.05-7.37), and that colour feedback supported pilots and non-pilots reduce this error by 1.85 deg.

Table 2: Performance LME AIC model results

Model	Param N	AIC	Δ AIC	AICw
<i>Indicated AirSpeed (IAS)</i>				
Generating Model: Colour \times Manoeuvre \times Group				
Random Effects structure: (1 + Manoeuvre Participant)				
Colour + Manoeuvre + Group	6	1762.3	0.0	0.42
Colour \times Group + Manoeuvre	7	1762.9	0.6	0.31
Colour + Manoeuvre \times Group	9	1764.6	2.3	0.13
Manoeuvre \times Group + Colour	9	1766.7	4.4	0.04
<i>Bank Angle (BNK)</i>				
Generating Model: Colour \times Manoeuvre \times Group				
Random Effects structure: (1 + Manoeuvre Participant)				
Colour \times Manoeuvre \times Group	16	1363.2	0.0	0.62
Colour \times Manoeuvre + Group	9	1364.3	1.1	0.37
Colour + Manoeuvre \times Group	9	1372.0	8.8	< 0.00
<i>Altitude (ALT)</i>				
Generating Model: Colour \times Manoeuvre \times Group				
Random Effects structure: (1 + Manoeuvre Participant)				
Intercept Only	1	1830.2	0.0	0.57
Colour + Manoeuvre + Group	4	1832.6	2.4	0.17
Colour + Manoeuvre \times Group	5	1833.2	3.0	0.13

Note: Shaded rows represent “preferred” models. Strong evidence for (Δ AIC < 2) and fewest parameters.

3.1.3 Altitude (ALT)

Comparison of main effect and interaction models for altitude revealed that the best model was an intercept only model (Table 2). There was no difference in altitude deviation between pilots and non-pilots across the flight manoeuvres. Colour feedback had no effect.

3.2 Subjective Results

There was no clear difference in NASA TLX scores between colour conditions for pilots and non-pilots (Appendix 3). For TLX scores, non-pilots and pilots reported moderate (mean = 9.6) and minimal levels (mean = 4.6) of subjective workload, respectively. Model comparison results supported the above observations. The best model for NASA TLX included only a main effect for *Group* (AIC = 0.0, AICw = 0.58).

Pilot and non-pilot usability ratings of the colour coded symbology are shown in Figure 4. A preference for colour coded symbology was displayed by both pilots (colour vs. mono mean diff = 0.43) and non-pilots (mean diff = 0.38). Overall, usability scores of non-pilots were higher than pilot usability scores (mean diff = 0.57). AIC model comparisons revealed the best model to include only the main effect for *Colour* (Table 3); pilots and non-pilots rated the usability of symbology 0.36 PSUQ points (CI: 0.27 – 0.46) when colour coding was present.

Figure 4 about here

Table 3: Usability LME AIC model results

Model	Param N	AIC	ΔAIC	AICw
<i>Post-Study Usability Questionnaire (PSUQ)</i>				
Generating Model: Colour × Group				
Random Effects structure: (1 + Item Participant)				
Colour	2	2349.5	0.0	0.60
Colour + Group	3	2351.2	1.7	0.25
Colour × Group	4	2352.2	2.7	0.15

Note: Shaded rows represent “preferred” models. Strong evidence for (ΔAIC < 2) and fewest parameters.

4. Discussion

The current study aim was to investigate if benefits provided by colour in head-down displays could be extended to different types of head up displays. Specifically, we evaluated the utility of a head-up colour coded symbology set to support the manual flying performance of professional commercial pilots and non-pilots.

As shown in this human-in-the-loop study, the performance of both professional and non-professional pilots benefitted from the presence of a colour coded symbology set. Specifically, colour coding of the bank indicator and airspeed tape minimised error during turning and altitude change manoeuvres, respectively. Therefore, the use of colour served a cognitive purpose beyond only facilitating the learning process in a novice group [30]. In agreement with Dudfield (1991), colour coding did not improve either professional or non-professional pilot performance during straight and level flight. The specificity of the performance benefits associated with manoeuvres that require a change of heading and /or altitude is likely to be due to their increased complexity over straight and level flight. In contrast to a wings-level profile, pilots are required to monitor one or more dynamic flight parameters (e.g. the aircraft's current heading during a heading change) whilst maintaining other static parameters (e.g. airspeed and/or bank angle). In this context, where pilots are deploying more attentional focus to dynamic flight parameter, colour feedback may support the relocation of attention to neglected static parameters whenever their associated error margins are breached. Essentially, neglected parameters "pop-out" of the display when colour coding feedback is provided. The functional benefit of colour in this regard has been reported widely in the visual attention literature [1–5]. Indeed, in the post-study usability survey the majority of professional and non-professional pilots highlighted the value of colour in supporting the shifting of attention during instances where the integrity of their scan pattern had deteriorated.

NASA TLX scores indicated that pilots and non-pilots reported minimal and moderate levels of workload on the current task, respectively. However, while participant flight performance benefitted

from the presence of colour feedback, these benefits occurred in isolation of any reported subjective workload benefit. A possibility for this outcome is that participants were required to report their average workload experienced over the entire flight scenario. Since the colour related-performance gains were manoeuvre specific, any associated workload benefits may have been masked with this averaging approach. Alternatively, the current task did not sufficiently challenge pilots in order to produce a discernible change in experienced work load. In Dudfield's (1991) HUD colour study, workload benefits were reported for a flying task which included a secondary task. Therefore, future studies should evaluate the value of colour coding during scenarios that include a greater emphasis on visual attention and planning (e.g. taxiing, precision approach and landing), and/or include a secondary task to simulate information processing burden.

Participant subjective usability ratings revealed a preference for the colour coding format. This supports the large body of research that suggests colour enhances usability of avionic displays. For example, Kaufmann and Eaton (1994) found that the users preferred the introduction of colour coding into an established radar navigation system. Colour has also benefited systems used in visualising experiments in physics [37]. Overall, the aesthetic qualities associated with colour appear to appeal to the user, and they tend to express preferences for colour over monochromatic displays.

A consideration of how colour coding effects attentional capture is important, where attention is allocated to one source of visual information at the expense of others, to the detriment of performance [38]. In the context of HUDs and HMDs, attentional capture has been associated with instances where the pilot neglects far-domain information to attend to the more compelling near domain flight symbology. The likelihood of attentional capture occurring has been shown to increase as a function of the degree of experienced perceptual load, the likelihood of an expected event, and the salience/compelling nature of the symbology [38]. For the latter, according to feature integration theory [1–3,5], colour implementation on a HUD and HMD would substantial increase the salience of near domain symbology information. Therefore, future research will need to examine

whether the implementation of colour impacts the prevalence of the attentional capture effects. In particular, understanding how the introduction of colour-coded head-up symbology might influence the risk of missing unexpected events during high workload flight tasks, for example runway incursion detection whilst landing in high turbulence.

Notwithstanding the attentional factors mentioned above, the current low-fidelity study did not consider the impact of other aspects that are important in the visual processing of HUD and HMD symbology. Issues relating to brightness and contrast are key considerations for HUD and HMD systems as these directly influence the visual perception of the symbols and information being presented [39]. Other issues related to HUD and HMD visual processing, while outside of scope for this study, should also be considered in future higher fidelity research; such as depth of field, focal distance, visual acuity, eye box and off-axis viewing.

Colour could also have deleterious effects if overused. For example, Carter and Cahill (1979) warned against the adverse effects of the 'unrestrained use' of irrelevant colour and argued that it could interfere with the processing of other symbol dimensions (especially when this type of coding antagonised the other symbol features). Similarly, Teichner (1979) found that instead of improving performance, too much use of colour increased mental workload and the incidence of visual strain [42,43]. Further research by Nagy and Sanchez (1992) showed that search times were slower when the number of colours was increased. Indeed, McFadden, Kaufmann, and Janzen (1994) suggested that shifts in colour appearance (as a function of surround colours) impaired the accurate interpretation of information and suggested the use of limited colour combinations. This suggestion is echoed by the FAA, who advise implementing a restricted set of colours on displays in order to avoid many of the above effects (FAA Advisory circular: 25-11B [14]). Nevertheless, it will be important for future research to discern the optimal set size for colour coding head-up symbology that simultaneously maximises performance and minimises pilot workload.

5. Conclusion and Recommendations

The main finding of this study was that head-up colour coded symbology improved the performance of both professional commercial pilots and non-professional pilots during a low workload manual flying task. Specifically, colour coding of the bank indicator and airspeed tape minimised error during turning and altitude change manoeuvres, respectively. In this context, such a finding has not been previously reported. Despite the absence of a colour-related workload improvement, the results of this study provide an encouraging basis for the future review of FAA HUD design guidelines regarding the design and development of HUD/HMD colour implementation (FAA Advisory circular: 25-11B (Federal Aviation Administration, 2014)). Pertinent questions remain regarding where and how head-up colour coding would be best utilised in-flight deck operations. In particular, the benefit of colour coding needs to be examined during high workload flight conditions that place emphasis on visual attention and planning (e.g. taxiing, precision approach and landing). In addition, higher fidelity evaluations of how head-up colour coding interacts with attentional capture effects must be addressed, specifically, would colour coding promote the allocation of attentional resources to near domain information at the detriment of far domain information?

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] A.M. Treisman, G. Gelade, A feature-integration theory of attention., *Cogn. Psychol.* 12 (1980) 97–136. doi:10.1016/0010-0285(80)90005-5.
- [2] 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. doi:10.1162/089892900564073.
- [3] J.M. Wolfe, Visual Attention, in: *Seeing*, CA: Academic Press., San Diego, 2000: pp. 335–386.
- [4] J. Blundell, S. Frisson, A. Chakrapani, S. Kearney, S. Vijay, A. Macdonald, P. Gissen, C. Hendriks, Markers of cognitive function in individuals with metabolic disease : Morquio syndrome and tyrosinemia type III, 3294 (2018). doi:10.1080/02643294.2018.1443913.
- [5] K. Nakayama, G. Silverman, Serial and Parallel processing of visual feature conjunctions, *Nature.* 320 (1986).
- [6] U. Neisser, Visual search., *Sci. Am.* 210 (1964) 94–102. doi:10.1038/scientificamerican0664-94.
- [7] R.C. Carter, Visual search with color., *J. Exp. Psychol. Hum. Percept. Perform.* 8 (1982) 127–136. doi:10.1037/0096-1523.8.1.127.
- [8] M. D’Zmura, Color in visual search, *Vision Res.* 31 (1991) 951–966.
- [9] W. Schneider, R.M. Shiffrin, Controlled and automatic human information processing: I. Detection, search, and attention., *Psychol. Rev.* 84 (1977) 1–66. doi:10.1037/0033-295X.84.1.1.
- [10] R.W. Backs, L.C. Walrath, Ocular measures of redundancy gain during visual search of colour symbolic displays, *Ergonomics.* 38 (1995) 1831–1840. doi:10.1080/00140139508925230.
- [11] N. Dahlstrom, S. Dekker, R. Van Winsen, J. Nyce, Fidelity and validity of simulator training, *Theor. Issues Ergon. Sci.* 10 (2009) 305–314.
- [12] 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. doi:10.1177/001872089303500201.
- [13] J.E. Melzer, K.W. Moffitt, COLOR HELMET DISPLAY FOR THE MILITARY COCKPIT, *Proc. IEEE/AIAA 11th Digit. Avion. Syst. Conf.* (1992) 538–542.
- [14] Federal Aviation Administration, Advisory Circular AC 25-11B Transport Category Airplane Electronic Flight Displays., Washington, DC US Dep. Transp. (2014).
- [15] S. DeMars, Human factors considerations for the use of color in display systems, *Sci. Tech. Aerosp. Reports.* (1975) 36.
- [16] G. Spenkeliink, K. Besuijen, The perception and measurement of contrast: the influence of gaps between display elements, *Behav. Inf. Technol.* 13 (1994) 320–327. doi:10.1080/01449299408914612.
- [17] D. Funke, M. Lloyd, Guidelines for color coding in cockpits, *Proc. Tenth Symp. Psychol. Dep. Def.* (1986).
- [18] T.J. Tredici, D.J. Ivan, Ophthalmology in aerospace medicine, in: T. and F. Davis, Johnson (Ed.), *Fundam. Aerosp. Med.*, Lippincott Williams and Wilkins, New York, 2008.

- [19] R.C.J. Ruigrok, J.M. Hoekstra, Human factors evaluations of Free Flight. Issues solved and issues remaining, *Appl. Ergon.* 38 (2007) 437–455. doi:10.1016/j.apergo.2007.01.006.
- [20] J.M. Hoekstra, R.N.H.W. van Gent, R.C.J. Ruigrok, Designing for safety: the ‘free flight’ air traffic management concept, *Reliab. Eng. Syst. Saf.* 75 (2002) 215–232. doi:http://dx.doi.org/10.1016/S0951-8320(01)00096-5.
- [21] M. Velger, *Helmet-mounted displays and sights*, Artech House, Boston, MA, 1998.
- [22] R.J. Nichol, Airline Head-Up Display Systems: Human Factors Considerations, *Int. J. Econ. Manag. Sci.* 04 (2015). doi:10.4172/2162-6359.1000248.
- [23] C. Bigler, P. Blanche, K. Sarma, Holographic waveguide heads-up display for longitudinal image magnification and pupil expansion, *Appl. Opt.* 57 (2018).
- [24] G.L. Martinsen, P.R. Havig, Luminance Contrast Requirements for Colored Symbols in Helmet-Mounted Displays, *Proc. SPIE.* 4711 (2002).
- [25] 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. doi:10.1117/12.2274736.
- [26] J. Capo-Aponte, H. Task, Visual perception and cognitive performance, in: C. Rash, M. Russo, T. Letowski, E. Schmeisser (Eds.), *Helmet-Mounted Displays Sensation*, Percept. Cogn. Issues, U.S. Army Aeromedical Research Laboratory, AL, 2009: pp. 335–390.
- [27] 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. doi:https://doi.org/10.1016/S0141-9382(99)00022-0.
- [28] 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. doi:10.1518/001872099779656798.
- [29] H.J. Dudfield, Colour head-up displays: help or hindrance ?, (1991) 146–150.
- [30] 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.
- [31] 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.
- [32] J.R. Lewis, IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use, *Int. J. Human–Computer Interact.* 7 (1995) 57–78. doi:10.1080/10447319509526110.
- [33] S.G. Hart, NASA-TASK LOAD INDEX (NASA-TLX); 20 YEARS LATER, (2006) 904–908.
- [34] P. Diggle, ed., *Analysis of longitudinal data*, Second Pap, Oxford University Press, Oxford, 2013.
- [35] K. Burnham, D. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd ed., Springer, New York, 2002. doi:10.1002/1521-3773(20010316)40:6<9823::AID-ANIE9823>3.3.CO;2-C.
- [36] R. Kaufmann, R.M. Eaton, User trials on the colours and symbols for electronic charts, *Appl. Ergon.* 25 (1994) 366–372. doi:https://doi.org/10.1016/0003-6870(94)90055-8.
- [37] H. Drevermann, D. Travis, Visualization using colour: Visual presentation of events in particle

- physics, *Behav. Inf. Technol.* 17 (1998) 18–26. doi:10.1080/014492998119643.
- [38] G.W. Stuart, K.I. McAnally, J.W. Meehan, Head-up displays and visual attention: Integrating data and theory., *Hum. Factors Aerosp. Saf.* 1 (2001) 103–124.
- [39] R. Patterson, M.D. Winterbottom, B.J. Pierce, Perceptual issues in the use of head-mounted visual displays, *Hum. Factors.* 48 (2006) 555–573. doi:10.1518/001872006778606877.
- [40] R.C. Carter, M.-C. Cahill, Regression Models of Search Time for Color-Coded Information Displays, *Hum. Factors.* 21 (1979) 293–302. doi:10.1177/001872087902100304.
- [41] W.H. Teichner, Temporal Measurement of Input and Output Processes in Short/Term Memory, *J. Gen. Psychol.* 101 (1979) 279–305. doi:10.1080/00221309.1979.9920082.
- [42] M. Bruce, J.J. Foster, The Visibility of Colored Characters on Colored Backgrounds in Viewdata Displays, *Visible Lang.* 4 (1982) 382–390.
- [43] M.L. Matthews, The influence of colour on CRT reading performance and subjective comfort under operational conditions., *Appl. Ergon.* 18 (1987) 323–328.
- [44] A.L. Nagy, R.R. Sanchez, Chromaticity and luminance as coding dimensions in visual search., *Hum. Factors.* 34 (1992) 601–614. doi:10.1177/001872089203400507.
- [45] S. McFadden, R. Kaufmann, M. Janzen, Classification of colours presented against different coloured backgrounds, *Displays.* 15 (1994) 203–214. doi:https://doi.org/10.1016/0141-9382(94)90069-8.

Figure 1. Examples of the head-up symbology format and colour coding implementation: A) Straight and level flight profile requirements within limits (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.

Figure 2. Mean IAS RMSE grouped by flight manoeuvre, presence of colour coding feedback, and pilot group. IAS RMSE standard deviations shown as error bars.

Figure 3. Mean Bank Angle (BNK) RMSE grouped by flight manoeuvre, presence of colour coding feedback, and pilot group. BNK RMSE standard deviations shown as error bars.

Figure 4: Mean Post-Study Usability Questionnaire (PSUQ) scores by presence of colour coding feedback and pilot group. PSUQ standard deviations shown as error bars.

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 difficult did you find the flight scenario you flew?

(Manoeuvre / element / Events / Other / Changes)

Appendix 2: AIC LME analysis Notes

Model selection using AIC is different from that using p-values, but not difficult to understand. AIC is preferred for model comparison because, unlike p-values, AIC balances fit and the number of model parameters when choosing models. In brief, better models produce smaller AIC values, but the absolute AIC values are not interpretable. Instead, the change in AIC (ΔAIC) between models is meaningful and captures the weight of evidence for each model (rather than being subject to a cut-off, like p-values). Evidence for a model starts to be clear if the ΔAIC exceeds 2. If ΔAIC between the “best” model and alternative models is less than 2 then the two models are substantially equivalent. When ΔAIC is between 2 and 10 there is decreasing support for an alternative model. A model with a $\Delta AIC > 10$ has essentially no support. For models where the ΔAIC is less than 2, it is reasonable to favour the least complex model (i.e. model with fewest parameters/variables). Favoured models contain terms that are important in accounting for data. This is parallel to significant effects in an analysis using hypothesis testing. For example, if a highly rated model has a term for group but no interaction, this is parallel to a significant main effect of group and a non-significant interaction. Comparisons can be assisted by calculating Akaike weights (AICw; Burnham & Anderson, 2002). AICw expresses the relative probability that a model is the best in a particular set, considering only the models from that set. It measures the weight of evidence for the models being compared. When values are relatively equal across two or more models, they are all relatively good models of the data. If one model has a high value, and the others are low, there is a model that is clearly better.

Appendix 3: Subjective Workload and Situation Awareness Figures / Tables and AIC Model Results

Figure A3.1. Mean TLX score grouped by presence of colour coding feedback and pilot group. TLX score standard deviations shown as error bars.

Table A3.2. NASA Task Load Index (TLX) LME AIC model results

Model	Param N	AIC	Δ AIC	AICw
Generating Model: Colour \times Group				
Random Effects structure: $(1 + Item/Participant)$				
Group	2	1740.6	0.0	0.58
Colour + Group	3	1742.1	1.5	0.27
Colour \times Group	4	1743.2	2.6	0.16
Manoeuvre \times Group + Colour	1	1766.1	25.5	< 0.01

Note: Shaded rows represent "preferred" models. Strong evidence for (Δ AIC < 2) and fewest parameters.

Table A3.3. Mean inter-item NASA-TLX score grouped by presence of colour coding feedback and pilot group. NASA-TLX score standard deviations shown in parentheses.

<i>TLX Items</i>	Pilot		Non-Pilot	
	Colour	Mono	Colour	Mono
1 – Mental	5.10 (2.81)	5.67 (3.16)	13.67 (2.70)	12.78 (2.86)
2 – Physical	3.20 (2.25)	4.11 (4.31)	8.06 (4.76)	8.59 (5.10)
3 – Temporal	3.90 (2.33)	3.44 (1.51)	9.11 (3.34)	9.11 (3.71)
4 – Performance	4.40 (1.65)	4.22 (1.64)	7.50 (4.13)	7.72 (4.97)
5 – Effort	6.50 (2.64)	7.11 (2.98)	12.61 (3.88)	10.78 (4.62)
6 - Frustration	4.30 (3.83)	4.44 (2.30)	7.94 (3.62)	7.79 (3.85)

Table A3.4. Mean inter-item PSUQ score grouped by presence of colour coding feedback and pilot group. SART score standard deviations shown as error bars.

<i>PSUQ Items</i>	Pilot				Non-Pilot			
	Colour		Mono		Colour		Mono	
1 – Ease	5.56	(0.88)	5.30	(1.34)	6.47	(0.51)	5.94	(0.77)
2 – Simplicity	5.89	(0.60)	5.80	(0.79)	6.24	(0.75)	5.94	(0.77)
3 – Speed	6.00	(0.50)	5.80	(0.92)	6.13	(0.72)	5.69	(0.60)
4 – Comfort	5.67	(0.87)	5.50	(1.35)	6.12	(0.70)	6.00	(0.89)
5 – Learnability	6.00	(1.00)	6.20	(0.63)	6.12	(1.05)	6.06	(1.00)
6 – Productivity	6.00	(1.22)	5.44	(1.67)	6.12	(0.93)	5.88	(0.81)
7 – Error Message	4.63	(1.69)	3.33	(1.50)	5.56	(1.59)	4.38	(1.12)
8 – Error Correction	5.43	(0.53)	5.10	(1.29)	5.47	(1.55)	5.13	(0.96)
9 – Clear Comm	5.50	(0.53)	5.00	(1.32)	5.71	(1.36)	5.53	(0.99)
10 – Info Find	6.22	(0.67)	5.70	(1.25)	5.82	(1.24)	5.75	(0.86)
11 – Info Detail	5.89	(0.78)	5.20	(1.32)	6.18	(0.88)	5.63	(0.96)
12 – Info Organisation	5.89	(1.27)	5.30	(1.06)	6.18	(0.73)	5.56	(0.81)
13 – Interface	5.00	(1.50)	4.67	(1.50)	5.88	(1.36)	5.56	(0.89)
14 - Like	5.33	(1.12)	4.70	(1.57)	6.12	(0.93)	5.50	(1.03)
15 – Expectation	3.44	(1.59)	2.90	(1.52)	5.82	(1.38)	5.38	(1.26)
16 - Satisfaction	4.78	(1.20)	4.30	(0.95)	6.06	(0.90)	5.75	(0.86)