DOT/FAA/TC-19/31

Federal Aviation Administration
William J. Hughes Technical Center
Atlantic City International Airport, NJ 08405

Symbol Salience Augments Change-Detection Performance in Cockpit Weather Displays

Ulf Ahlstrom, AvMet Applications, Inc
Nicole Racine, DSoft Technology, Engineering, and Analysis Inc.

August 2019

Technical Report

This document is available to the U.S. public through the National Technical Information Services (NTIS), Springfield, Virginia 22161.

This document is also available from the Federal Aviation Administration William J. Hughes Technical Center at actlibrary.tc.faa.gov.

U.S. Department of Transportation
Federal Aviation Administration

NOTICE

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The United States Government assumes no liability for the contents or use thereof. The United States Government does not endorse products or manufacturers. Trade or manufacturers' names appear herein solely because they are considered essential to the objective of this report. This document does not constitute Federal Aviation Administration (FAA) certification policy. Consult your local FAA aircraft certification office as to its use.

This report is available at the FAA William J. Hughes Technical Center's full-text Technical Reports Web site: http://actlibrary.tc.faa.gov in Adobe® Acrobat® portable document format (PDF).

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/TC-19/31		

4. Title and Subtitle	5. Report Date
Symbol Salience Augments Change-Detection Performance in Cockpit Weather Displays	August 2019
	6. Performing Organization Code: ANG-E25

7. Author(s)	8. Performing Organization Report No.
Ulf Ahlstrom, AvMet Applications, Inc Nicole Racine, DSoft Technology, Engineering, and Analysis Inc.	DOT/FAA/TC-19/31

9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
Federal Aviation Administration Human Factors Branch William J. Hughes Technical Center Atlantic City International Airport, NJ 08405	11. Contract or Grant No.

12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
Federal Aviation Administration 800 Independence Avenue, S.W. Washington, DC 20591	Technical Report
	14. Sponsoring Agency Code: ANG-C61

15. Supplementary Notes

16. Abstract

The goal of the study, conducted under the Weather Technology in the Cockpit (WITC) program, is to develop Minimum Weather Service rendering recommendations to resolve/reduce the previously identified gap of change blindness with the intent of reducing/resolving the shortfall of associate safety risks. This study evaluated the effect of symbol enhancement, symbol shapes, line thickness, and age-related effects on change blindness for pilots in relation to weather displays. Ninety-seven private general aviation pilots participated in this study. The participants experienced two part-task experiments for which they had to determine whether images presented to them were the same or different. Both experiments used a one-shot technique. We found an overall effect of participant age on discrimination accuracy and an increase in response time as age increases. We found a main effect of salience for detection of the line, lightning, and METAR symbols. There was also a main effect of line color discriminability, with blue lines having the highest discriminability, followed by the red and then black lines. In general, enhancing the display symbols increases the discriminability accuracy and reduces the response time. Although this part-task study shows that enhancing symbol salience may help improve change-detection capabilities, it does not evaluate whether these enhancements would be adequate to draw attention to displays in a dynamic cockpit environment.

17. Key Words	18. Distribution Statement
Age effects. Change blindness, Change detection, Cockpit weather displays, Salience, Symbol discrimination, Weather symbols,	This document is available to the public through the National Technical Information Service, Alexandria, Virginia, 22312. A copy is retained for reference at the William J. Hughes Technical Center Library.

19. Security Classification (of this report)	20. Security Classification (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	72	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

THIS PAGE IS BLANK INTENTIONALLY.

Table of Contents

List of Illustrations

Tables **Page**

Figures **Page**

Acknowledgments

We would like to thank the personnel at the FAA William J. Hughes Technical Center Cockpit Simulation Facility for their support and for the use of their laboratory for this project. Special thanks to Vicki Ahlstrom for comments and help in editing the document. We thank Albert Rehmann, the Simulation Branch manager; Robert Kusza and Cynthia Truong for networking and experimental computer support; and Kevin Tracy for developing the change-detection program.

Executive Summary

The overarching goal of the study, conducted under the Weather Technology in the Cockpit (WITC) program, is to develop Minimum Weather Service rendering recommendations to resolve/reduce the previously identified gap of change blindness with the intent of reducing/resolving the shortfall of associated safety risks. During this study, researchers used the FAA William J. Hughes Tech Center (WJHTC) Cockpit Simulator Facility to perform two change-detection experiments. The change-detection experiments focused on pilot sensitivity to weather symbol changes in cockpit weather presentations.

Pilots may use cockpit weather displays to receive weather information updates during their flights. It is important that pilots perceive changes in the weather display because the weather changes can have safety implications for the flight. The ability for the pilot to detect changes in the display is called change detection, whereas the inability to detect any change is called change blindness. Previous research has found change blindness effects for symbol and color variations and revealed that these effects are more pronounced for older participants (age 60 and older) than younger participants (age 59 and younger). These results are important and point to several gaps in current weather display research. First, weather display developers should aim to develop display symbols that allow rapid encoding and detection. This is especially important considering the large number of weather symbols and different backgrounds used on modern weather displays. Second, as part of the symbol-evaluation process, care must be taken to ensure that the weather symbology is equally effective for all age groups. This is especially important because approximately 30% of all U.S. private pilots are 60 years of age or older (FAA, 2016).

In the present study, we addressed this symbol-detection gap by evaluating the effect of symbol enhancement on pilot visual performance. Specifically, we used symbol salience enrichments to increase pilot change-detection performance. The specific goals of the study were: to evaluate how enhancements to weather symbol salience affect pilot change-detection performance (Experiment 1); to evaluate how line segment salience affects pilot change-detection performance (Experiment 2); and to assess the presence of age-related effects on change-detection performance and response time.

Ninety-seven private general aviation pilots participated in this study. The participants experienced two part-task experiments in which they had to determine whether images presented to them were the same or different. Both experiments utilized a one-shot technique, which briefly presented two images in succession that were either the same or with enhanced symbol salience.

The result of our symbol enhancements was a main effect of symbol salience on change-detection performance. For the line, lightning, and METAR symbol trials, pilot detection performance was credibly higher for high-salience lines and credibly higher for the salience-enriched symbols (Enhanced) compared to the original symbols (Control). For time-stamp information, however, there was no credible difference in discriminability between the Enhanced and Control conditions. We conclude that the lack of an effect of salience for time-stamp information is due to the time-stamp location. The result for both the Control and the Enhanced time-stamp conditions show near-chance discriminability, similar to the result reported by Ahlstrom and Suss (2014). Currently, many commercial weather products display time-stamp

information for displayed weather data (FAA, 2010), and this information is frequently presented in the menu bar or in one of the upper corners of the display (Latorella & Chamberlain, 2002). Because this time-stamp location precludes efficient detection and encoding, display developers need to design time-stamp information that overcomes this deficiency.

There is now sufficient evidence of credible differences in pilot change-detection performance depending on the symbol shape and color. There is also evidence that the detection time and encoding time of weather symbols vary as a function of pilot age. These are important findings that should guide future weather display developments. First, all weather display symbols should allow rapid encoding and detection for pilots of all ages. This is especially important because of the large number of different weather elements that are overlaid on modern multifunctional displays (FAA, 2010) and the relatively large (~30%) population of private pilots that are 60 years of age or older (FAA, 2016). If not taken into consideration, weather symbol changes or weather symbol updates can lead to salience problems where important information fails to visually segregate from less-critical background information

We believe that the current study provides a general framework for how to enhance weather display symbols for efficient use in a multitasking cockpit environment. If display developers would use this framework, it would support GA pilots and possibly increase the efficiency and safety of operations by providing information that can be used to avoid adverse weather conditions during flight.

THIS PAGE IS BLANK INTENTIONALLY.

1 INTRODUCTION

Flights in clear weather conditions, where the visibility is clear enough for pilots to see where the aircraft is going, are called Visual Flight Rules (VFR) flights. During VFR flights, General Aviation (GA) pilots perform many tasks that are vital to aviating, navigating, and communicating. These tasks range from navigating a route, performing "see and avoid," operating the radio and navigational instruments, viewing charts and approach plates, and scanning and interpreting the flight instruments. Single-pilot operations are highly dynamic environments that require multitasking on part of the pilot.

When GA pilots experience weather situations, the level of multitasking is likely to increase. This can happen, for example, when weather conditions at the destination airport or along the route of flight impact or prevent VFR operations, but instead require instrument fly rules (IFR) operations. In these situations, pilots need to stay informed about weather changes that could affect the safety of flight as well as the tasks and scan patterns necessary for aviating, navigating, and communicating. Many GA pilots choose to maintain their weather situation awareness through use of cockpit weather displays, receiving in-flight weather updates for relevant areas along the route.

When using cockpit-mounted or handheld weather displays, pilots need to extend their visual scan pattern to include the weather display. Previous research has shown that pilot scan patterns are highly variable, but that most pilots sample visual information from the out-the-window view, the cockpit glass display, the weather presentation, and the cockpit instrument console (Ahlstrom & Dworksy, 2012). Pilots generally include at least four separate locations during their visual scan. While the pilot is scanning one of the other areas, the information on the weather display may change. On modern electronic weather displays, weather updates are done in various ways, using text, line, symbol, or color changes. The ability of humans to detect such changes is called "change detection" (Rensink, 2002). The opposite of change detection is "change blindness," a phenomenon in which humans are unable to detect changes in their field of view. Simons and Ambinder (2005) suggest that the underlying cause of change blindness is a limitation in the ability to encode, retain, and compare information from one glance to the next. This is particularly relevant for pilots because of their divided attention and visual scan pattern. Pilots' information extraction is often restricted to a summary encoding of information from one glance to the next. This means that pilots might be susceptible to change blindness, caused by a failure to encode and to compare the weather display status from only a few visual samples.

Previous research has documented change blindness effects for symbols similar to the ones used for modern digital weather displays. For example, Durlach (2004) found that operators of Army digital systems showed change blindness to icon position changes and color changes. Research has also found that change blindness is much more pronounced for background image changes than for symbol foreground changes, even when the background changes are greater than the symbol changes (Turatto, Angrilli, Mazza, Umiltà & Driver, 2002). Similarly, a recent study by Ahlstrom and Suss (2014) used flight simulations and change-detection tasks to evaluate the effect of shape and color variations in MEteorological Terminal Aviation Routine weather reports (METARs) symbols on pilots' change-detection performance. The result showed credible differences in pilot change-detection accuracy depending on the symbol shape and

color. In addition to these results, there are even more peculiar effects on change-detection performance from variations in symbol color and shape. Using color and luminance changes for rectangular symbols, Ball and Busch (2015) found that participants could detect changes that fell outside their focus of attention, but were unable to tell what had changed or where it had happened.

In addition to symbol variables, there are also individual variables that affect the detection of change (Sekuler & Sekuler, 2000). Previous research found an effect of participants' age on detection accuracy and encoding time. For example, Costello, Madden, Mitroff and Whiting (2010) found change-detection accuracy to be worse for older participants (age 60 and older) than younger participants. Ratcliff, Thapar, and McKoon (2001) examined the effects of participant age on response time in a simple detection task and found that older participants (age 60 and older) had a longer response time than younger participants (age 59 and younger). There is also research on encoding and comparing information from only a few visual samples that has relevance to pilot scan patterns. Using a measure of stimulus recall fidelity, Peich, Husain and Bays (2013) examined the exactness by which participants could reproduce (from memory) the orientation and color of a single rectangle from an array of colored rectangles. Their result indicated an age-related decrease in the resolution by which visual information can be maintained in working memory. The older participants (age 66 and older) exhibited wider error distributions, indicating less accurate memory.

Taken together, previous research has found change blindness effects from symbol and color variations and revealed that these effects are more pronounced for older participants (age 60 and older) than younger participants (age 59 and younger). These results are important and point to several gaps in current weather display research. First, weather display developers should aim to develop display symbols that allow rapid encoding and detection. This is especially important considering the large number of weather symbols and different backgrounds used on modern weather displays. Second, as part of the symbol-evaluation process, care must be taken to ensure that the weather symbology is equally effective for all age groups. This is especially important because approximately 30% of all U.S. private pilots are 60 years of age or older (FAA, 2016).

Therefore, the purpose of the present study is to assess detection performance of weather symbols and provide recommendations to reduce change blindness.

2 PURPOSE

The present study aimed to evaluate the effect of symbol enhancement on pilot visual performance to reduce or resolve the change blindness gap found by previous research. The specific goals of the study were:

1. to evaluate how enhancements to weather symbol salience affect pilot change-detection performance (Experiment 1);
2. to evaluate how line segment salience affects pilot change-detection performance (Experiment 2); and
3. to assess the presence of age-related effects on change-detection performance and response time.

3 EXPERIMENT 1

The purpose of Experiment 1 was to empirically evaluate the effect of enhancements to weather symbology salience on change-detection performance. Based on previous research, we predict that change-detection performance will increase as symbol salience increases.

3.1 Participants

A total of 97 private GA pilots volunteered to participate in Experiment 1. We present their background information in Table 1.

Table 1. Descriptive Characteristics of Study Participants

			Flight hours accrued					
	Age (years)		Total		Instrument		Instrument- last 6 mo.	
N	Median	Range	Median	Range	Median	Range	Median	Range
97	48	18 - 82	325	65 - 725	12	0 - 1000	1	0 - 90

We divided the N=97 participants into two groups: younger pilots (n=67; age 59 and younger) and older pilots (n=30; age 60 and older) similar to previous research (Costello, Madden, Mitroff and Whiting, 2010; Ratcliff, Thapar and McKoon, 2001).

Because a wide range of visual stimulation (including flicker sequences) can lead to seizures in epileptics, we planned to exclude participants from participation in Experiment 1 if they had a personal and/or familial history of epilepsy. No pilots declined to participate because of a history of epilepsy.

3.2 Informed Consent Statement

Each participant read and signed an informed consent statement before beginning the experiment. Informed consent statements describe the study, the foreseeable risks, and the rights and responsibilities of the participants, including that their participation in the study is voluntary. All the information that the participant provided, including Personally Identifiable Information (PII) is protected from release except as may be required by statute. Signing the form indicated that participants understood their rights as participants in the study and their consented to participation.

3.3 Research Personnel

The research team consisted of personnel from the FAA Human Factors Branch. We developed stimuli, performed initial testing, briefings, data collection, and analysis of the data.

3.4 Facilities

We conducted the study at the William J. Hughes Technical Center Cockpit Simulation Facility (CSF). The CSF has a dedicated computer section for change-detection experiments.

3.5 Stimulus Experiment System

We used the Stimulus Experiment System (SES) software installed on Intel Corei5 HP® desktop computers equipped with a Dell® P2212H Liquid Crystal Display monitor (as shown in Figure 1) to display the change-detection stimuli to participants. This system enables researchers to administer the experimental tasks to the participant by: 1) selecting the experimental protocol (i.e., one-shot technique); 2) assigning a coded identifier to the participant; and 3) present a set number of experimental trials to the participant. Participant response data were recorded automatically and written to a data file.

Figure 1. Experiment desktop computer

3.6 Data Handling Procedure

We assigned a coded identifier to each participant pilot and tagged all other data-collection forms, computer files, electronic recordings, and storage media containing participant information only with the coded identifier, not the name or personal identifying information of the participants.

3.7 Change-Detection Technique

Change detection has been defined as "the apprehension of change in the world around us" (Rensink, 2002, p. 246). In GA operations using weather presentations, it is imperative that pilots can perceive symbol locations correctly and also detect important symbol changes. This is especially important because pilots often need to capture display information "at a glance."

The goal of Experiment 1 was to examine GA pilots' abilities to detect changes in weather presentations in a part-task environment (i.e., change detection as an isolated task). One experimental technique used to assess of change-detection performance is the one-shot technique (Rensink, 2002). Using this technique, researchers present participants with a pair of images (i.e., Image 1 and Image 2) sequentially. Image 1 is displayed for a short duration, then a blank screen masks the display briefly, and then Image 2 is displayed. The participant presses a button to indicate whether Image 2 was different than Image 1.

During Experiment 1, we used two types of trials:

- Signal trials are those trials in which a change occurs between Image 1 and Image 2 (i.e., there is a change to detect).
- Noise trials are trials in which Image 1 is identical to Image 2 (i.e., there is no change to detect). In the present experiment, half of the images are noise trials. By including noise trials, participants must make a decision between two options (i.e., change detected/no change detected).

For Experiment 1, we used the one-shot technique, as shown in Figure 2. During the experiment, participants initiated each trial by pressing the spacebar on the computer keyboard. First, a grey screen with a central fixation cross appeared for 1000 ms, and then Image 1 was displayed for 400 ms. Image 1 was then replaced by a blank, grey screen for 1000 ms, after which Image 2 was displayed for 400 ms (Droll, Gigone, & Hayhoe, 2007; Turatto, Angrilli, Mazza, Umiltà, & Driver, 2002). Participants provided a response by pressing one of two buttons on the keyboard to indicate either "Yes" (change detected) or "No" (no change detected). The software waited for a response for 60 seconds; if the participant did not enter a response within that period, the trial ended automatically and recorded a "no response."

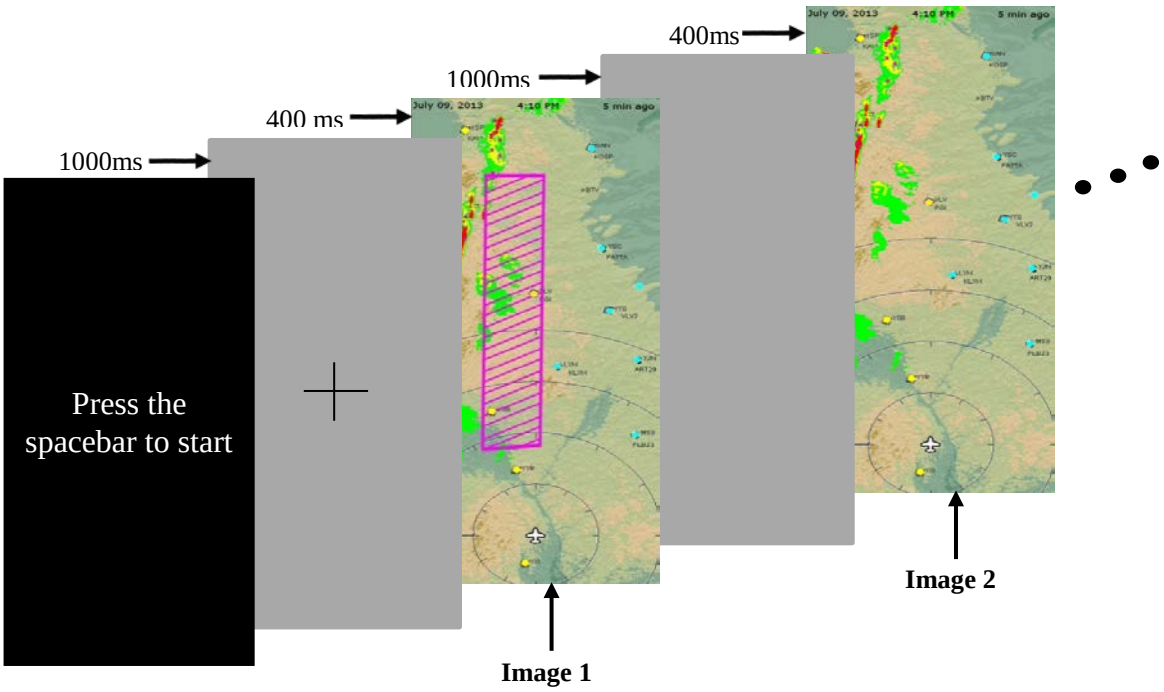


Figure 2. Example trial using the one-shot change-detection paradigm

3.8 Analysis and Manipulation of Salience Information

Modern cockpit weather displays can be very complex because the large number of weather elements that are presented on top of a detailed background map. It is therefore imperative that individual weather elements are clearly discernable (i.e., salient), allowing pilots to detect both small and large changes.

A salient object (element) stands out against neighboring elements in a natural scene or an image. Salient features capture our attention and thereby enhance detection. Researchers have developed models that predict visual attention and fixation patterns from salience, as defined by measurable stimulus characteristics. According to these models, the visual detection process is driven by fast, pre-attentive processing, as well as slow, top-down processing. A well-known computational model is an algorithm by Itti, Koch, and Niebur (1998) that first integrates image features into a topological salience map. The model then selects image regions in order of decreasing saliency. In a final processing stage, the salient (i.e., conspicuous) regions are selected for a detailed analysis. In this way, a complex scene is rapidly broken down into areas of interest.

Similar to the analysis of a natural scene, saliency-based approaches can be used to analyze weather display images. In general, salience algorithms determine the contrast of image regions relative to their nearest-neighbor regions, using features of color, luminance, and intensity. Therefore, a salience analysis can break down a complex weather image into salient features or conspicuous regions. However, one drawback with most saliency algorithms is that the resulting salience map is of very low resolution and would, therefore, not be suitable to analyze detailed and complex weather maps. One exception is a frequency-tuned salience algorithm proposed by Achanta, Hemami, Estrada, and Süsstrunk (2009). In their adaptation, the output of the saliency-detection analysis (i.e., the salience map) has more frequency content from the original image and can therefore produce a salience map in full resolution. In addition to their published research, Achanta et al. (n.d.) also provide free online software that implements their salience.

Here, we present our general procedure for performing a salience-based analysis of weather display images. For all salience analyses, we used the Achanta et al. software to generate salience maps, and the GNU Image Manipulation Program (GIMP) (www.gimp.org) for creating change-detection images and for measuring symbol intensities on salience maps. Becker, Sundar, Bello, Alzahabi, Weatherspoon, and Bix (2016) used a similar approach for an analysis of the salience of front-of-pack nutrition labels.

In the Achanta et al. (2009) framework, the salience map (S) for a weather image (I) of a given width (W) and height (H) is given by:

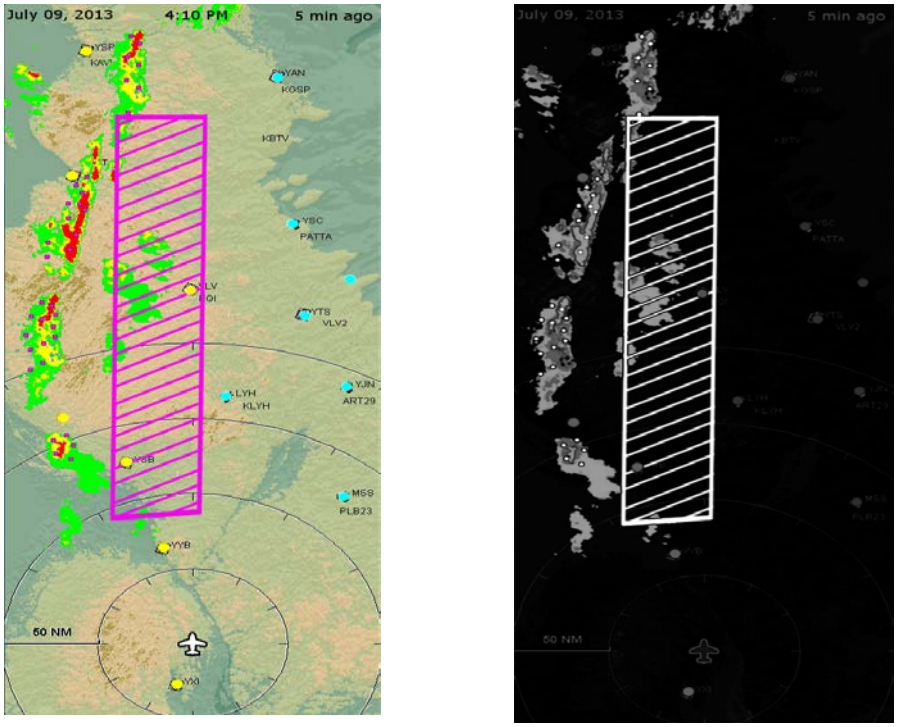
$$S(x, y) = |I_{\mu} - I_{whc}(x, y)| \quad (1)$$

where I_{μ} is the mean image pixel value and I_{whc} is the Gaussian blurred version of the original image. In the Achanta et al. application, Equation 1 is extended to use features of color and luminance using the *Lab* color space:

$$S(x, y) = \|\mathbf{I}_{\mu} - \mathbf{I}_{whc}(x, y)\| \quad (2)$$

where $\mathbf{I}_{\mu}$ is the average image feature vector, $\mathbf{I}_{whc}(x, y)$ is the Gaussian blurred version's image pixel vector value, and $\| \|$ is the L_2 norm. Using the *Lab* color space, each pixel location is therefore an $[L, a, b]^T$ vector with the L_2 norm being the Euclidean distance.

Figure 3(a) shows a weather display image that is similar to some of the images used in the Ahlstrom and Suss (2014) study. The image has weather elements like precipitation, lightning strikes (purple dots), Significant Meteorological Advisory (SIGMET) area (solid purple lines), and METAR information (yellow IFR and blue VFR circles). When we analyze the image using the Achanta et al. software, we acquire the salience map that is shown in Figure 3(b). In the salience map, the brightest regions have the highest salience, and the "darkest" regions have the lowest salience. As we can see in the salience map, the SIGMET area and the lightning strikes have the highest salience. Other image features like the time-stamp text (top of image) and the METAR symbols have low salience, with the precipitation areas falling somewhere in-between the salience of the SIGMET and the METARs. The topological background map has the lowest salience of all the image features.



(a) (b)

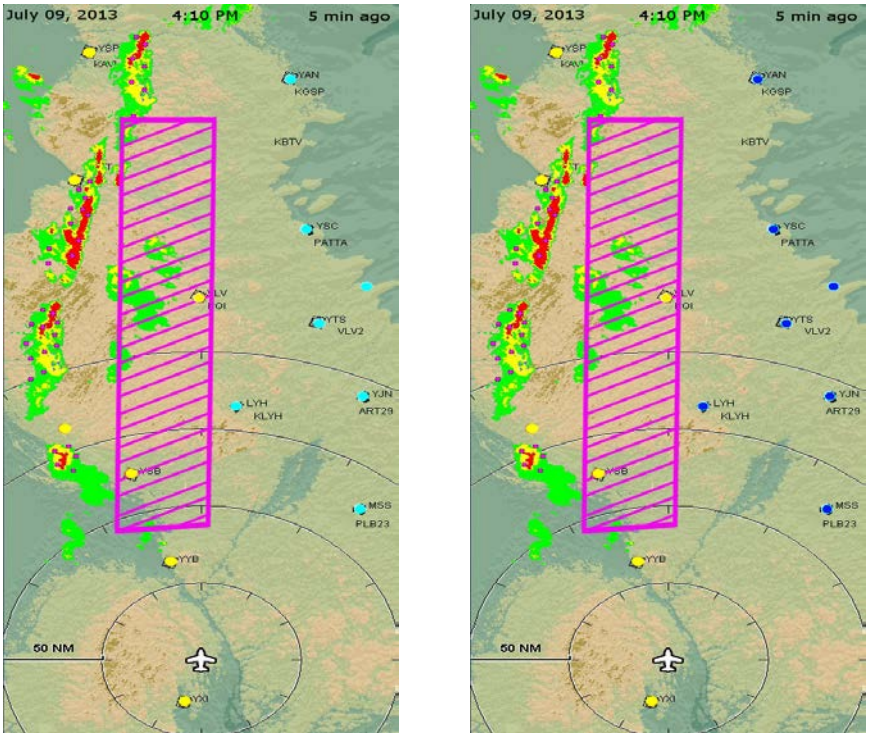
Figure 3. (a) an original weather display similar to the images used by Ahlstrom and Suss (2014); (b) the salience map produced by analyzing the image to the left with the Achanta, Hemami, Estrada, and Süsstrunk (2009) software

With regard to our use of the word "color" during our preparation of the change-detection images for Experiment 1 and Experiment 2, we manipulated the colors used to define line segments, lightning symbols, METARs, and the text in the time-stamp information. These color adjustments were done to increase the salience of features in the enhanced images. This does not mean that we are proposing that weather display developers should implement the colors we used for the enhanced images. We are simply using our symbology palettes as an example. Weather display images are unique as developers use different symbols, color palettes, and background maps for their weather products. Therefore, there is no single color adjustment that would work for all weather images as the salience of a particular feature is determined by many

factors including color, luminance, spatial frequency, and feature orientation of surrounding elements. Our color adjustments show how we can manipulate the salience of weather image features—with the hypothesis that the relative salience drives the detection of image symbol change. Whereas the particular color adjustments are unique for different weather images, our framework for manipulating image salience is general. The same procedure can be used for other similar weather images that were created with different color palettes and different weather symbology.

As an example of the process, we present a scenario that exemplifies a detection problem and how our methodology can be used to enhance individual weather symbols. We run a change-detection experiment and find that pilots have difficulty seeing color changes for METAR symbols (e.g., VFR to IFR—a blue circle changes to a yellow circle). We analyze the weather images using the Achanta et al.(2009) software to acquire salience maps. We then analyze the salience maps for METAR intensities and find that the intensities for the blue and yellow circles are very close, which means that the salience is also very similar. If the salience for different METAR colors is very similar, it means that a change from one color to another will be difficult for pilots to detect. Therefore, we want to increase the salience difference between the METAR colors to make the changes more visible to pilots.

To increase the salience of the VFR METAR symbols (see Figure 4, light blue circles) we performed the following steps:

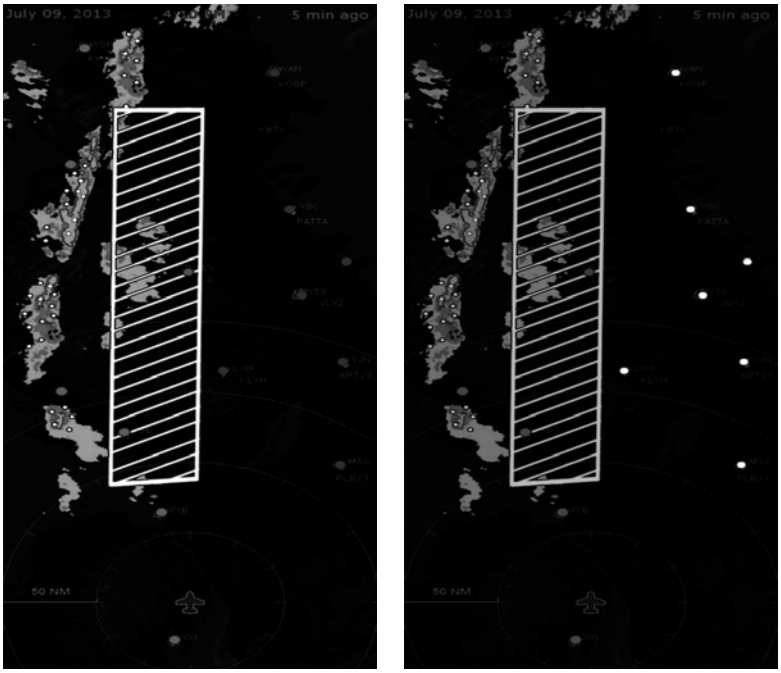


(a) (b)

Figure 4. (a) original image—blue METARs have RGB(0, 254, 254);
(b) adjusted image—the center of the blue METARs have RGB(4, 27, 245)

First, we used GIMP to make a change to the blue VFR METAR colors in the image—either by adjusting the original color value or by using a different color. In Figure 4, the original image is shown on the left, and the adjusted image is shown on the right

Next, we performed a salience analysis of the original image and the adjusted image using the Achanta et al. software. The output of this analysis is a salience map for each image, as shown in Figure 5. Note the difference between the VFR METARs in the (a) original image and (b) the adjusted image.



(a) (b)

Figure 5. (a) salience map from the original image; (b) salience map from the enhanced (color manipulated) image

In the next step, we loaded each of the two salience maps into GIMP. In GIMP, we can either use the Free Select Tool or the Fuzzy Select Tool (see Figure 6) to select the METAR symbol.

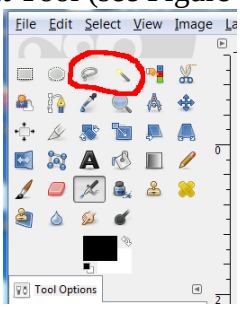
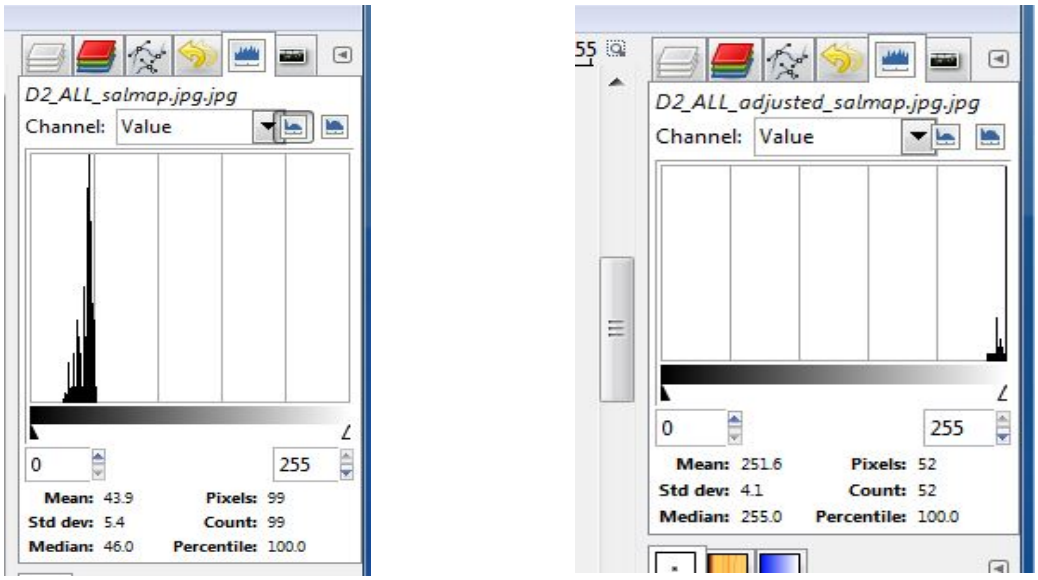


Figure 6. The Free Select Tool and the Fuzzy Select Tool in GIMP

Once a symbol is selected, we can read off the channel-intensity parameters in the Histogram Dialog (see Figure 7; here we only use the mean value). The intensity levels range from 0–255 with a black pixel being 0 and a white pixel being 255. In Figure 7(a), we can see that the mean intensity for a VFR METAR symbol in the original image is 43.9, whereas it is 251.6 for the color-adjusted VFR METAR (see Figure 7(b)). The mean intensity value for an IFR METAR symbol (yellow circle) in the originl image = 91.7. Therefore, the intensity difference for a color change (yellow to blue) in the original image is 91.7 - 42.3 = 49.4 intensity units. For the adjusted image, the intensity difference is 251.6 - 67 = 184, which implies a 3.7 times higher intensity change for a yellow/blue color change compared with the original image.



(a) (b)

Figure 7. (a) the mean intensity value for a METAR in the original image = 43.9; (b) the mean intensity value for a METAR in the adjusted image = 251.6

In the final step, we verified that the adjusted METAR symbols improve the detection of the blue/yellow color change. Depending on the task at hand, this could amount to an informal visual evaluation by one or more observers, or the design and execution of a formal change-detection experiment.

The procedure outlined in steps above can be used for an adjustment of any weather image feature such as lines, symbols, solid infills, semi-transparent infills, and text.

3.9 Experimental Stimuli

Modern weather displays present a multitude of information to pilots. Figure 8 shows three different presentations that reveal the same weather information, but in different symbols and colors. In addition to weather symbols, most weather applications also allow the user to select different backgrounds like topological maps or detailed VFR/IFR charts, as shown in Figure 9.

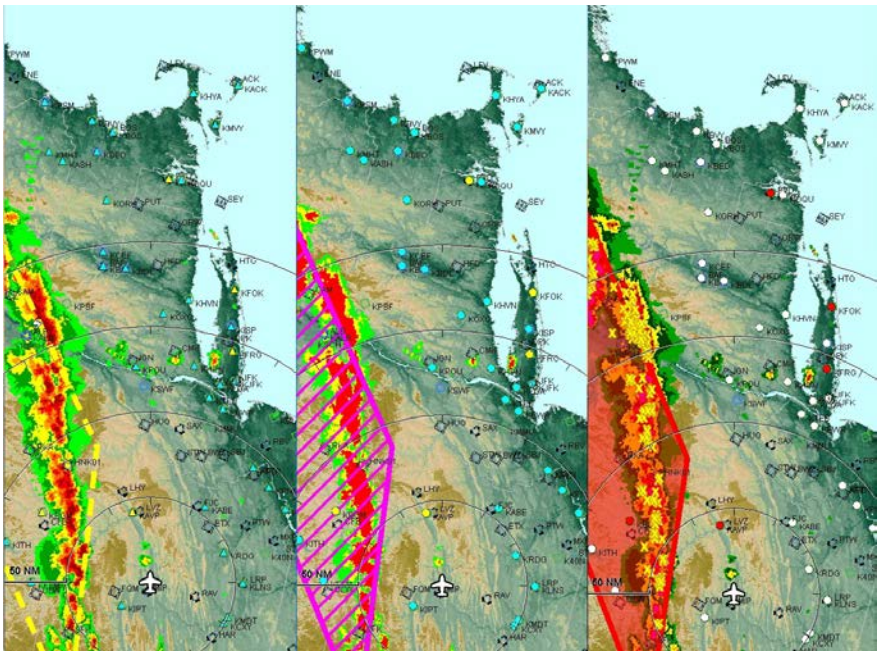


Figure 8. Example weather displays using a topological background map

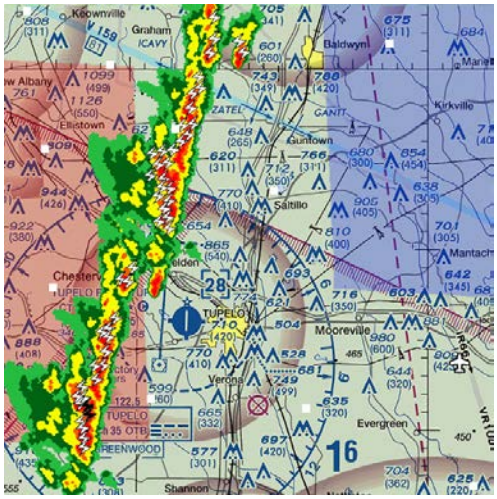


Figure 9. Example weather display using a VFR background map

Modern cockpit weather displays typically incorporate the following types of weather information (see Figure 10):

- METARs for a specific location—Small, color-coded symbols are used to summarize METAR as either VFR or IFR flight conditions, according to visibility and ceiling;
- SIGMET information—Shows advisories on weather that is significant to the safety of all aircraft. Regions affected by the SIGMET are typically enclosed by a polygon (e.g., rectangle);
- Lightning strikes—Regions affected by lightning strikes are marked by small symbols in the shape of a lightning bolt;

- Precipitation—Shows the intensity of precipitation overlaid on the active map; and
- The time stamp—Contains the current date and time, and length of time in minutes since the display was last updated.

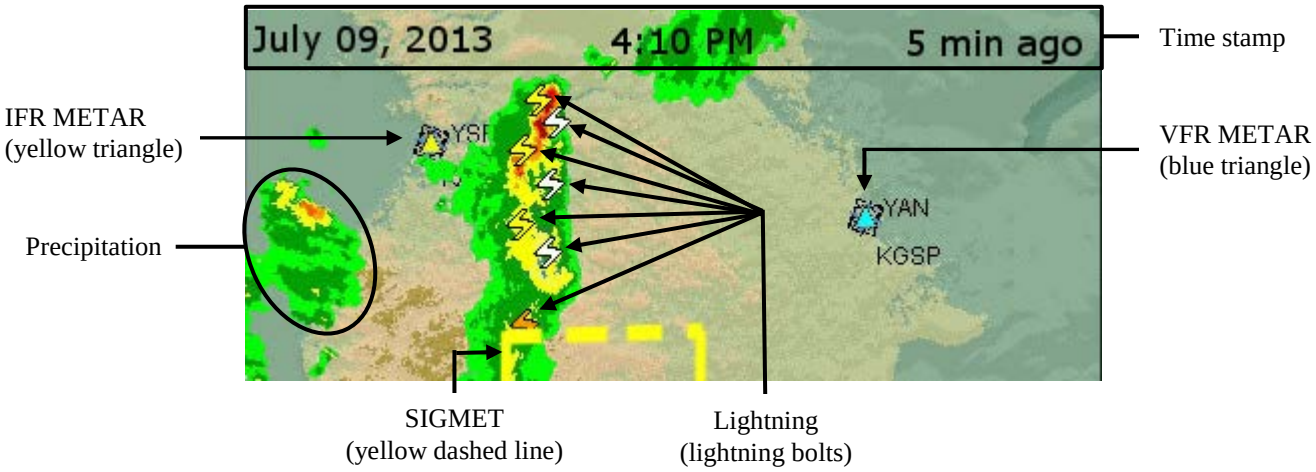
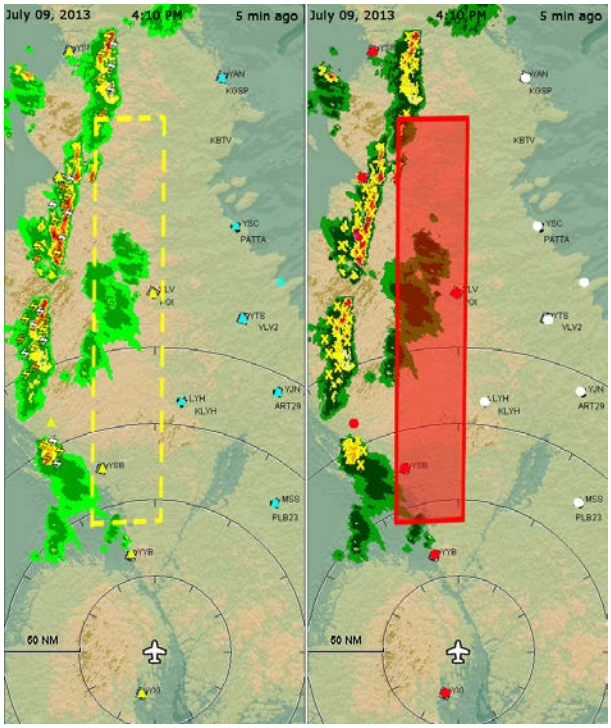


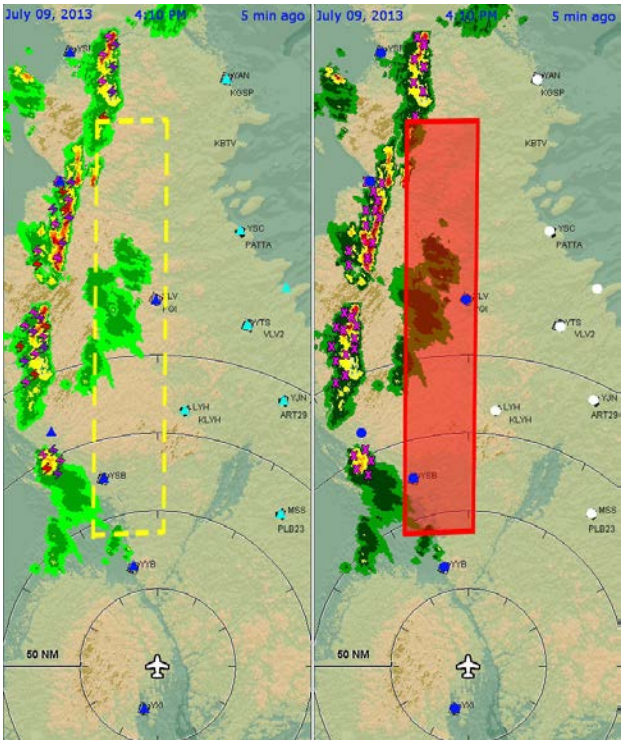
Figure 10. Example of a weather display showing different weather-information symbols

To get some stimulus variability in Experiment 1, we used two symbol sets (D1 and D3) previously used by Ahlstrom and Suss (2014). The main focus of Experiment 1 is to evaluate whether an increase in symbol salience affects symbol discrimination. For this purpose, we used the original D1 and D3 displays for a Control condition and enhanced versions of the same displays for the Enhanced condition. The display enhancement implies that we increased the salience for three symbols; METAR, lightning, and time-stamp information. Figure 11 shows the Control D1 and D3 displays, and Figure 12 shows the Enhanced D1 and D3 displays.



(a) (b)

Figure 11. The (a) original D1 and (b) D3 displays used in the Control condition



(a) (b)

Figure 12. The (a) D1 and (b) D3 displays used in the Enhanced condition

The Enhanced color palette was used to increase the symbol salience compared with the Control color palette. Table 2 provides the color changes for each symbol element.

Table 2. Color Values for the Control and Enhanced Weather Symbols

Symbol	Control	Enhanced
D1 - IFR METAR symbol	Yellow RGB (231, 255, 7)	Blue RGB (6, 24, 244)
D1 – Lightning symbols	White RGB (255, 255, 255)	Dark Purple RGB (150, 1, 243)
	Yellow RGB (255, 252, 1)	Light Purple RGB (254, 1, 250)
	Orange RGB (252, 151, 0)	Red RGB (254, 0, 0)
D1 - Time stamp text	Black RGB (0, 0, 0)	Blue RGB (6, 26, 247)
D3 - IFR METAR symbol	Red RGB (255, 0, 21)	Blue RGB (6, 26, 247)
D3 – Lightning symbols	Yellow RGB (255, 251, 6)	Light purple RGB (254, 1, 250)
D3 - Time stamp text	Black RGB (0, 0, 0)	Blue RGB (6, 26, 247)

In Experiment 1, we manipulated a change of the following weather symbols:

1. METAR—The onset/offset of 14 METAR symbols in the display;
2. METAR—A change in the color of two groups of 7 METAR symbols;
3. Lightning—The onset/offset of lightning symbols in the display; and
4. Time stamp—The onset/offset of the time stamp text.

Table 3 presents the result of the enhancement in terms of symbol salience increases as measured from the symbol intensity on the salience maps. For the D1 images, we increased the salience of the IFR METAR symbols from the mean Control salience of 111.5 to a mean Enhanced salience of 252.5 (dark blue triangles). For the D1 METAR onset/offset displays, the empty METAR stations had a mean salience of 8.9. We also increased the salience of the D1 lightning bolt symbols (white, yellow, and orange lightning bolts) from a Control mean salience of 79 to an Enhanced mean salience of 159 (shown as red, light purple, and dark purple lightning bolts).

Table 3. Salience Intensities in the Control and Enhanced Images (D1 and D3 Displays)

	D1		**D3**	
Weather element	*Control*	*Enhanced*	*Control*	*Enhanced*
METAR VFR	No change	No change	No change	No change
METAR IFR	111.5 (*SD*=1.7)	252.5 (*SD*=1.7)	150 (*SD*=0.53)	252 (*SD*=0.48)
METAR empty background	8.9 (*SD*=3.6)	8.9 (*SD*=3.6)	22.8 (*SD*=22.2)	22.8 (*SD*=22.2)
Lightning	79 (*SD*=38.3)	159 (*SD*=39.3)	86.4 (*SD*=3.1)	178.7 (*SD*=4.5)
Time stamp text	97.3 (*SD*=8.24)	151.25 (*SD*=18.8)	55.6 (*SD*=6.24)	147.9 (*SD*=15.7)

For the D3 images, we increased the Control IFR METAR symbol (red circles) from a mean salience from 150 to a mean Enhanced salience of 252 (shown as blue circles). For the D3 METAR onset/offset displays, the empty METAR stations had a mean salience of 22.8. We also

increased the salience for the D3 lightning symbols (yellow X symbols) from a mean of 86.4 for the Control displays to a mean of 178.7 for the Enhanced displays (light purple X symbols).

Finally, we enhanced the display of time-stamp information by increasing the salience of the text in the enhanced displays. The time-stamp information displayed the time when the weather display was last updated. This information was the same for all images (it never changed), with "July 09, 2013," "4:10 PM," and "5 min ago." In the Control D1 images, we increased the mean salience of the black time-stamp information from 97.3 to a mean salience of 151.25 in the Enhanced D1 image. For the D3 images, we increased the Control black time-stamp text from a mean salience of 55.6 to a mean salience of 147.9 in the Enhanced image.

At a viewing distance of 64 cm, the viewing angle of the experimental images (428 × 1021 pixels) subtended 9 degrees horizontally and 20 degrees vertically.

A manipulation of the image variables in Experiment 1 gives the following image design: 2 (color; Control and Enhanced) x 4 (symbol change) x 2 (symbology; D1 and D3) x 2 (replication) = 32 change trials. In addition to these 32 change trials, we also created 32 noise trials (no change; the same image is shown twice) and 13 practice trials (divided into eight change trials and five noise trials). This yields a total of 77 trials per participant and session. During each session, each participant was shown both the D1 and D3 images in random order. However, the independent variable color (Control and Enhanced) is a between-subjects variable, so each participant belonged to either the Control group or the Enhanced group.

3.10 Independent Variable

The independent variable in Experiment 1 is color (Control and Enhanced).

3.11 Dependent Variables

The dependent variables are the observed counts of hits, false alarms, misses, correct rejections, and the time it took participants to respond.

3.12 Procedure

After reading and signing the Informed Consent Statement, participants completed a biographical questionnaire before performing the experiment. We explained to the participant that the task instructions would be presented on-screen in a self-paced manner. After following the on-screen instructions, participants first completed the practice trials, followed by the experimental trials. We instructed the participants to respond as quickly as possible without making any errors. Participants could pause for as long as they wanted after each trial before continuing the experiment. There was no feedback during the practice or experimental trials.

Participants began each trial of the change-detection task by pressing the space bar on the keyboard. For each trial, the participant responded to whether or not ("YES" or "NO") a change occurred between Image 1 and Image 2. The first 13 trials were practice trials, divided into eight signal (change) trials and five noise (no change) trials. Following the practice trials, the same process was repeated for the 64 experimental trials. Half of the 64 trials were noise trials (i.e., no change) with the remainder of the trials being signal trials (i.e., change). For each participant, we presented the trials in random order.

3.13 Data Analysis

For Experiment 1, the noise trials were created by displaying the same image for both Image 1 and Image 2. Because we were using "YES" and "NO" responses, we labeled each response as a "hit" if the participant responded "YES" to a signal trial, a "false alarm" if the participant responded "YES" to a noise trial, a "miss" if the participant responded "NO" to a signal trial, and "correct rejection" if the participant responded "NO" to a noise trial.

From the observed counts of hits, false alarms, misses, and correct rejections, we derived indexes of discriminability (*d*) and bias (*c*) using an equal-variance Signal Detection Theory (SDT) model (Lee, 2008) and Markov Chain Monte Carlo (MCMC) sampling. The model infers *d* and *c* from the hit and false alarm rates using a cumulative standard normal distribution. The observed counts of hits and false alarm rates are binomially distributed. The priors on *d* and *c* are Gaussian distributions, corresponding to uniform priors over the hit and false alarm rates. The discriminability index, *d*, measures how easily participants can distinguish signal trials (change) from noise trials (no change). The higher the *d* value is, the easier it is for participants to detect a change, whereas a *d* value of 0 corresponds to random guessing. The bias index, *c*, is a measure of the participant decision-making criterion. If a participant has a positive value of the bias index, the participant has a bias to respond "NO." This will result in an increase in the number of correct rejections but will also increase the number of misses. If a participant has a negative bias index, the participant displays a bias towards answering "YES," and this leads to an increase in the number of hits and an increase in the number of false alarms.

During the analysis, we used Just Another Gibbs Sampler (JAGS) (Plummer, 2003, 2011) that we called from R (R Development Core Team, 2011) via the package rjags. All software for the analysis and figure generation is adapted program code from Lee (2008) and Kruschke (2014).

The Bayesian analysis generates a posterior distribution, which is a distribution of credible parameter values. We can use this large distribution of representative parameter values to evaluate certain parameters or to compare differences between parameters. Here, we use a separate decision rule to convert our posterior distributions to a specific conclusion about a parameter value. When plotting the posterior distribution, we include a black horizontal bar that represents the 95% High Density Interval (HDI). The HDI has a higher probability density compared to values that fall outside the HDI. When we compare conditions (i.e., perform contrasts), we compute differences at each step in the MCMC chain and present the result in a histogram along with the HDI. These histograms show both credible differences and the uncertainty of the outcome. If the value 0 (implying zero difference) is not located within a 95% HDI, we say that the difference is credible. If the 95% HDI includes the value 0, the difference is not credible because it means that a difference of 0 is a possible outcome.

For response time effect size analyses, we are using a Region of Practical Equivalence (ROPE). The ROPE contains values that, for all practical purposes, are the same as a null effect (i.e., no meaningful difference). If the 95% HDI falls completely within the ROPE margins for an effect size, we can declare the presence of a null effect, and unlike traditional analyses, we can accept the null outcome. If, however, the entire ROPE falls outside the 95% HDI, we can reject the presence of a null effect.

To derive the posterior distributions, we used 200,000 samples. For all analyses, we used priors that were vague and noncommittal on the scale of the data.

3.14 Results

3.14.1 Age Effects on Symbol Discrimination Accuracy

In the following sections we present SDT analyses on the lightning, METAR, and time stamp data. First, we assessed if participant age affects symbol discrimination accuracy in the pooled dataset (i.e., we combined the data from all METAR, lightning, and time-stamp trials). Second, we assessed the effect of color (i.e., salience) on discriminability. Third, we repeated the age-affect analysis for each symbol condition separately. Fourth, we performed ex-Gaussian response time analyses to assess potential differences in symbol response time.

One way to assess the main effect of age (x) on discrimination accuracy (y) is to use simple linear regression, allowing a prediction of participant discrimination accuracy from their age. As a first step in this analysis, we computed a simple discrimination index d' (Stanislaw & Todorov, 1999) for each participant and symbol condition:

$$d' = \min(0, \Phi\left(\frac{h+1}{h+m+2}\right) - \Phi\left(\frac{fa+1}{fa+cr+2}\right))$$

where h is the number of hits, m the number of misses, fa the number of false alarms, and cr the number of correct rejections. We then used one d' value (per symbol condition) per participant for the regression analysis.

We used a robust linear regression model (Kruschke, 2014) where each predicted y value is computed as $y = \beta_0 + \beta_1 x$, where β_0 is the y-intercept (where the regression lines intersect the y-axis when x=0), and β_1 is the slope (indicates how much y increases when we increase x by 1). To be robust against outliers, the model uses a t-distribution for the noise distribution instead of a normal distribution (i.e., Gaussian distribution). At the lowest level of the model, each datum comes from a t-distribution with a mean μ, a scale parameter (i.e., standard deviation) σ, and a normality parameter v. The prior on the scale parameter is a broad uniform distribution, and the normality parameter v has a broad exponential prior. Both β_0 and β_1 have broad normal priors that are noncommittal and vague on the scale of the data.

Figure 13 shows the outcome of the regression analysis. On the x-axis, we have participant age and on the y-axis, discrimination accuracy (d'). The black circles are the data points; one d' value per participant and symbol condition (i.e., METAR, lightning, and time-stamp conditions). The lines represent regression lines, and the three vertical distributions are superimposed t-distributed noise distributions.

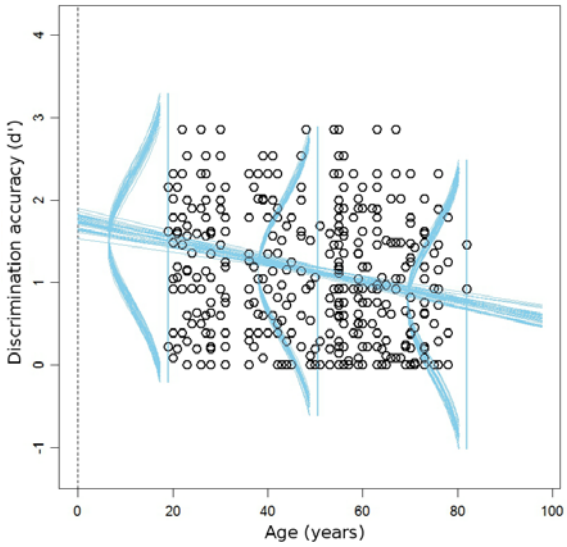


Figure 13. Regression lines and noise distributions for the prediction of d' from age

As we can see in the figure, the regression lines are tightly clustered and have a negative slope, meaning that discrimination accuracy (*y*) for weather symbols decreases with an increasing age (*x*).

Figure 14 shows the mean posterior outcome for β_0 (intercept), β_1 (slope), and the scale parameter σ. The intercept has a mode of 1.7, which is the value of *y* when *x* = 0. The credible slope has a posterior mode of -0.0113. This means that as we increase *x* (age) by 1, *y* will increase by the value of β_1. In this case, the slope is negative, so for each year of *x* (Age), *y* (*d'*) will decrease by the value of β_1.

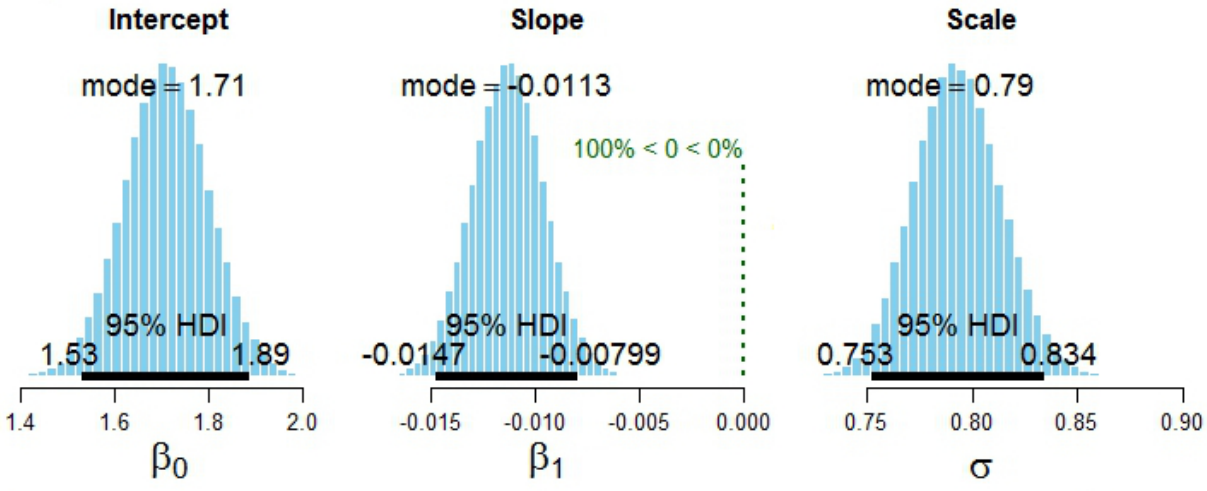


Figure 14. The posterior intercept (β0), slope (β1), and scale (σ) parameters from the robust linear regression analysis

There are several ways to use the predictive outcomes of the linear regression model. For example, we can compute a simple point estimate of the predicted *d'* from a given value of age. In the following, we will use examples for age = 20 years and age = 70 years because these values are on the opposite sides of the age scale (*x*). The following is a prediction for age = 20: If we have *y*=1.7 (intercept) + -0.0113 (slope) * 20 (age), the outcome is *y*=1.472. This means that for age = 20, the model predicts a *d'* of 1.474. Likewise, for age = 70: *y* = 1.7 (intercept) + -0.0113 (slope) * 70 (age) = 0.909. That means for age = 70, the model predicts a *d'* of 0.90, which is a substantially lower discriminability than the predicted *d'* of 1.472 for age =, 20.

A drawback with the simple point estimates above is the fact that we only derive one *d'* value without any distributional information or estimates of uncertainty. However, using the power of Bayesian estimation, we can compute posterior distributions of predicted values of *d'* (*y*) for a given value of age (*x*). We do this by randomly simulating a *y* value from the model using parameter values at each step in the MCMC chain. The resulting distribution is the *y* values for the posterior predictive distribution at *x*.

Figure 15 shows the posterior distributions from the MCMC sampling of predicted mean values of *d'* (*y*) for x_1 = 20 and x_2 = 70. Besides providing an estimate of the mean predicted *d'*, we get distributional information with uncertainty specified by the width of the 95% HDI. As we can see in the figure, there is a fair amount of uncertainty in our posterior predictions. This is expected because we are using *d'* values from three different symbol conditions (i.e., METAR, lightning, and time-stamp conditions), and we know from previous research that discriminability varies for different symbols (Ahlstrom & Suss, 2014; Ahlstrom, Caddigan, Schulz, Ohneiser, Bastholm & Dworsky, 2015a, 2015b). Nevertheless, the model predicts age-related differences for which the average discriminability for 20-year olds is approximately 62% higher than the average discriminability for 70-year olds.

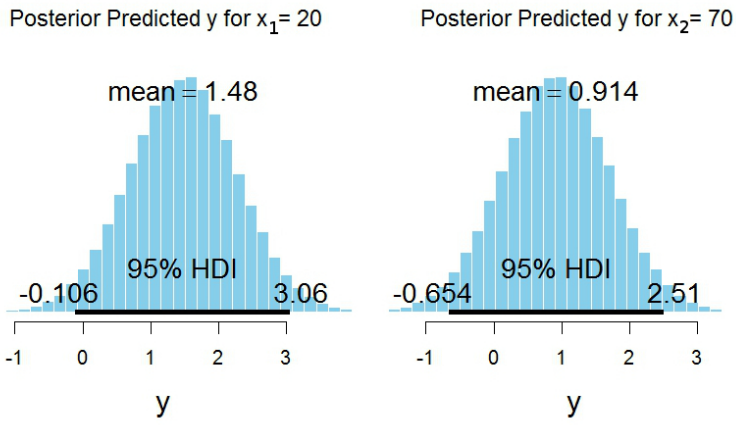


Figure 15. The posterior predicted distributions and mean d' values for x1 = 20 years and x2 = 70 years

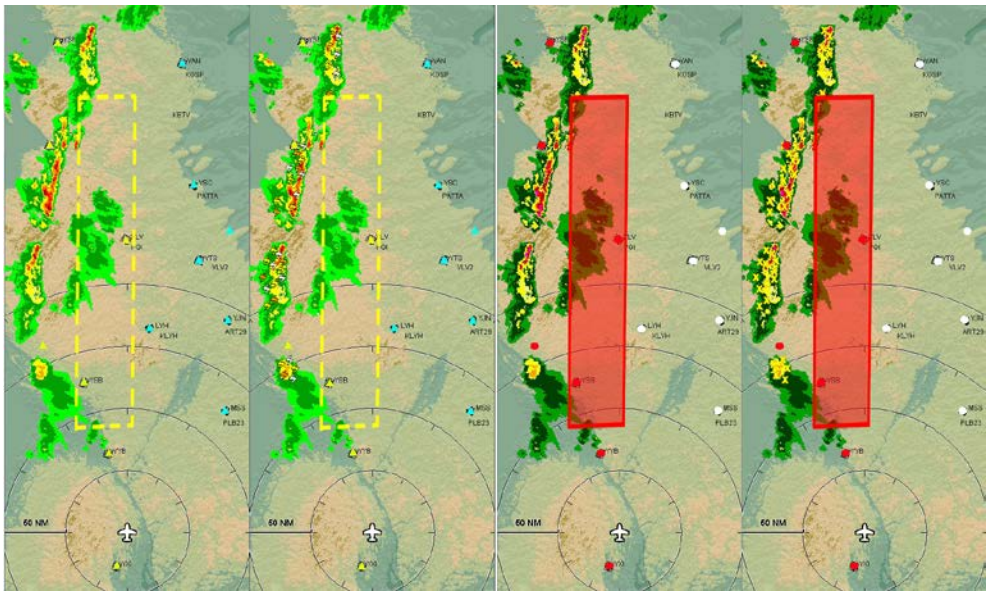
In summary, a robust linear regression on all the *d'* symbol data revealed an overall effect of participant age on discrimination accuracy. Therefore, we performed the same analysis separately for each one of the three symbol conditions.

3.14.2 Lightning Symbols

In Figures 16 and 17, we show the displays used to evaluate lightning changes in the Control and Enhanced conditions, respectively.

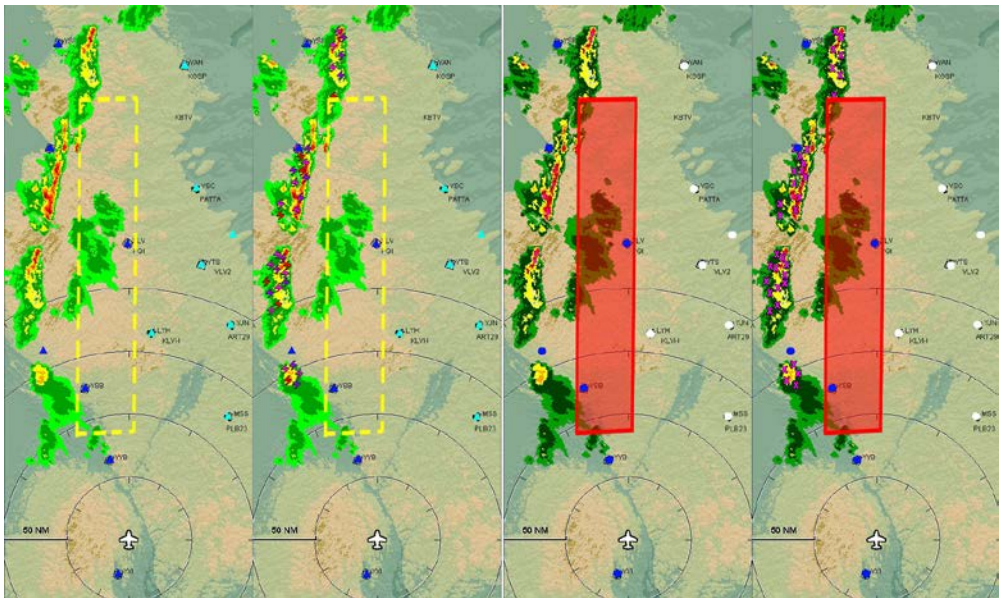
During Experiment 1, the signal trials (change) for the Control condition displayed the two left-most images in Figure 16 for D1 trials (see Figure 16 (a) and (b)), and the two right-most images for the D3 trials (see Figure 16 (c) and (d)). The D1 and D3 signal trials for the Enhanced condition were displayed in the same manner and are shown in Figure 17.

The left side of the D1 and D3 image pairs shows the experimental image without lightning symbols, and the right side of the D1 and D3 image pairs shows the image with lightning symbols. During half of the trials, the left image was displayed as Image 1, and the right image was displayed as Image 2, and vice versa. For the D1 and D3 noise trials (no change), Image 1 and Image 2 both displayed the same left or the same right image. For each trial during the experiment, the participant responded to whether or not ("YES" or "NO") a change occurred between Image 1 and Image 2.



(a) (b) (c) (d)

Figure 16. Illustration of the lightning Control condition stimuli: (a) control D1 display without lightning symbols; (b) control D1 display with lightning symbols; (c) control D3 display without lightning symbols; and (d) control D3 display with lightning symbols



(a) (b) (c) (d)

Figure 17. Illustration of the lightning Enhanced condition stimuli: (a) enhanced D1 display without lightning symbols; (b) enhanced D1 display with lightning symbols; (c) enhanced D3 display without lightning symbols; and (d) enhanced D3 display with lightning symbols

Figure 18 shows the posterior discriminability (*d*) and bias (*c)* for the main effect of salience on the change-detection performance for lightning symbols in the Control and Enhanced conditions. The points in Figure 18 represent individual subjects where the Enhanced condition (2/Blues) are clustered in the higher discriminability and lower bias than the Control conditions (1/Reds) with little overlap between clusters.

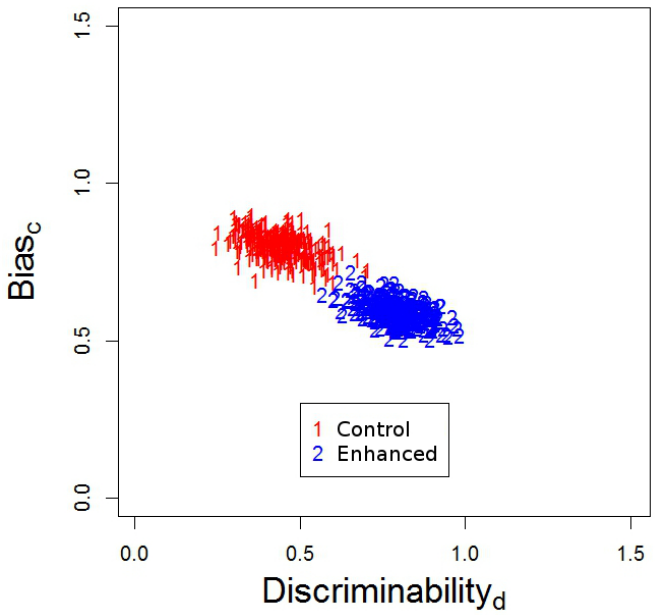


Figure 18. The posterior discriminability index (d) and bias (c) for the lightning Control and Enhanced conditions

Figure 19 shows the posterior *d* and contrast for the Control and Enhanced conditions. As predicted, the mean *d* is higher for the Enhanced displays (*d* = 0.8) than the Control displays (*d* = 0.44). The contrast on the Control versus Enhanced differences is credible, with a mean *d* difference of -0.35.

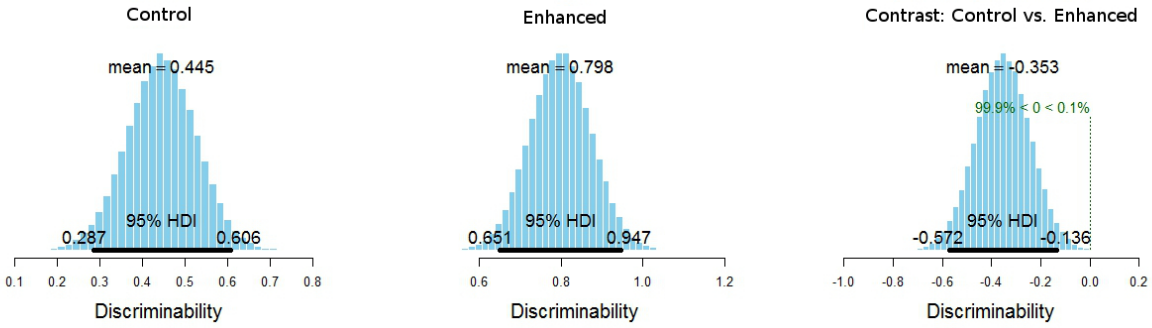


Figure 19. The posterior d and contrast for the Control and Enhanced lightning conditions

In addition to the credible difference in discriminability, there was also a difference in bias between the Control (mean *c* = 0.80; 95% HDI from 0.72–0.88) and the Enhanced (mean *c* = 0.6; 95% HDI from 0.52–0.67) conditions. Participants in the Control conditions had a greater propensity for responding "NO" during the trials. The contrast on the *c* difference was credible, with a mean *c* difference of 0.20 (95% HDI from 0.10–0.31).

3.14.3 Age Effects on Discriminability for Lightning Symbols

To assess the effect of age on lightning symbol discriminability, we performed a linear regression on the *d'* data (using all trials for each participant). Figure 20 shows the outcome of the linear regression analysis. On the *x*-axis, we have participant age, and on the *y*-axis is discrimination accuracy. As shown by Figure 20, the credible regression lines have a negative slope, meaning that discrimination accuracy for lightning symbols decreases with an increasing age.

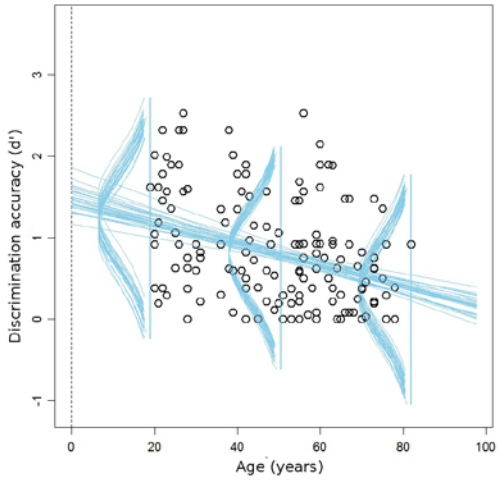


Figure 20. Credible regression lines and noise distributions for the prediction of lightning d' from age

Figure 21 shows the mean posterior outcome for β_0, β_1, and σ. The intercept has a mode of 1.46, and the credible slope has a posterior mode of -0.0133. This means that for each year of age, the model predicts that discrimination performance (*d'*) for lightning symbols decreases by -0.0133.

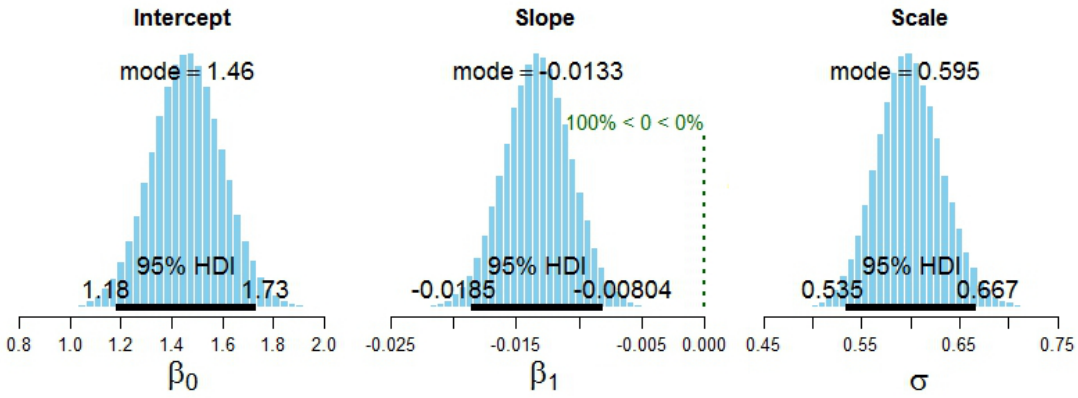


Figure 21. The posterior intercept (β0), slope (β1), and scale (σ) parameters for lightning trials

Figure 22 shows the posterior distributions of predicted mean values of *d'* (*y*) for $x_1 = 20$ and $x_2 = 70$. For $x_1 = 20$ years, the model predicts a mean *d'* of 1.19, and for $x_2 = 70$ years, the model predicts a mean *d'* of 0.524.

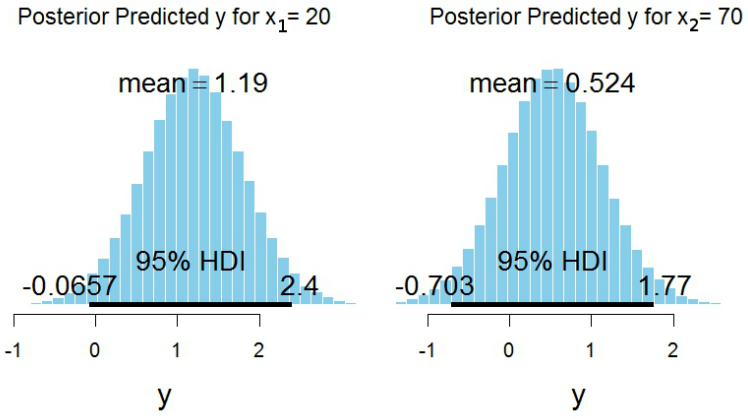


Figure 22. The posterior predicted distributions and mean d' values for lightning symbols with x1 = 20 years and x2 = 70 years

When analyzing the D1 lightning condition separately, we find a mean discriminability of 0.95 (95% HDI from 0.7–1.2) for the younger pilots (age 59 and younger) and a mean discriminability of 0.38 (95% HDI from 0.05–0.7) for the older pilots (age 60 and older). The contrast of this difference is credible, with a mean discriminability difference of 0.6 (95% HDI from 0.15–0.99). For the D3 condition, however, there is no credible difference in discriminability between the younger (*M* = 0.45, 95% HDI from 0.2–0.7) and the older (*M* = 0.24, 95% HDI from -0.09–0.6) pilots (*M* difference = 0.2, 95% HDI from -0.22–0.64).

In summary, the overall discriminability of lightning symbols is low, meaning that participants had a difficult time discriminating lightning signal trials from lightning noise trials. However, there is a credible main effect of color (i.e., salience) on discriminability where participants exhibited a greater discriminability for lightning symbols in Enhanced displays than Control displays. In addition to this main result, we also find credible effects of age on discriminability, with younger pilots (59 years of age and younger) exhibiting an increased discriminability for D1 lightning displays compared with older pilots (60 years of age and older).

3.14.4 Response Time for Lightning Displays

Response time distributions are not symmetrical Gaussian distributions (i.e., normal distributions). On the contrary, response time data are skewed distributions with a steep rise on the left side and a long tail on the right side (Whelan, 2008). Despite this, it is common to base a response time analysis on the mean response time for each participant. However, this type of analysis does not take the response time distribution's shape into account and will, therefore, obscure important aspects of participant response behavior (Heathcote, Popiel, & Mewhort, 1991). We would like to use all the response time responses (i.e., use all the data from the underlying response distribution). Furthermore, because of the skewed distribution for response time data, we cannot use an analysis method that is based on the normal or Gaussian likelihood.

A distribution that fits response time data well is the ex-Gaussian distribution (Matzke & Wagenmakers, 2009). This distribution is a combination of a Gaussian distribution and an exponential distribution. Figure 23 shows a response time dataset in the form of histograms. The left histogram shows an analysis using a normal distribution; the right histogram shows a response time analysis using the ex-Gaussian distribution. The blue lines superimposed on the data represent a posterior predictive check. This is to test that our model provides a good fit to the data. As we can see in Figure 23(a), the normal likelihood model provides a poor fit to the response time data. On the contrary, the ex-Gaussian distribution in Figure 23(b) provides a good fit to the response time data.

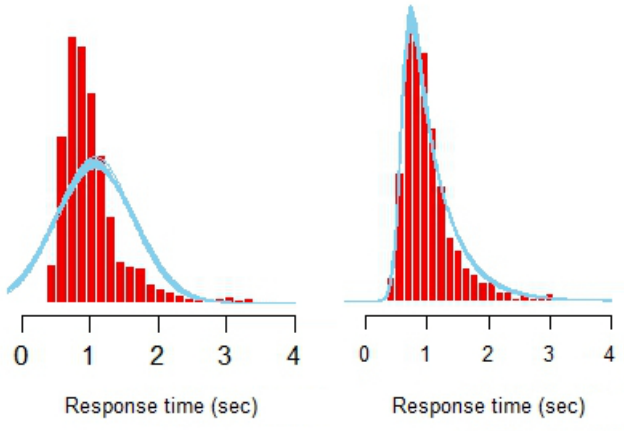


Figure 23. A comparison of response time data analyses using a (a) normal distribution vs. (b) an ex-Gaussian distribution; the blue lines represent a posterior predictive check, verification that the model provides a good fit to the response time data

The Gaussian portion of the ex-Gaussian distribution is described by the parameters μ (mean) and σ (standard deviation), whereas the parameter τ describes both the mean and the standard deviation of the exponential component. The mean of the ex-Gaussian is given by $\mu + \tau$ and the variance by $\sigma^2 + \tau^2$.

For the present ex-Gaussian response time analysis, we used a Bayesian between-subjects model. At the lowest level of the model, we have the individual response times, coming from an ex-Gaussian distribution with a mean, μ, and a standard deviation, σ. The prior on μ is a normal distribution, and the prior on σ is a uniform distribution, with both priors being vague and noncommittal on the scale of the data.

Before running the analyses on our experimental data, we first used the *rexgauss* function (Massidda, 2013) to generate synthetic response time data with known parameter values (i.e., μ = 3.7, σ = 0.8, τ = 2.19). Subsequently, we ran three separate ex-Gaussian analyses on the synthetic datasets (10,000 response time data points each). If our ex-Gaussian model works correctly, it should be able to recover the known parameter values. The result from the analyses

showed mean parameter values for μ = 3.71 (SD = 0.01), σ = 0.80 (SD = 0.01), and τ = 2.17 (SD = 0.03), which imply that our ex-Gaussian model correctly recovered the generating parameters.

For the present analysis, we are interested in the main effect of color on response time. Therefore, we pooled the D1 and D3 data for the Control and Enhanced conditions separately, using all response times for each subject for the signal trials only. Figure 24 shows the outcome of the analysis for the lightning Control (top) and lightning Enhanced (bottom) conditions. As is clear from the figure, the mean response times for the two conditions are virtually identical (*M* of ex-Gaussian=1.01; 95% HDIs from 0.95–1.07). Consequently, for the lightning symbols used in Experiment 1, there is no main effect of symbol salience (Control versus Enhanced) on symbol response time.

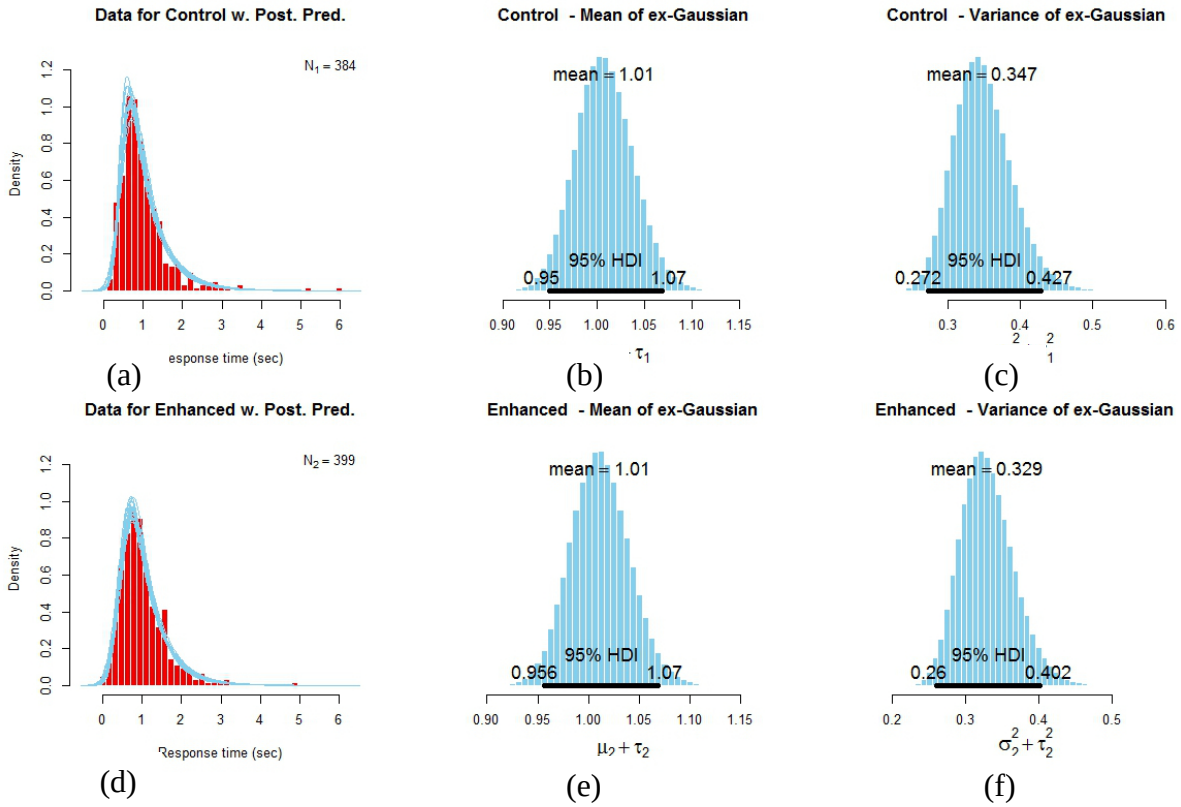


Figure 24. Response time data (a & d) with the mean (b & e) and variance (c & f) of the ex-Gaussian distribution for the Control (a–c) and Enhanced (d–f) conditions

3.14.5 Age Effects on Response Time for Lightning Displays

In addition to the main effect of color on lightning symbol response time, we also assessed the effect of age on response time. We performed a linear regression on the response time data (using all lightning trial response times for each participant). As we can see from Figure 25, there is a slight positive slope on the regression lines indicating that the response times increase with increasing age.

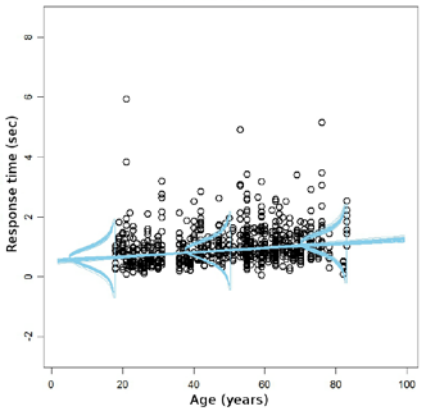


Figure 25. Credible regression lines and noise distributions for the prediction of lightning response time (sec) from age (years)

Figure 26 shows the mean posterior outcome for β_0 (intercept), β_1 (slope), and the scale parameter, σ. The intercept has a mode of 0.5, which is the value of y when $x = 0$. The credible slope has a posterior mode of 0.00753. This means that as we increase x (age) by 1, y will increase by the value of β_1.

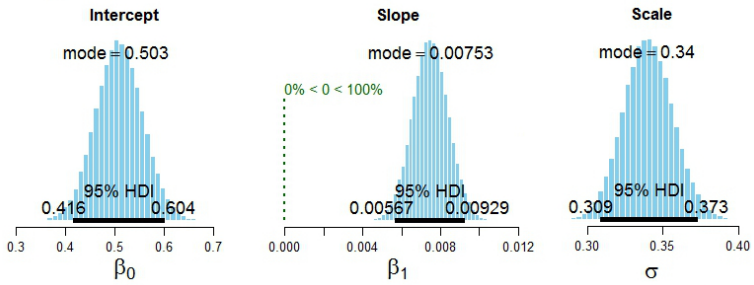


Figure 26. The posterior intercept (β0), slope (β1), and scale (σ) parameters for response time on lightning trials

Figure 27 shows the posterior distributions of predicted mean values of response time (y) for $x_1 = 20$ and $x_2 = 70$. For $x_1 = 20$ years, the model predicts a mean response time of 0.66 sec, and for $x_2 = 70$ years, the model predicts a mean response time of 1.0 sec.

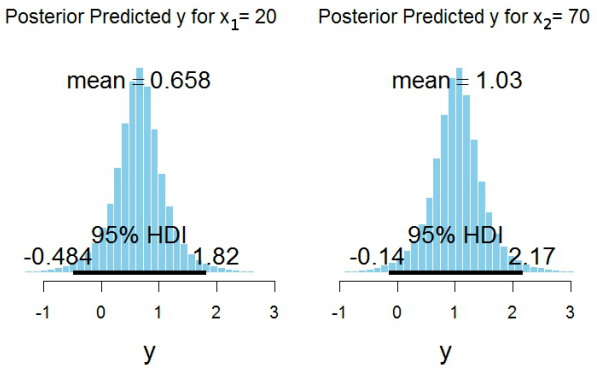


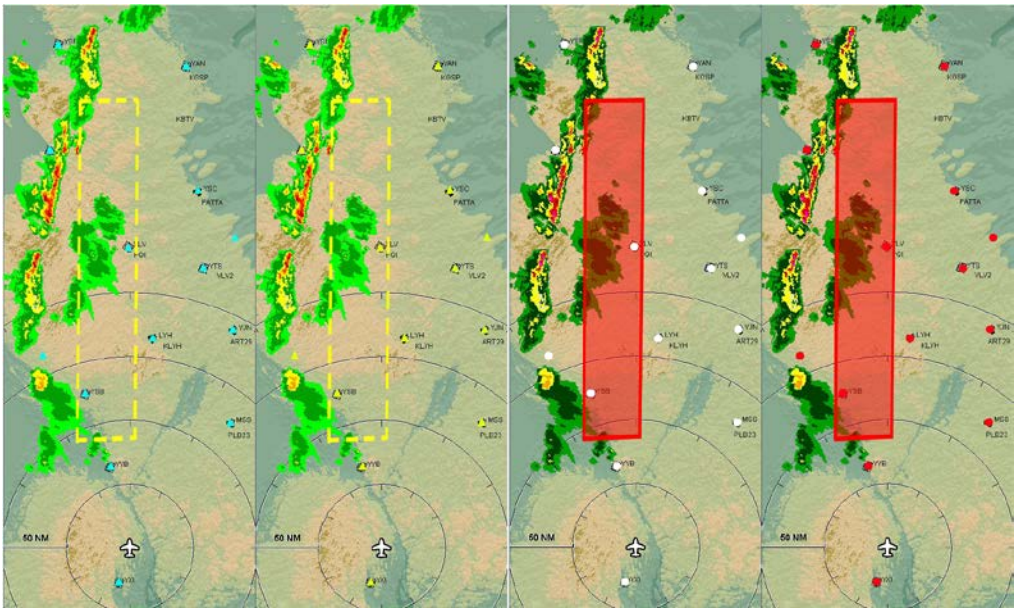
Figure 27. The posterior predicted distributions and mean response time values for lightning symbols with x1 = 20 years and x2 = 70 years

3.14.6 METAR Symbols

We used two different METAR symbol color changes for both the Control and the Enhanced conditions. The first is a color change in which all the 14 METAR symbols changes color between Image 1 and Image 2 for signal (change) trials or stay the same between Image 1 and Image 2 for noise (no change) trials. The second METAR condition is an on/off change for which there are no METAR symbol colors (i.e., empty METAR symbol background) in the first image, but there are in the second image (and vice versa). Figure 28 shows the METAR symbol color changes for the D1-D3 Control condition, and Figure 29 shows the symbol changes for the D1-D3 Enhanced condition. Figure 30 shows the METAR symbol on/off changes for the D1-D3 Control condition, and Figure 31 shows the METAR on/off changes for the D1-D3 Enhanced condition.

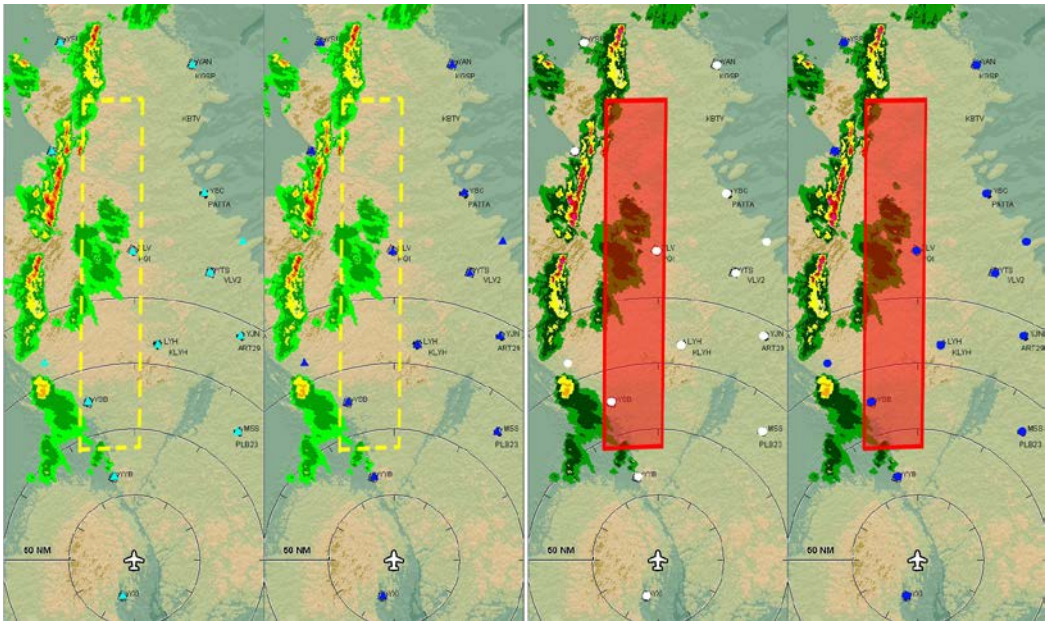
During the experiment, color changes were introduced in the D1 and D3 signal (change) trials by displaying, for example, the leftmost image in Figure 28 (light blue triangles) as Image 1 and the second leftmost image as Image 2 (yellow triangles). Here, there is a METAR symbol color change between Image 1 and Image 2. For the noise (no change) trials, we displayed the same image for Image 1 and Image 2.

For the D1 and D3 METAR symbol on/off changes shown in Figure 30, Image 1 contained no METAR color symbols (i.e., empty METAR backgrounds), whereas Image 2 displayed 14 METAR symbols in two different colors (and vice versa). For example, for the D3 Control condition in Figure 30 (the two rightmost images), Image 1 could display empty METAR stations, whereas Image 2 displayed 7 white and 7 red METAR symbols (and vice versa). Again, for the noise (no change) trials, we displayed the same image for Image 1 and Image 2.



(a) (b) (c) (d)

Figure 28. Illustration of the METAR Control condition stimuli for color changes: (a) Control D1 display with light blue METAR symbols; (b) Control D1 display with yellow METAR symbols; (c) Control D3 display with white METAR symbols; and (d) Control D3 display with red METAR symbols



(a) (b) (c) (d)

Figure 29. Illustration of the METAR Enhanced condition stimuli for color changes: (a) Enhanced D1 display with light blue METAR symbols; (b) Enhanced D1 display with dark blue METAR symbols; (c) Enhanced D3 display with white METAR symbols; and (d) Enhanced D3 display with dark blue METAR symbols

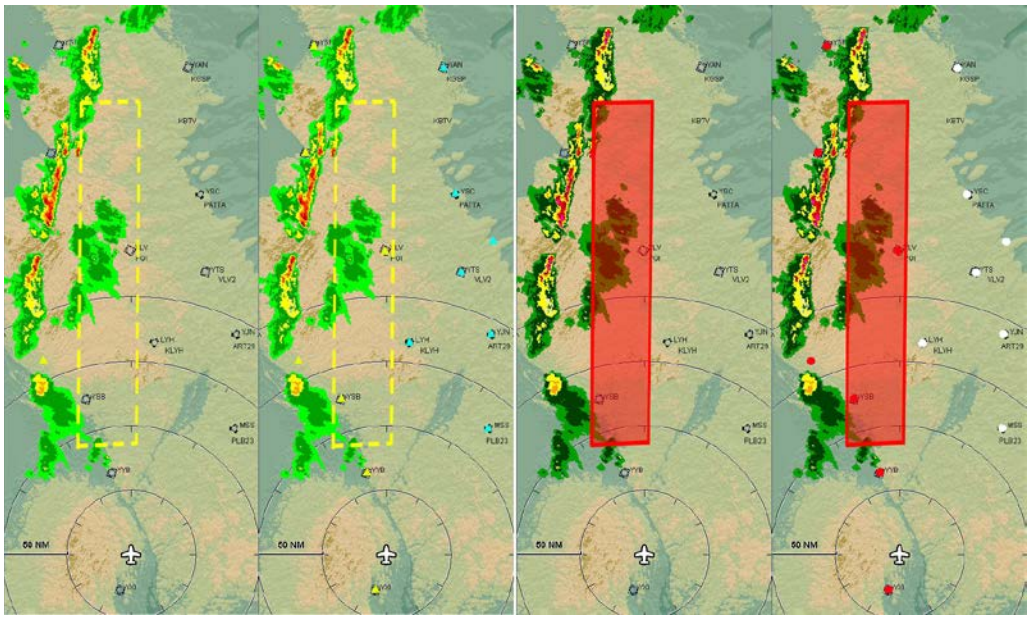


Figure 30. Illustration of the METAR Control condition stimuli for on/off METAR changes: (a) Control D1 display with unfilled METAR symbols; (b) Control D1 display with 7 yellow and 7 light blue METAR symbols; (c) Control D3 display with 7 red and 7 white METAR symbols; and (d) Control D3 display with red METAR symbols

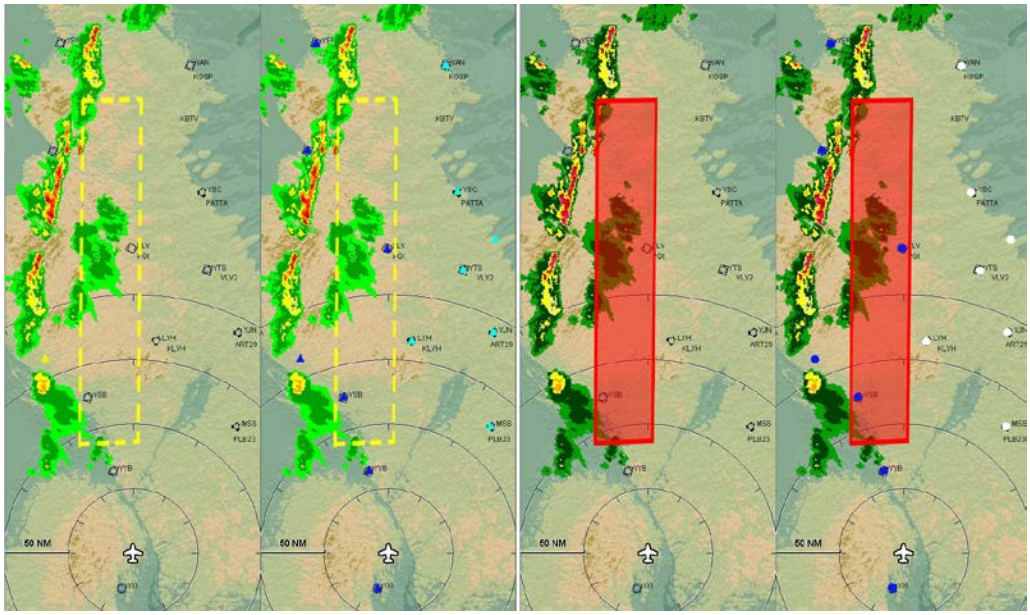


Figure 31. Illustration of the METAR Enhanced condition stimuli for on/off METAR changes: (a) Enhanced D1 display with unfilled METAR symbols; (b) Enhanced D1 display with 7 dark blue and 7 light blue METAR symbols; (c) Enhanced D3 display with unfilled METAR symbols; and (d) Enhanced D3 display with 7 dark blue and 7 white METAR symbols

In a first analysis, we assessed the main effect of salience on discriminability. Figure 32 shows the posterior discriminability (*d*) and bias (*c)* for the effect of salience on the change-detection performance for METAR symbols (color change plus on/off trials) in the Control and Enhanced conditions.

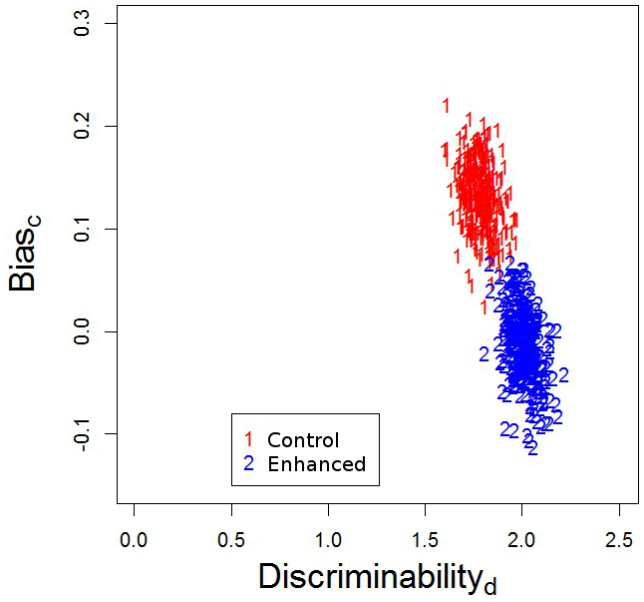


Figure 32. The posterior discriminability (d) and bias (c) for the Control and Enhanced METAR conditions

As shown in Figure 33, the mean posterior *d* for the Control condition is 1.8, whereas it is 2.0 for the Enhanced condition. The contrast on the difference is credible (i.e., the value 0 is not located within the 95% HDI), with a mean posterior difference of -0.22. There was also a difference in bias between the Control (mean *c* = 0.13; 95% HDI from 0.06–0.2) and the Enhanced (mean *c* = -0.01; 95% HDI from -0.08–0.05) conditions. Participants in the Control conditions had a greater propensity for responding "NO" during the trials. The contrast on the *c* difference was credible, with a mean *c* difference of 0.14 (95% HDI from 0.05–0.23).

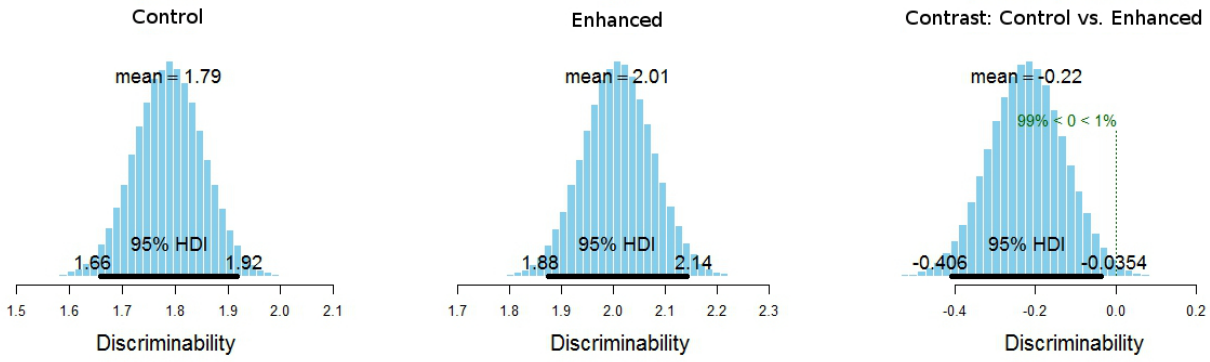


Figure 33. The posterior discriminability and contrast for the Control and Enhanced METAR symbols

In the Ahlstrom and Suss (2014) study, researchers found a credible difference in discriminability between METAR triangle symbols and METAR circle symbols, with circles, on average, yielding higher discrimination performance than triangles. Here, we also assessed whether the discriminability was lower for D1 trials (triangles) compared with D3 trials (circles). The SDT analysis revealed that the mean discriminability was 1.82 (95% HDI from 1.7–1.9) for triangles and 1.92 (95% HDI from 1.8–2.0) for circles. Although the discriminability is slightly higher for circles, the contrast on triangles and circles was not credible (*M* difference = -0.1, 95% HDI from -0.25–0.05). Therefore, in the present experiment, participant discrimination accuracy was the same for METAR triangle and METAR circle symbols.

3.14.6.1 Age Effects on Discriminability for METAR Symbols

Figure 34 shows the outcome of a regression analysis using participant age (*x*) and *d'* (*y*), based on the METAR data. The credible regression lines are tightly clustered, and they all have a negative slope. This means that the discrimination accuracy for METAR symbols decreases with an increasing age.

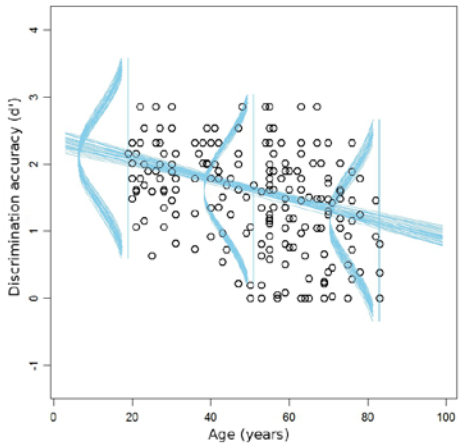


Figure 34. Credible regression lines and noise distributions for the prediction of METAR d' from age

Figure 35 shows the mean posterior outcome for the intercept (β_0), the slope (β_1), and the scale (σ) parameters. The intercept has a mode of 2.38, and the credible slope has a posterior mode of -0.0152. This means that for each year of age, the model predicts that discrimination performance (*d'*) for *METAR* symbols decreases by -0.0152.

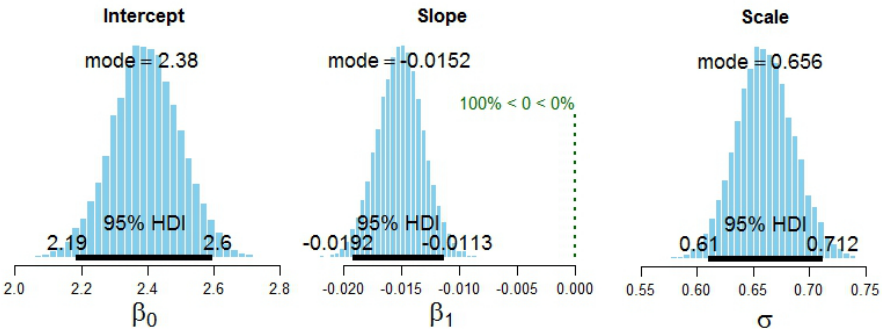


Figure 35. The posterior intercept (β0), slope (β1), and scale (σ) parameters for METAR trials

Figure 36 shows the posterior distributions of predicted mean values of *d'* (*y*) for $x_1 = 20$ and $x_2 = 70$. For $x_1 = 20$ years, the model predicts a mean *d'* of 2.09, and for $x_2 = 70$ years, the model predicts a mean *d'* of 1.33.

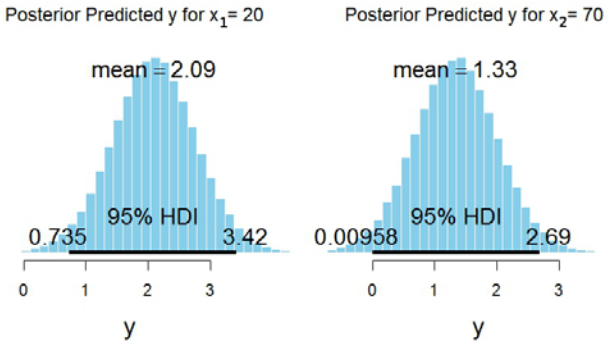


Figure 36. The posterior predicted distributions and mean d' values for METAR symbols with x1 = 20 years and x2 = 70 years

When analyzing the D1 color change trials separately, we find a differential effect for the younger and older pilots. Although there was no credible difference in discriminability between the METAR symbol D1 Control and METAR symbol D1 Enhanced versions for the older pilots (Control: *M* = 1.6, 95% HDI from 1.39–1.86; Enhanced: *M* = 1.5; 95% HDI from 1.17–1.8; mean difference = -0.15, 95% HDI from -0.22–0.53), the younger pilots exhibit an increased discriminability in the Enhanced condition (*M* = 2.94, 95% HDI from 2.38–3.5) compared with the Control condition (*M* = 2.2, 95% HDI from 1.93–2.5). This difference was credible with a mean difference of -0.70 (95% HDI from -1.35–-0.06).

We also find a differential effect for the D3 METAR color change trials. Here, there was no credible difference between the Control versus Enhanced trials for the older pilots (Control *M* = 1.6, 95% HDI from 1.3–1.9; Enhanced *M* = 1.5, 95% HDI from 1.12–1.9; *M* difference = 0.10, 95% HDI from -0.41–0.62). For the younger pilots, however, there was a higher discriminability for the Control condition stimuli (*M* = 3.1, 95% HDI from 2.46–3.75) compared with the Enhanced condition stimuli (*M* = 2.17, 95% HDI from 1.9–2.45) with a credible mean difference of 0.92 (95% HDI from 0.23–1.63).

For the D1 METAR on/off changes, the older pilots exhibited an increased discriminability for the Enhanced condition (*M* = 1.6, 95% HDI from 1.2–2.08) compared with the Control condition (*M* = 1.05, 95% HDI from 0.7–1.36) with a credible mean posterior difference of -0.6 (95% HDI from -1.11–-0.05). For the younger pilots, however, there was no credible difference in discriminability between the Control (*M* = 2.1, 95% HDI from 1.8–2.4) and the Enhanced (*M* = 2, 95% HDI from 1.7–2.2) conditions (contrast *M* =-.14, 95% HDI from -0.24–0.5).

For the D3 METAR on/off trials, there was no credible difference between the Control (*M* = 1.4, 95% HDI from 1.08–1.7) and Enhanced (*M* =1.24, 95% HDI from 0.8–1.6) conditions for the older pilots (*M* difference = 0.148, 95%HDI from -0.36–0.65) or the younger pilots (Control *M* = 2, 95% HDI from 1.8–2.4; Enhanced *M* = 1.9, 95% HDI from 1.7–2.2; *M* difference = 0.13, 95% HDI from -0.26–0.52).

In summary, there is a credible main effect of color (i.e., salience) on discriminability where participants exhibited a greater discriminability for METAR symbol changes in Enhanced displays than METAR symbol changes in Control displays. However, we also find a main effect of age on discriminability where the discrimination accuracy for METAR symbols decreases with an increasing age. In addition, there are also differential effects of age on discriminability for METAR symbols. For the D1 (METAR triangles) and D3 (METAR circles) color trials, there is no credible effect of the Enhanced stimuli on discriminability for older pilots. Conversely, younger pilots exhibit an increased discriminability for the Enhanced symbols in the D1 condition but not in the D3 condition. For the METAR D1 on/off trials, the older pilots exhibit an increased discriminability for the Enhanced condition. For the younger pilots, there is no credible difference between the Control and Enhanced condition. For the D3 on/off condition, there is no effect of the Enhanced METAR stimuli for the older pilots or the younger pilots. Finally, we found no credible difference in discriminability between the METAR triangle (D1) and the METAR circle (D3) conditions.

3.14.6.2 Response Time for METAR Displays

For the response time analysis, we were mainly interested in the main effect of color on response time, and, therefore, we pooled the D1, D3, color change, and on/off change data for the Control and Enhanced conditions separately using all response times for each subject (signal trials only). Figure 37 shows the outcome of the analysis for the METAR Control (see Figure 37(a–c) and METAR Enhanced (see Figure 37[d–f]) conditions. As is clear from the figure, the mean response times for the two conditions are very similar, with a mean of ex-Gaussian = 0.84 sec for the Control condition and 0.88 sec for the Enhanced condition. The contrast on this difference was not credible, because the value 0 was located within the 95% HDI (mean difference = -0.032; 95% HDI from -0.08–0.021).

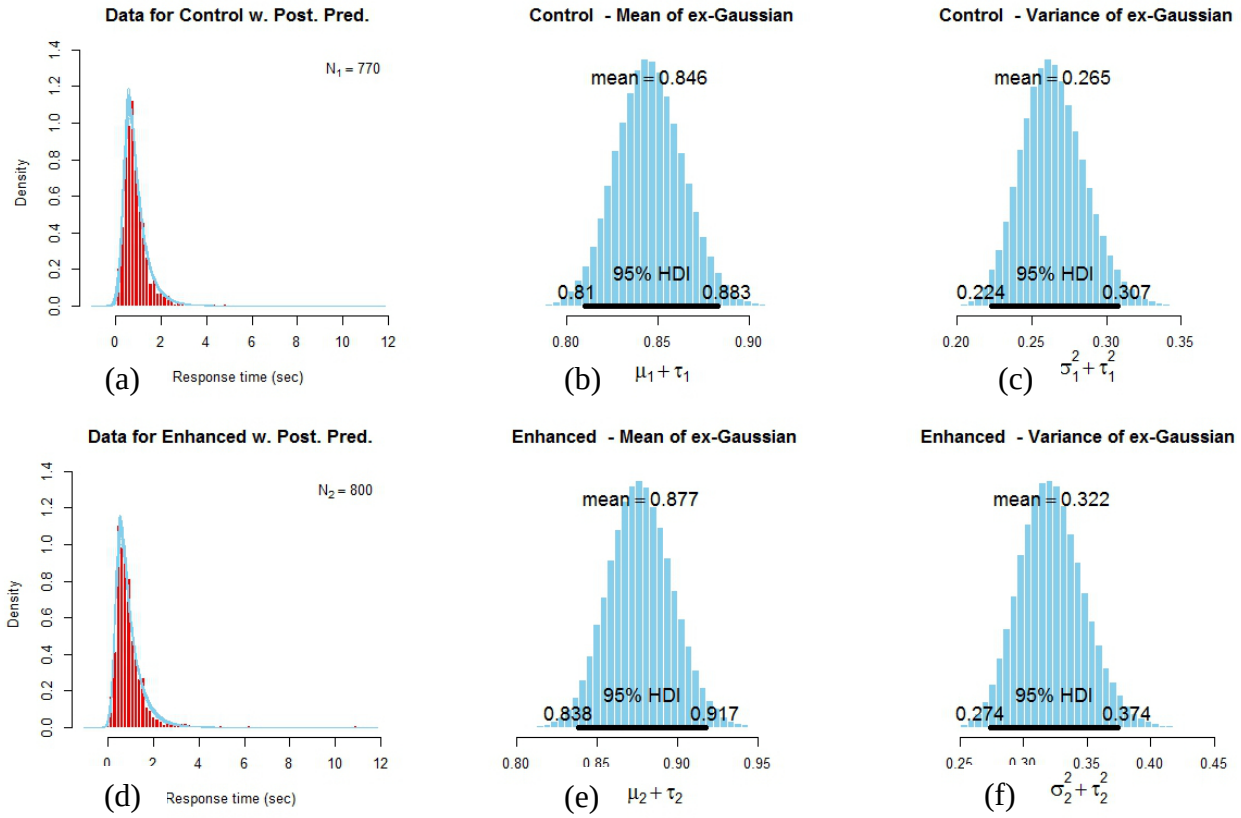


Figure 37. METAR response time data (a & d) with the mean (b & e) and variance (c & f) of the ex-Gaussian distribution for the Control (a–c) and Enhanced (d–f) conditions

In summary, there are no credible differences in response time between the METAR Control and Enhanced conditions.

3.14.6.3 Age Effects on Response Time for METAR Symbols

Figure 38 shows the outcome of a METAR regression analysis using participant age (*x*) and response time (*y*).

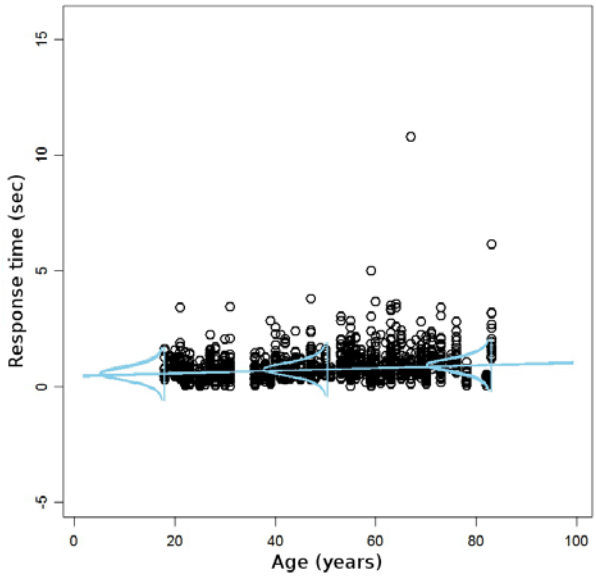


Figure 38. Regression lines and noise distributions for the prediction of METAR response time from age

As show in Figure 39, the intercept has a mode of 0.45, and the credible slope has a posterior mode of 0.006. This means that for each year of age, the response time for discriminability of METAR symbols increases by 0.006 sec.

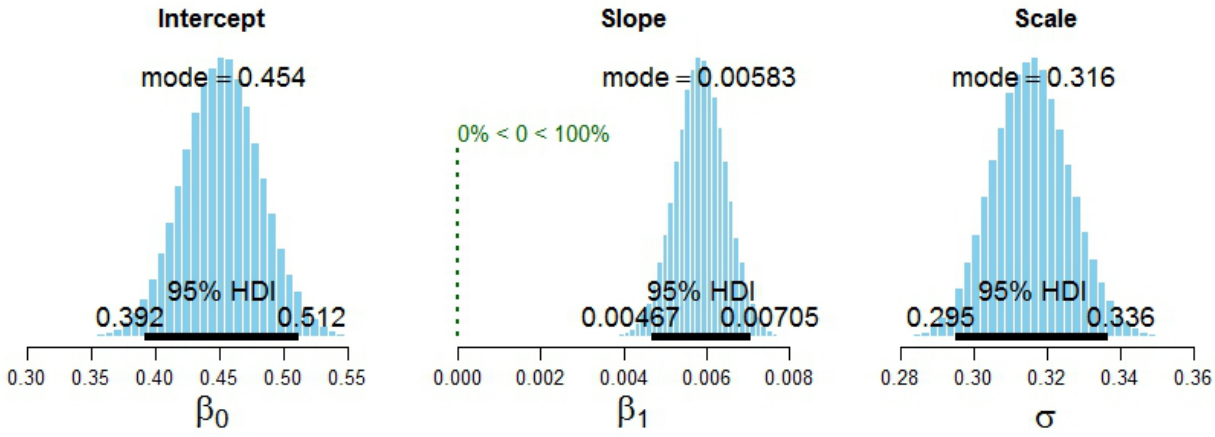


Figure 39. The posterior intercept (β0), slope (β1), and scale (σ) parameters for METAR response time

Figure 40 shows the posterior distributions of predicted mean values of response time (y) for x_1 = 20 and x_2 = 70. For x_1 = 20 years, the model predicts a mean response time of 0.6, and for x_2 = 70 years, the model predicts a mean response time of 0.86.

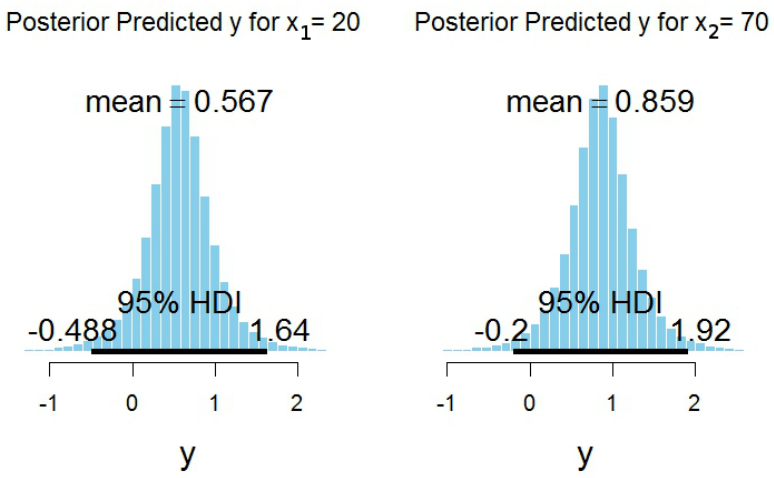


Figure 40. The posterior predicted distributions and mean response time values for METAR symbols with x1 = 20 years and x2 = 70 years

In summary, a regression analysis using participant age and response time revealed that for each year of age, there is an associated increase in response time for discriminability of METAR symbols.

3.14.7 Time-Stamp Information

During Experiment 1, we used two different time stamp colors. For the D1 and D3 Control conditions, time-stamp information was presented in black text at the top of the displays. For the Enhanced D1 and D3 conditions, the time-stamp information was presented in a blue color (enhanced salience) at the top of the displays. Figure 41 shows the *Control* ((a): D1; (b): D3) and Enhanced ((c): D1; (d): D3) time-stamp conditions. The salience differences between the Control and Enhanced conditions are shown in Table 3.

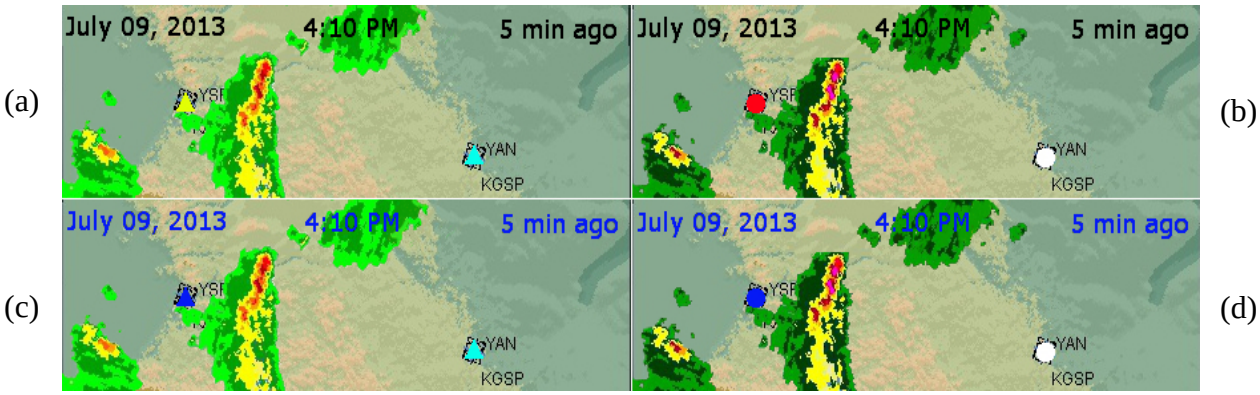


Figure 41. The Control ((a): D1; (b): D3) and Enhanced ((c): D1; (d): D3) time-stamp conditions

During Experiment 1, the signal trials (change) for the Control condition displayed the top-left image (D1) or the top-right image (D3) as Image 1 or Image 2, and compared this to an image that looked the same but was lacking a time stamp. The same pairwise comparison was used for

the D1 and D3 signal trials for the Enhanced condition. Noise trials for both conditions were created by displaying the same image for Images 1 and 2.

Figure 42 shows the posterior discriminability (*d*) and bias (*c)* for the main effect of salience on the change-detection performance for time stamps in the Control and Enhanced conditions. As we can see in the figure, the discrimination performance is very low for both conditions with performance for the Control condition at random guessing.

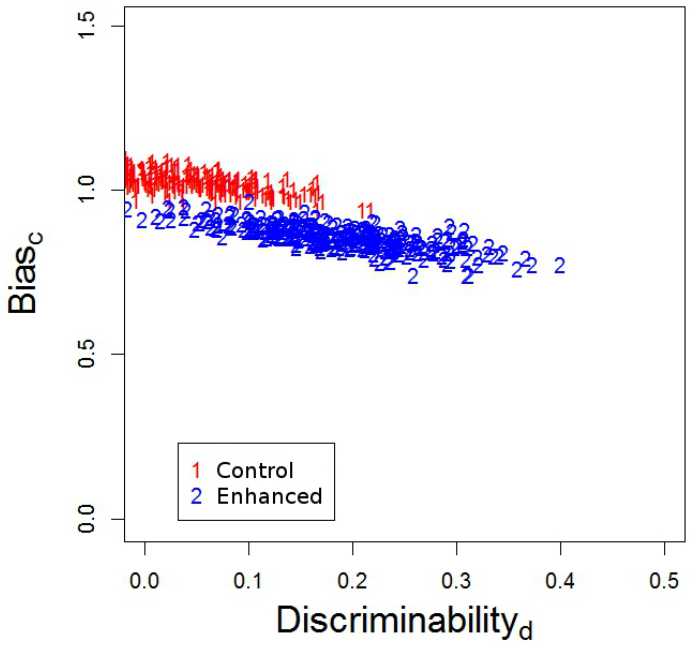


Figure 42. The posterior discriminability (d) and bias (c) for the Control and Enhanced time-stamp conditions

As shown in Figure 43, the mean posterior *d* for the Control condition is 0.009, whereas it is 0.184 for the Enhanced condition. Although the mean discriminability for the Enhanced condition is higher than the Control condition, the contrast is not credible because the value 0 is located within the 95% HDI (mean posterior difference of -0.18).

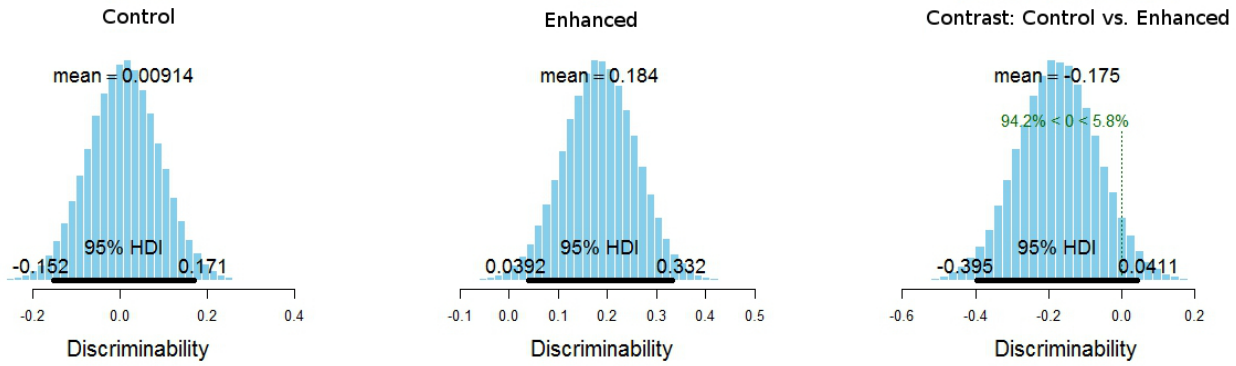


Figure 43. The mean posterior d and contrast for the Control and Enhanced time-stamp conditions

In summary, participants had great difficulty in discriminating signal and noise trials for the time-stamp displays, regardless of the time-stamp color. The discrimination performance is very low for both the Control and Enhanced conditions, with performance for the Control condition at random guessing.

3.14.7.1 Age Effects on Discriminability for Time-Stamp Information

Figure 44 shows the outcome of a regression analysis using participant age (*x*) and *d'* (*y*) for all the time-stamp trials.

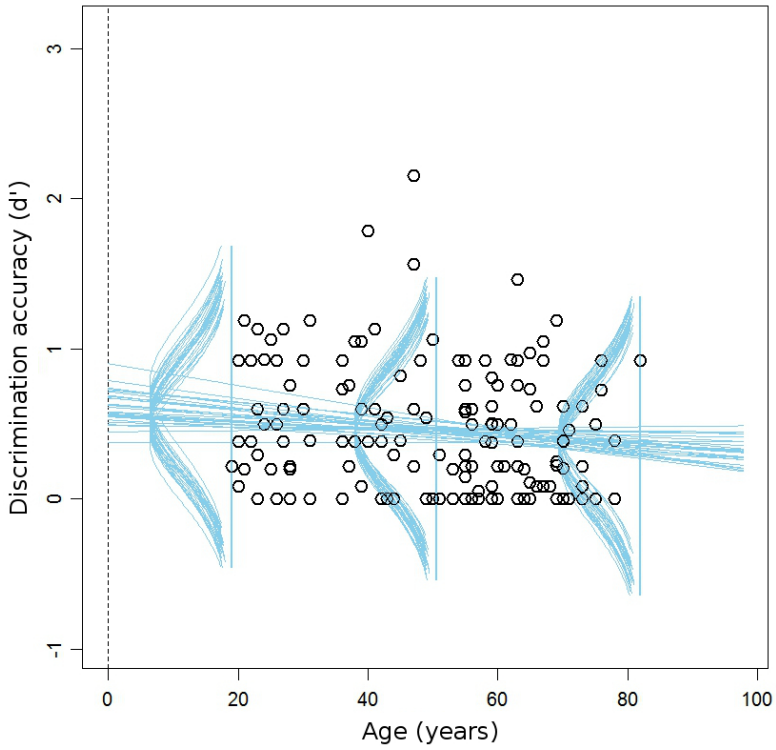


Figure 44. Linear regression lines and noise distributions for the prediction of d' from age for time-stamp trials

As shown in Figure 45, the intercept has a mode of 0.61, and the slope has a posterior mode of -0.003. However, because the value 0 is included in the 95% HDI, the slope is not credible. A slope of 0 is among the credible outcomes. Therefore, there are no credible effects of age on time-stamp discriminability. The discriminability is equally low for all levels of age.

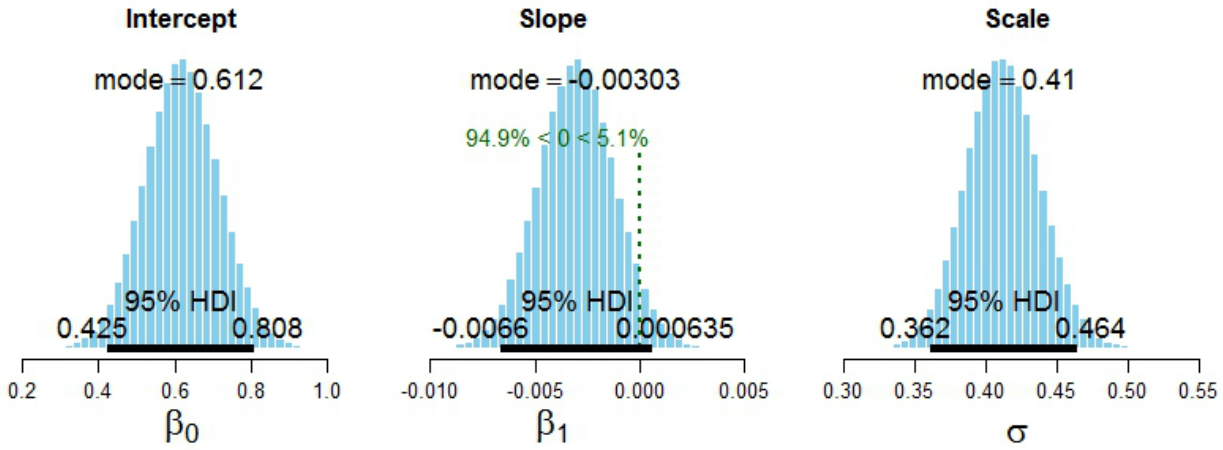


Figure 45. The posterior intercept (β0), slope (β1), and scale (σ) parameters for the prediction of time-stamp d'

In summary, a linear regression analysis on the effect of age on time-stamp discriminability showed no credible effect of age. Participant performance is equally low regardless of age.

3.14.7.2 Response Time for Time-Stamp Information

For the time stamp response time analysis, we were mainly interested in the main effect of color on response time, and we therefore pooled all the data for the Control and Enhanced conditions separately using all response times for each subject (signal trials only). Figure 46 shows the outcome of the analysis for the time-stamp Control (see Figure 46(a–c)) and time-stamp Enhanced (see Figure 46(d–f)) conditions. As is clear from the figure, the mean response times for the two conditions are similar with a mean of ex-Gaussian = 1.03 sec for the Control condition and 1.12 sec for the Enhanced condition.

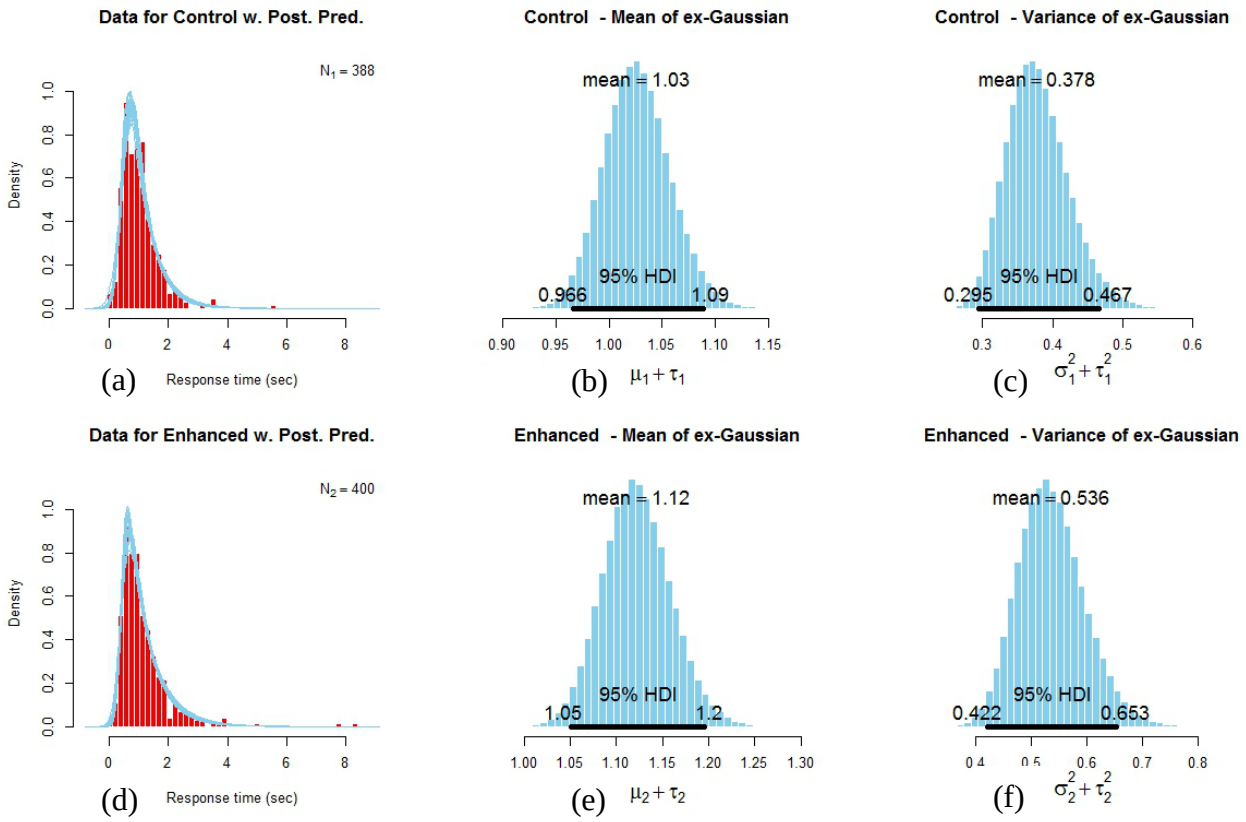


Figure 46. Time stamp response time data (a & d) with the mean (b & e) and variance (c & f) of the ex-Gaussian distribution for the Control (a–c) and Enhanced (d–f) conditions

However, a contrast showed this difference to be credible with a mean posterior difference of -0.10 (95% HDI from -0.19 to -0.001). Although credibly different, the mean posterior effect size for the difference was only -0.14 (95% HDI from -0.28 to -0.004) with 80% of the posterior distribution located within the ROPE (-0.2–0.2). Therefore, the difference in response time between the Control and the Enhanced displays are essentially the same for practical purposes.

In summary, there is a tiny but credible effect of color on time-stamp response time. However, because of the small effect size for this difference—and the fact that performance is at or close to random guessing—there is no meaningful difference between the Control and the Enhanced condition response times.

3.14.7.3 Age Effects on Response Time for Time-Stamp Information

We also assessed the presence of age effects on response time for the time-stamp change-detection trials. Figure 47 shows the outcome of a regression analysis using participant age (*x*) and *d'* (*y*) on all the time-stamp signal trials.

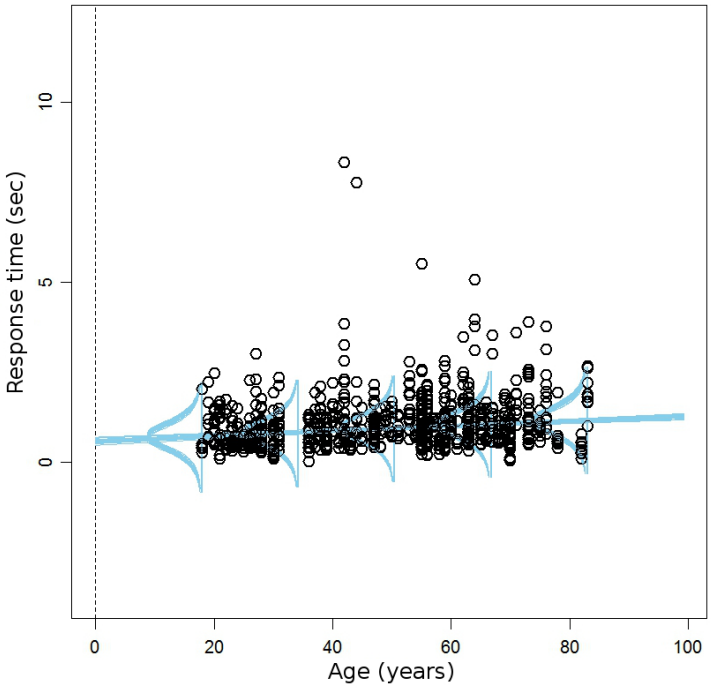


Figure 47. Time-stamp regression lines and noise distributions for the prediction of response time from age

Figure 48 shows the mean posterior outcome for the intercept (β_0), the slope (β_1), and the scale (σ) parameters. The intercept has a posterior mode of 0.60, and the credible slope has a posterior mode of 0.006.

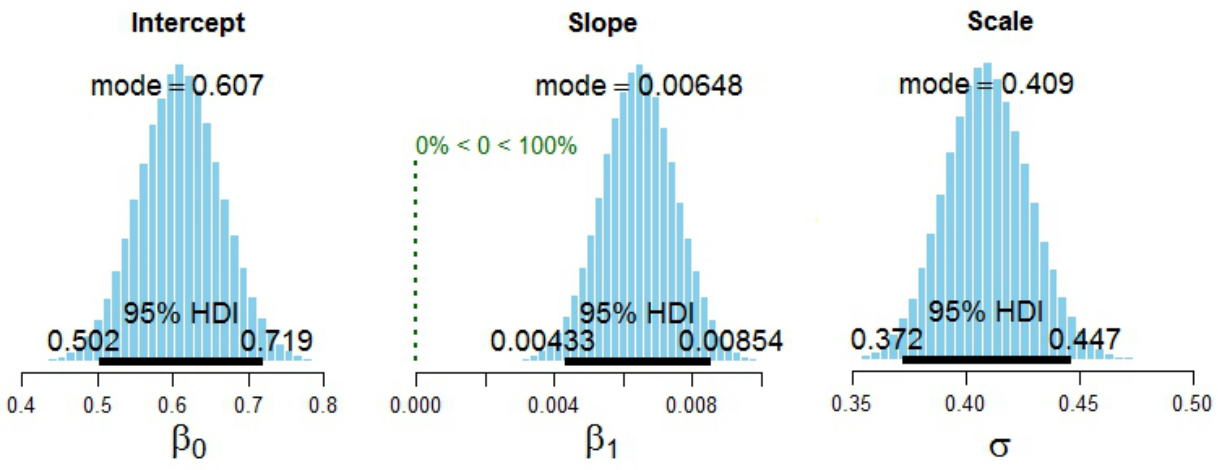


Figure 48. The posterior intercept (β0), slope (β1), and scale (σ) parameters for the prediction of time-stamp response time

Figure 49 shows the posterior distributions of predicted mean values of response time (y) for x_1 = 20 and x_2 = 70. For x_1 = 20 years, the model predicts a mean response time of 0.73, and for x_2 =70 years, the model predicts a mean response time of 1.06.

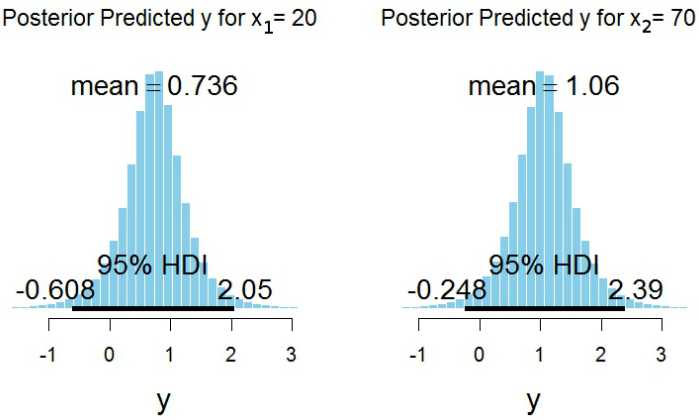


Figure 49. The posterior predicted distributions and mean response time values for time-stamp information with x1 = 20 years and x2 = 70 years

In summary, the mean *d* for discrimination of time-stamp signal and time-stamp noise trials is very low, equaling the performance from random guessing. Although the performance is nominally higher for the Enhanced condition, the difference is minimal, which shows that an increase in the time-stamp salience does not improve performance beyond random guessing. Furthermore, at this low level of discriminability, there is no meaningful effect of age on change-detection performance.

4 EXPERIMENT 2

Ahlstrom and Suss (Ahlstrom & Suss, 2015) examined change-detection performance for weather symbols, like METARs, lightning strikes, precipitation areas, and SIGMET outlines. They found that change-detection accuracy varies depending on the symbol color and shape. As shown in Figure 9, however, there can be many other features in complex weather displays that provide important information. For example, solid lines or line segments are important because they define different classes of airspace and airspace boundaries. Line segments also define routes, airports, runways, airways, state lines, latitude and longitude scales, and military operations areas, to mention a few. Furthermore, the various lines and line segments are depicted in varying colors, thicknesses, and orientations.

Although we have empirical data on change-detection performance for weather symbols, less is known about observer sensitivity to changes in line segments in complex weather displays. This is an important topic because line segments are used to display user-defined routes and can be an important part of graphical display notifications for both air traffic controllers and pilots (Ahlstrom, 2015; Ahlstrom & Jaggard, 2010).

The purpose of Experiment 2 was to assess the effect of color (i.e., salience) on discriminability of line segments in complex weather displays.

4.1 Method

4.2 Participants

Twenty-four males volunteered to participate in Experiment 2. None of the participants participated in Experiment 1. Thirteen of the twenty-four participants were pilots, and we present their background information in Table 4.

Table 4. Descriptive Characteristics of Study Participants

			Flight hours accrued					
	Age (years)		Total		Instrument		Instrument – last 6 months	
N	Mdn	Range	Mdn	Range	Mdn	Range	Mdn	Range
13	55	26–65	1490	230–9,000	60	4–800	12	0–80
11	37	24–60	N/A	N/A	N/A	N/A	N/A	N/A

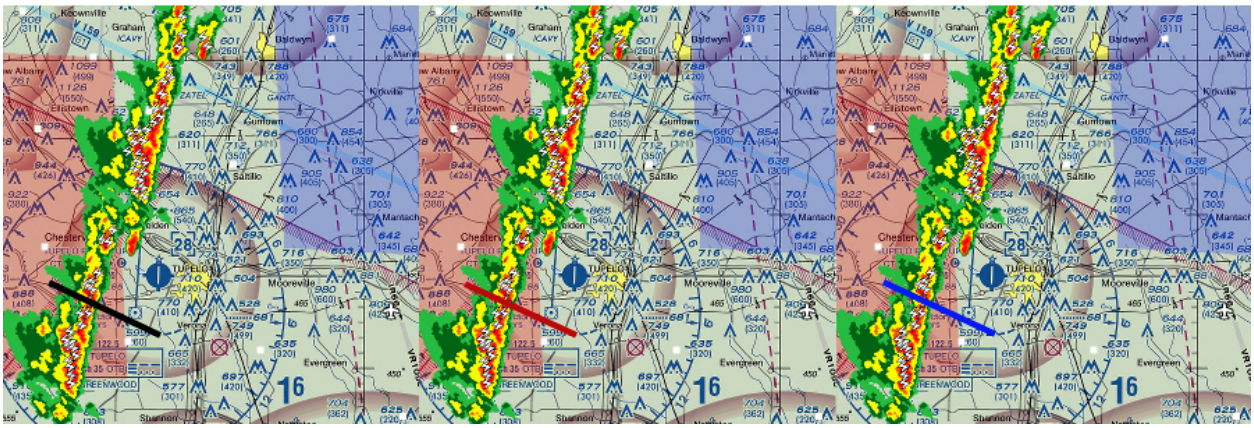
Similar to Experiment 1, participants were excluded from participation in Experiment 2 if they had a personal or familial history of epilepsy. No participants reported a history of epilepsy.

4.3 Experimental Stimuli

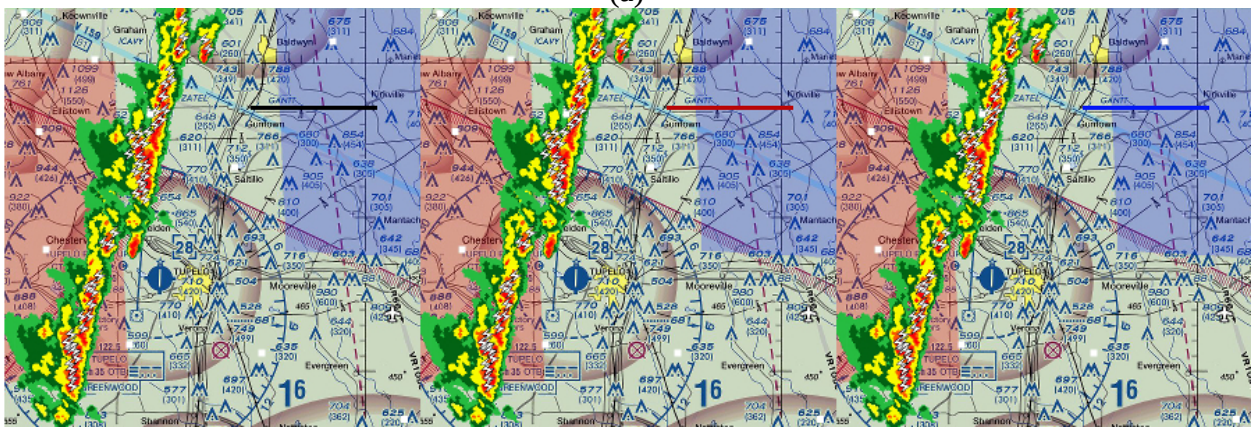
In Experiment 2, we evaluated change-detection performance for line segments that were superimposed on a complex weather display. That is, the line segments were superimposed on a VFR background map that also contained precipitation, special-use airspace (SUA) areas, and lightning information. At a viewing distance of 64 cm, the viewing angle of the experimental images (600 × 600 pixels) subtended 13 degrees horizontally and 13 degrees vertically.

To create some variation in the stimuli, we varied the line segment color (i.e., black, red, and blue), thickness (i.e., thin = 3 pixels, medium = 6 pixels, and thick = 9 pixels), orientation (i.e., vertical, horizontal, and oblique), and position (i.e., upper left, lower left, upper right, and lower right quadrants). To create some variation in the background maps, we rotated an original map in steps of 90 degrees to create four background configurations. These image manipulations resulted in a 4 (Background) x 3 (Line width) x 3 (Line Orientation) x 4 (Line Position) x 2 (Trial type) = 288 images for each of the 3 line colors. There was no replication of the signal trials; participants only saw each signal (change) trial once. Only the 288 noise (no change) trials were replicated throughout the experiment.

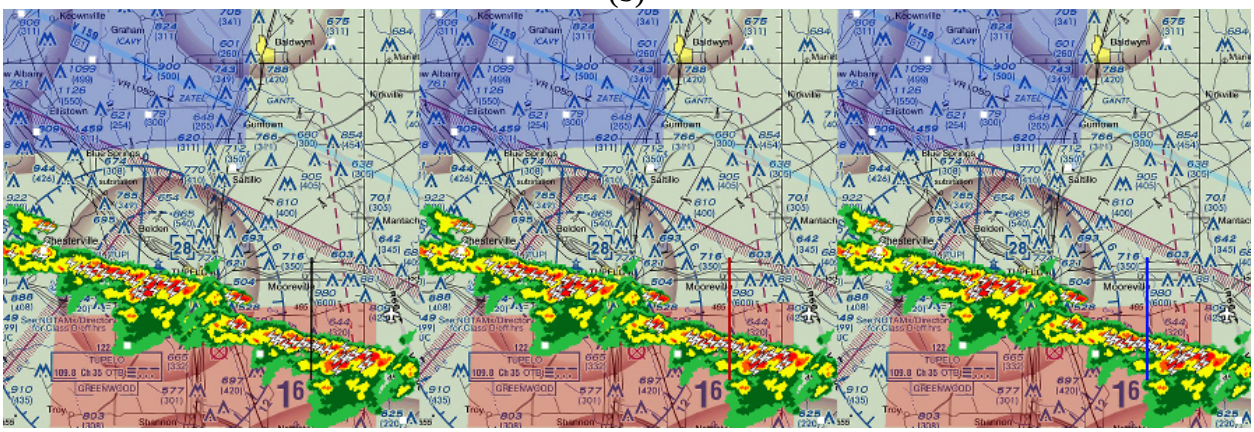
Although these stimuli manipulations allow us to analyze the effect of background, orientation, and trial types, for the present analysis, we are only interested in the main effects of line color and thickness on change-detection performance. Figure 50 shows examples of the line stimuli in different colors, locations, orientations, and thicknesses.



(a)



(b)



(c)

Figure 50. Example weather display using a VFR background map: (a) thick oblique lines, lower-left quadrant; (b) medium-thickness horizontal lines, upper-right quadrant; and (c) thin vertical lines, lower-right quadrant

4.4 Line Salience

Using the procedure outlined in section 3.8 above, we created three line colors that varied in salience. Table 5 shows the average salience for each line color used in Experiment 2.

Table 5. Color and Salience Values for the Three Line Colors

Line color name	RGB	Mean line salience
Black	RGB (0, 0, 0)	100 (SD=7.8)
Red	RGB (185, 9, 9)	165 (SD=2.5)
Blue	RGB (4, 27, 245)	251 (SD=1.3)

The black lines have the lowest salience followed by the red lines. The blue lines have the highest salience. Based on previous research and the color results from Experiment 1, we predict that discrimination performance will be highest for the blue lines and lowest for the black lines.

4.5 Independent Variable

The independent variable in Experiment 2 is the line segment color (i.e., black, red, and blue). Color is a between-subjects variable, meaning that each participant was randomly allocated to one of three line-color conditions.

4.6 Dependent Variables

The dependent variables are the observed counts of hits, false alarms, misses, correct rejections, and the time it took the participant to respond.

4.7 Procedure

After reading and signing the Informed Consent Statement, participants completed a biographical questionnaire. Next, the researcher explained to the participant that the task instructions would be presented on-screen in a self-paced manner. After following the on-screen instructions, participants first completed the practice trials, followed by the experimental trials. Participants were instructed to respond as quickly as possible without making any errors. Participants could pause for as long as they wanted after each trial before continuing the experiment. There was no feedback during the practice or experimental trials.

Participants began each trial of the change-detection task by pressing the space bar on the keyboard. For each trial, the participant responded to whether or not ("YES" or "NO") a change occurred between the two images. The first 20 trials were practice trials, divided into 16 signal (change) trials and 4 noise (no change) trials. Following the practice trials, the same process was repeated for an additional 576 trials. Half of the 576 trials were noise trials (i.e., no change) with the remainder of the trials being signal trials (i.e., change). For each participant, we presented the trials in random order.

For Experiment 2, we used the same one-shot technique as used for Experiment 1. During the experiment, participants initiated each trial by pressing the spacebar on the computer keyboard. First, a grey screen with a central fixation cross appeared for 1000 ms, and then Image 1 was

displayed for 100 ms. Image 1 was then replaced by a blank, grey screen for 1000 ms, after which Image 2 was displayed for 100 ms (Keshvari, van den Berg, & Ma, 2013; Wilken & Ma, 2004). Participants provided a response by pressing one of two buttons on the keyboard to indicate either "Yes" (change detected) or "No" (no change detected). The software waited for a response for 60 s; if the participant did not enter a response within that period, the trial ended automatically and recorded a "no response."

4.8 Data Analysis

We used the same analysis methods for Experiment 2 as used for Experiment 1.

4.9 Results

In the following section, we present the outcome of the line discrimination analysis. First, we assess the main effect of age on line discriminability. Second, we assess the main effect of line color (i.e., salience) on discriminability. Third, we present an analysis of the effect of line thickness on discriminability. Lastly, we present a response time analysis for each of the three line colors.

4.9.1 Effects of participant age on line discrimination accuracy

Similar to Experiment 1, we used a robust linear regression analysis to assess the main effects of participant age on line discriminability.

Figure 51 shows the outcome of the regression analysis. We have participant age on the *x*-axis and accuracy (*d'*) on the *y*-axis discrimination. The black circles represent participant *d'* values (three values per participant for the three line-thickness conditions). The blue horizontal lines represent regression lines, and the three vertical distributions are superimposed *t*-distributed noise distributions.

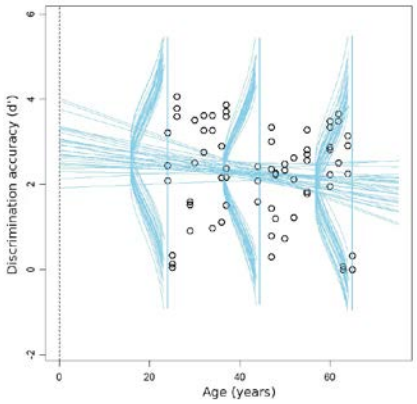


Figure 51. Regression lines and noise distributions for the prediction of d' from Age

Figure 52 shows the mean posterior outcome for β_0 (intercept), β_1 (slope), and the scale parameter (σ). The intercept has a posterior mode of 2.88, which is the value of *y* when $x = 0$. The slope has a mode of -0.0139. The scale has a mode of 1.13. However, because the value 0 is located within the 95% HDI, the slope is not credible. It means that a slope of 0 is a likely

predicted outcome, and a slope of 0 implies a complete lack of a relationship between variable *y* (*d'*) and predictor *x* (age). Therefore, there is no credible effect of participant age on discrimination accuracy in the line discrimination data.

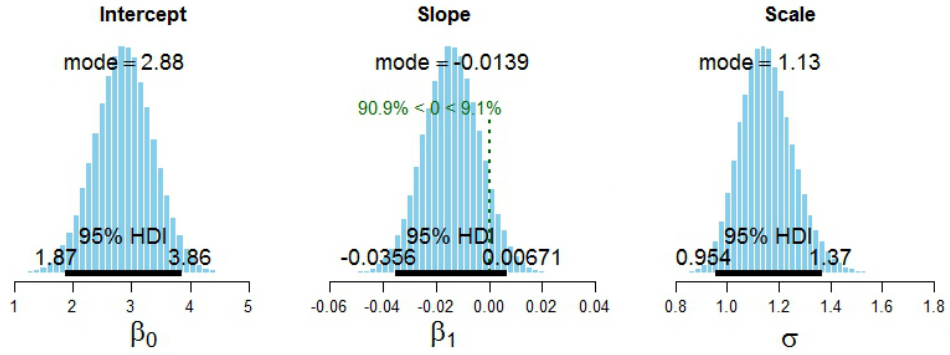


Figure 52. The posterior intercept (β0), slope (β1), and scale (σ) parameters from the robust linear regression analysis

4.9.2 Line color

Figure 53 shows the discriminability and bias outcome for the black, red, and blue line conditions. The blue lines have the highest discriminability (*d*), followed by the red lines. The black lines have the lowest discriminability.

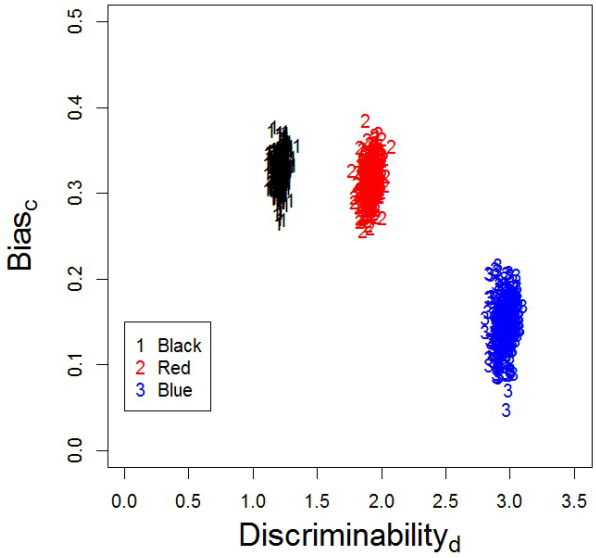


Figure 53. Posterior distributions for the black, red, and blue line colors in terms of discriminability (d) and bias (c)

The black and red lines also have a higher positive value of the bias (*c*) compared with the blue lines, meaning that participants displayed a greater bias in responding "NO" to the black and red line stimuli.

Figure 54 shows the mean posterior discriminability for the three line colors along with their contrasts. All three contrasts are credible (i.e., the value 0 is not included in the 95% HDI), with the discriminability of the red lines (mean d = 1.92) being higher than the black lines (mean d = 1.21, mean difference = -0.712 with the 95% HDI from -0.831 to -0.593), the discriminability of the blue lines (mean d = 2.96) being higher than the black lines (mean d = 1.21, mean difference = -1.75 with the 95% HDI from -1.89 to -1.61), and the discriminability of the blue lines (mean d = 2.96) being higher than the discriminability of the red lines (mean d = 1.92, mean difference = -1.04 with the 95% HDI from -1.18 to -0.894).

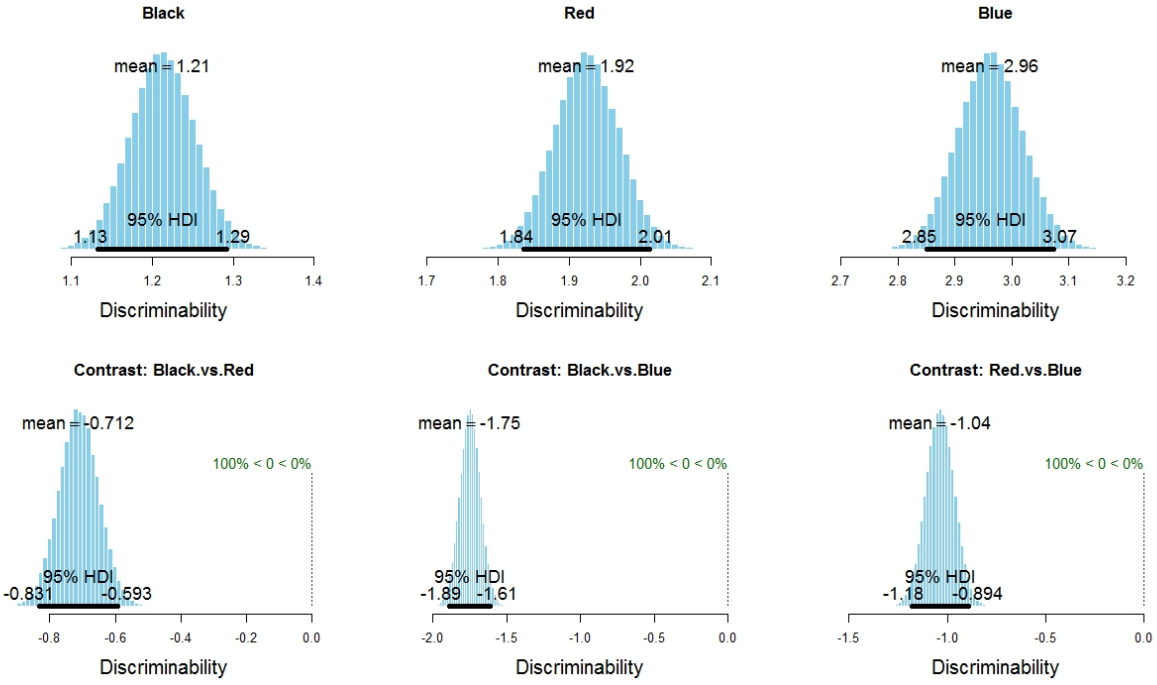
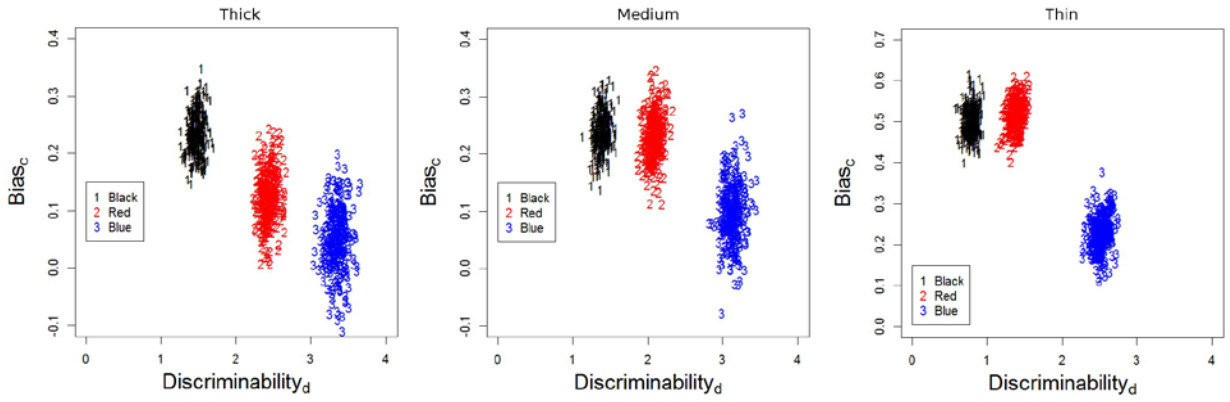


Figure 54. The mean posterior discriminability for the three line colors (top) and their contrasts (bottom)

In summary, there is a main effect of color on line discriminability. Participant discriminability performance exactly follows the relative salience for the black (salience; M = 100, SD = 7.8), red (salience; M = 165, SD = 2.5), and blue lines (salience; M = 251, SD = 1.3). Participants have the lowest discriminability performance for the black lines (d; M = 1.21), followed by the red lines (d; M=1.92), and shows the highest discriminability performance for the blue lines (d; M=2.96). Therefore, in the present change-detection experiment, symbol salience is a predictor of discriminability performance.

4.9.3 Line thickness

We also analyzed the data to assess whether the main effect of color (i.e., salience) is invariant across the three line thicknesses. Figure 55 shows the posterior distributions for the thick (left), medium (middle), and thin (right) line stimuli. As shown, the discriminability index d follows the same trend for all line thicknesses, with the *blue* line having the highest discriminability index, followed by the *red* lines, and with the *black* lines having the lowest discriminability.



(a) (b) (c)

Figure 55. The posterior discriminability and bias for the: (a) thick, (b) medium, and (c) thin line segments

For the thick-line stimuli in Figure 55(a), the discriminability (*d*) for the black lines has a posterior mean=1.48 (95% HDI from 1.34 to 1.62). For the *red* lines, the posterior mean = 2.44 (95% HDI from 2.28–2.61). For the blue lines, the posterior mean = 3.35 (95% HDI from 3.14 to 3.57). All contrasts between the three line colors are credible, with the red lines having a larger *d* than the black lines (mean difference = -0.97; 95% HDI from -1.18 to -0.75), the blue lines having a larger *d* than the black lines (mean difference = -1.88; 95% HDI from -2.13 to -1.62), and the blue lines having a larger *d* than the red lines (mean difference = -0.91; 95% HDI from -1.18 to -0.64). There was also a credible difference in the bias (*c*) between the blue and black lines, with a greater bias in responding "NO" for the black lines than the blue lines (mean difference = 0.19; 95% HDI from 0.059 to 0.32).

For the medium line thickness in Figure 55(b), the *d* for the black lines has a posterior mean = 1.39 (95% HDI from 1.26 to 1.53). The red lines have a mean *d* = 2.08 (95% HDI from 1.92 to 2.24). The blue lines have a mean *d* = 3.13 (95% HDI from 2.93 to 3.33). All contrasts between the three line colors are credible, with a larger *d* for the red lines compared with the black lines (mean difference = -0.69; 95% HDI from -0.90 to -0.48), a larger *d* for the blue lines compared with the black lines (mean difference = -1.73; 95% HDI from -1.98 to -1.49), and a larger *d* for the blue lines compared with the red lines (mean difference = -1.05; 95% HDI from -1.3 to -0.79). There was also a credible difference in the bias (*c*) between the blue and black lines, with a greater bias in responding "NO" for the black lines than the blue lines (mean difference = 0.14; 95% HDI from 0.02 to 0.26).

For the thin lines in Figure 55(c), the *d* for the black lines has a posterior mean = 0.787 (95% HDI from 0.65 to 0.92). The red lines have a mean *d*= 1.38 (95% HDI from 1.24 to 1.53). The blue lines have a mean *d* = 2.53 (95% HDI from 2.4 to 2.7). Again, all contrasts between the three line colors are credible, with a larger *d* for the red lines compared with the black lines (mean difference = -0.6; 95% HDI from -0.8 to -0.4), a larger *d* for the blue lines than the black lines (mean difference = -1.75; 95% HDI from -1.9 to -1.53), and a larger *d* for the blue lines compared with the red lines (mean difference = -1.15; 95% HDI from -1.4 to -0.92). There were also credible differences in the bias (*c*) between the blue and black lines with a greater bias in responding "NO" for the black lines than the blue lines (mean difference = 0.27; 95% HDI from

0.16 to 0.38), and a greater bias in responding "NO" for the red lines than the blue lines (mean difference = 0.29; 95% HDI from 0.17 to 0.4).

In summary, participant discriminability performance follows the same trend for all three line thicknesses (i.e., thin, medium, and thick) with the blue lines yielding the highest discriminability performance followed by the red lines, and with the black lines yielding the lowest discriminability performance.

4.9.4 Response time

The result of the present experiment has shown that line discriminability is affected by the line salience. However, because the discriminability varies for line segments with different salience, there might be a difference in the time it takes to visually encode the three line colors. The logic is that conspicuous line changes will be easier to detect and therefore yield shorter response time compared with less conspicuous symbol changes. Therefore, in the following section, we present an analysis of the main effect of color (i.e., salience) on response time. For the analysis, we used all response times for each subject for the signal trials.

Figure 56 shows the response time data, the mean of the ex-Gaussian distribution, and the variance of the ex-Gaussian distribution for the black (see Figure 56(a–c)), red (see Figure (d–f)), and blue (see Figure 56(g–i)) line conditions.

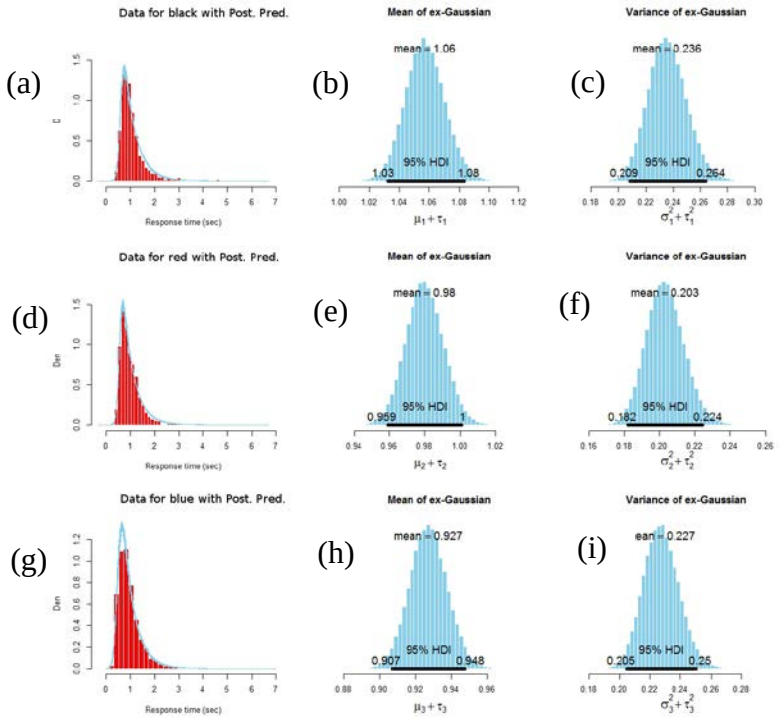


Figure 56. Response time data (a, d, g) with the mean (b, e, h) and variance (c, f, i) of the ex-Gaussian distribution for the black (a–c), red (d–f), and blue (g–i) line conditions

The posterior mean response time for the black lines is 1.06 sec (mean variance = 0.24), red lines 0.98 sec (mean variance = 0.20), and blue lines 0.92 sec (mean variance = 0.22). Figure 57 shows the difference of means for the comparison between the black, red, and blue response time times. All differences are credible (the value 0 is not included in the 95% HDI), showing a longer response time for the black lines compared with the red lines (left; mean = 0.08, 95% HDI from 0.04 to 0.11), a longer response time for the black lines than the blue lines (middle; mean = 0.13, 95% HDI from 0.1 to 0.16), and a longer response time for the red lines compared with the blue lines (right; mean = 0.05, 95% HDI from 0.02 to 0.08).

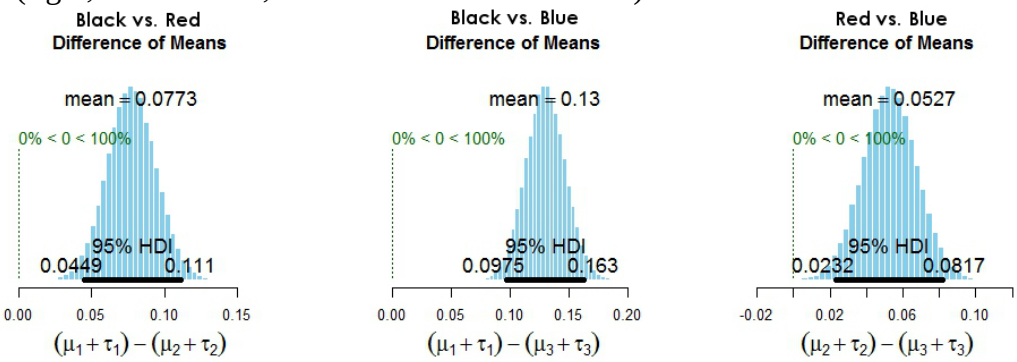
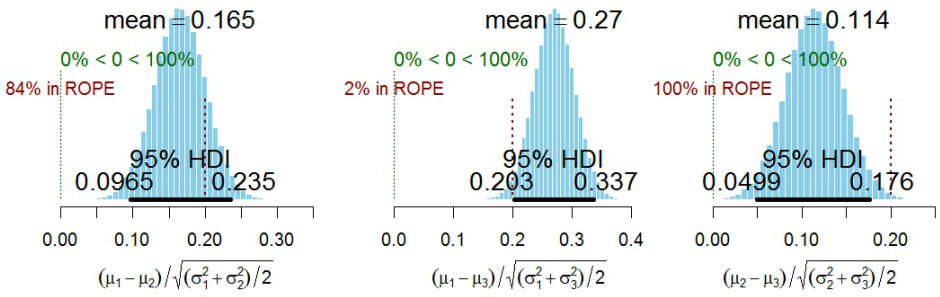


Figure 57. Differences of mean ex-Gaussian response time between the: (a) black and red lines; (b) black and blue lines; and (c) red and blue lines

However, although credibly different, these differences of means are rather small. Furthermore, we are using a large sample for our analysis. This means that we have to look at the effect sizes for the differences of means because it is the size of the effect that is important. Figure 58 shows the standardized effect sizes for the comparison between the black and red (see Figure 58(a); mean = 0.16), black and blue (see Figure 58(b); mean = 0.27), and the red and blue (see Figure 58(c); mean = 0.11).

For the red versus blue effect size (see Figure 58(c)), we have the 95% HDI completely within the ROPE margins (-0.2 to 0.2). Therefore, we declare a null effect for the difference of response time means for the red and blue conditions. Similarly, for the black and red effect size (see Figure 58(a)), we have 84% of the 95% HDI included within the ROPE. This implies a tiny effect, which is unlikely to have any important practical consequences. The effect size for the difference of means between the black and blue conditions (mean = 0.27), however, is small but only shows 2% of the 95% HDI within the ROPE. Therefore, this difference can have a practical importance with the response time for the black lines being 0.27 standard deviations longer than the response time for the blue lines.



(a) (b) (c)

Figure 58. Standardized effect sizes for the differences of means for the: (a) black and red lines; (b) black and blue lines; and (c) the red and blue lines. For this analysis, we are using a ROPE that spans from -0.2 to 0.2. If the 95% HDI is completely within the ROPE, we declare a null effect for practical purposes

In summary, the analysis of line discriminability response times showed credible differences between the black, red, and blue line colors. The blue lines have the shortest response time, followed by red response time, and finally the black lines with the longest response time. However, the effects are generally small, and only the difference between the black line response time and the blue line response time is deemed to have potential practical consequences.

4.9.5 Effects of participant age on line discrimination Response Time

To assess the effect of age on line discriminability response time, we performed a linear regression on the response time data using all trials for each participant. Figure 59 shows the outcome of the linear regression analysis. We have participant age in years on the *x*-axis and response time in seconds on the *y*-axis. As shown by Figure 59, the regression lines have a slight positive slope, indicating that line response times increase with an increasing age.

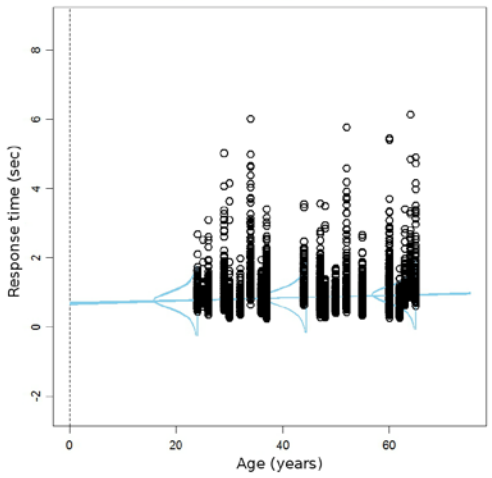


Figure 59. Regression lines and noise distributions for the prediction of line response time from age

Figure 60 shows the mean posterior outcome for β_0 (intercept), β_1 (slope), and the scale parameter (σ). The intercept has a mode of 0.7, which is the value of y when $x = 0$. The credible slope has a posterior mode of 0.00409. This means that as we increase x (age) by 1, y will increase by the value of β_1.

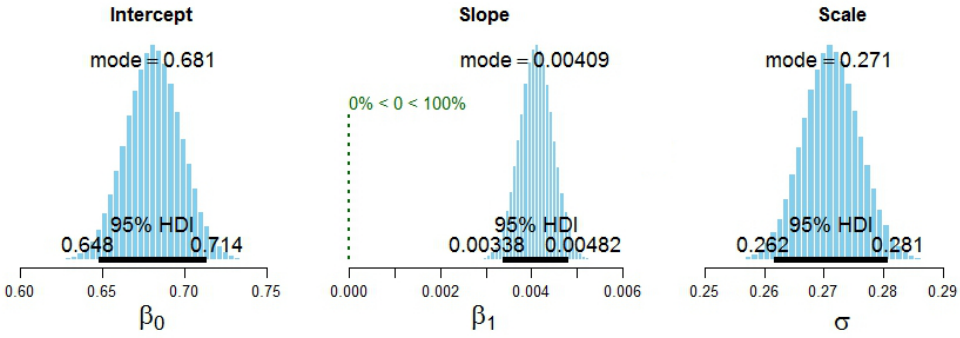


Figure 60. The posterior intercept (β0), slope (β1), and scale (σ) parameters from the linear regression of line response time data and age

Figure 61 shows the posterior distributions of predicted mean values of response time (y) for x_1 = 20 and x_2 = 70 from the MCMC sampling. The model predicts a mean response time of 0.8 seconds for age = 20 and a mean response time of 1.0 seconds for age = 70.

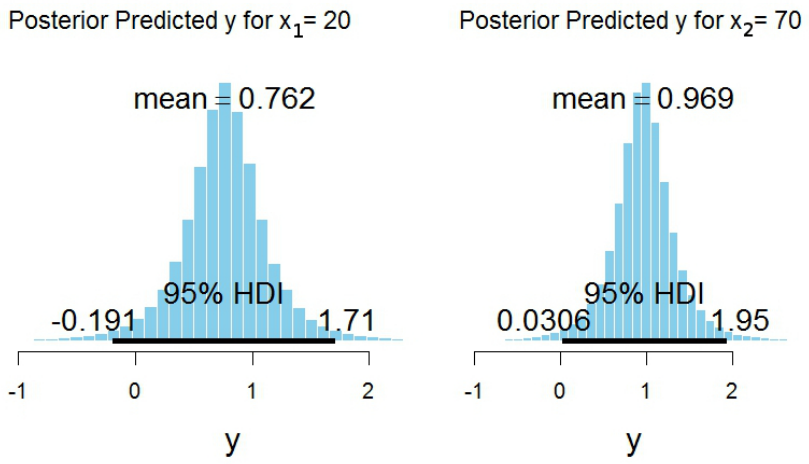


Figure 61. The posterior predicted distributions and mean response time values for line discrimination and x1 = 20 years and x2 = 70 years

In summary, a linear regression model on the response time data predicts an effect of age on response time, with a predicted posterior mean difference of 200 msec between age = 20 and age = 70.

5 DISCUSSION

Previous research has found change blindness effects for commonly used weather display symbols (Ahlstrom & Suss, 2014; Ahlstrom et al., 2015a, 2015b). Using simple change-detection tasks, researchers found that different symbology renderings of the same weather data led to varying performance. Moreover, the change-blindness effect was not restricted to simple change-detection tasks; it also accounted for pilot detection of symbol color changes during simulated flights.

In the present study, we addressed this symbol-detection gap by evaluating the effect of symbol enhancement on pilot visual performance. Specifically, we used symbol salience enrichments to increase pilot change-detection performance. This is in line with current frameworks for visual search and covert attention modeling that stress the importance of saliency for object and scene encoding (Achanta, Hemami, Estrada & Süsstrunk, 2009; Itti & Koch, 2000; Itti, Koch & Niebur, 1998). It is also consistent with previous research on the importance of salience for detection of changes in front-of-pack labels (Becker et al., 2016), improving information visualizations (Jänicke & Chen, 2010), predicting fixation locations (Parkhurst, Law, & Niebur, 2002), and estimating change-detection performance (Verma & McOwan, 2010).

The result of our symbol enhancements was a main effect of symbol salience on change-detection performance. This is because salient symbols stand out against neighboring symbols in an image, thereby capturing attention and enhancing detection. For the line, lightning, and METAR symbol trials, pilot-detection performance was credibly higher for high-salience lines and credibly higher for the salience-enriched symbols (Enhanced) compared with the original symbols (Control). For time-stamp information, however, there was no credible difference in discriminability between the Enhanced and Control conditions. We conclude that the lack of an effect of salience for time-stamp information is entirely due to the time-stamp location. The result for both the Control and the Enhanced time-stamp conditions show near-chance discriminability, similar to the result reported by Ahlstrom and Suss (2014). Currently, many commercial weather products display time-stamp information for displayed weather data (FAA, 2010), and this information is frequently presented in the menu bar or in one of the upper corners of the display (Latorella & Chamberlain, 2002). Because this time-stamp location precludes efficient detection and encoding, display developers need to design time-stamp information that overcomes this deficiency.

Overall, the outcome of the present study shows that we can use symbol salience to enhance discrimination performance. However, the outcome also shows that discrimination performance is not consistent across all age groups. For the lightning and METAR trials, we found that discrimination accuracy decreased with an increased pilot age. In addition, for each year of age there was an increase in response time for discrimination of lightning, METAR, time stamp, and line symbols. Similar effects of age on pilot performance have been documented for a variety of tasks. For example, Hardy and Parasuraman (1997) reviewed previous research of cognitive proficiency and flight performance and summarized pilot age-group differences for perceptual-motor skills, memory, attention, and problem solving. Similarly, Peich, Husain, and Bays (2013) report on cognitive age-related reductions in maintaining associations (i.e., bindings) between individual objects in visual arrays during change detection. Researchers have used random walk models to analyze age effects on information processing like the longer response times for older

adults compared with younger adults (Spieler, 2001). According to Sekuler and Sekuler (2000), older adults have particular difficulties in tasks that require rapid processing, high light sensitivity, and dynamic vision. Unfortunately, these are the very same visual requirements that underlie efficient scan patterns and symbol encoding during single-pilot operations.

Whereas age-related declines in memory and vision can affect many aspects of everyday life, these reductions are also likely to affect the encoding and detection of display changes in the cockpit. There is now sufficient evidence of credible differences in pilot change-detection performance depending on the symbol shape and color. There is also evidence that the detection time and encoding time of weather symbols vary as a function of pilot age. These are important findings that should guide future weather display developments. First, all weather display symbols should allow rapid encoding and detection for pilots of all ages. This is especially important because of the large number of different weather elements that are overlaid on modern multifunctional displays (FAA, 2010) and the relatively large (~30%) population of private pilots that are 60 years of age or older (FAA, 2016). If not taken into consideration, weather symbol changes or weather symbol updates can lead to salience problems where important information fails to visually segregate from less-critical background information.

We believe that the current study provides a general framework for how to counter these effects by enhancing weather display symbols for efficient use in a multitasking cockpit environment. During our stimulus creation, we used a step-wise procedure for salience-based analyses of the change-detection images (similar to the method used by Becker et al., 2016). First, we subjected our images to a frequency-tuned salience analysis using the algorithm by Achanta, Hemami, Estrada, and Süsstrunk (2009). The outputs of the algorithm are high-resolution salience maps on which bright objects have the highest salience, and dark objects have the lowest salience. Second, we analyzed the salience maps to determine each weather symbol's salience. Third, we enhanced symbols by adjusting the symbol color, thereby increasing the symbol salience. If display developers would use this framework, it would support GA pilots and possibly increase the efficiency and safety of operations by providing information that can be used to avoid adverse weather conditions during flight.

6 KEY FINDINGS

This study specifically evaluated the effect of symbol enhancement on pilot visual performance to reduce or resolve the change blindness gap. This study evaluated symbol color enhancements, symbol shapes, line thickness, and age-related effects.

In this study, we found following key results.

- We found an overall effect of participant age on discrimination accuracy and an increase in response time as age increases.
- We found a main effect of salience for detection of the line, lightning, and METAR symbols.
- We found a main effect of line color discriminability, with blue lines having the highest discriminability, followed by the red and then black lines.
- In general, enhancing the display symbols increases the discriminability accuracy and reduces the response time.

- Although this part-task study shows that enhancing symbol salience may help improve change-detection capabilities, it does not evaluate whether these enhancements would be adequate to draw attention to displays in a dynamic cockpit environment.

Companion research conducted by Ahlstrom et al. (2017) evaluated the effect of symbol salience in a dynamic, real-time simulation environment. The real-time simulation was needed to determine if salience alone will adequately enhance detection capabilities.

7 RECOMMENDATIONS

This part-task study specifically evaluated the effect of symbol enhancement on visual performance for GA pilots. We intentionally limited our pool of participants to GA with no commercial experience and limited instrument flight time. It is possible that commercial pilots who have more experience with cockpit weather technology may be more adept at detecting change. Evaluating differences in commercial versus GA pilots may help tailor display features for different user groups (i.e., GA, aging GA, or commercial) or identify training needs.

1. Conduct follow-up studies to assess the effect of symbol enhancement of weather displays on commercial pilots. This study evaluated GA pilots with a wide range of ages. Are there differences in the effects of symbol enhancement on pilots with a range of experience?
2. Conduct follow-up study on the placement of time-stamp information. This study found no credible difference in the detection of Enhanced versus Control time-stamp information. Is there a potential for detection improvement by changing the time-stamp location?

REFERENCES

Achanta, R., Hemami, S., Estrada, F., & Süsstrunk, S. (2009). *Frequency-tuned Salient Region Detection*. Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2009), 1597 - 1604. doi:10.1109/CVPR.2009.5206596

Achanta, R., Hemami, S., Estrada, F., & Süsstrunk, S. (n.d.). IVRG – Images and Visual Representation Group. *Frequency-tuned Salient Region Detection*. Retrieved from https://ivrlwww.epfl.ch/supplementary_material/RK_CVPR09/

Ahlstrom et al. (2019). *General Aviation Pilot Situation Assessment and Decision-Making During Flights in Deteriorating Visibility Conditions.* (DOT/FAA/TC-19/32). Atlantic City International Airport, NJ: FAA William Hughes Technical Center.

Ahlstrom, U. (2015). Experimental Evaluation of the AIRWOLF Weather Advisory Tool for En Route Air Traffic Controllers. *Aviation Psychology and Applied Human Factors, 5*(1), 18-35. https://doi.org/10.1027/2192-0923/a000070

Ahlstrom, U., & Dworsky, M. (2012). *Effects of Weather Presentation Symbology on General Aviation Pilot Behavior, Workload, and Visual Scanning* (DOT/FAA/TC-12/55). Atlantic City International Airport, NJ: FAA William Hughes Technical Center.

Ahlstrom, U., & Jaggard, E. (2010). Automatic Identification of Risky Weather Objects in Line of Flight (AIRWOLF). *Transportation Research Part C: Emerging Technologies, 18*(2), 187-192. https://doi.org/10.1016/S0968-090X(10)00003-3

Ahlstrom, U., & Suss, J. (2014). *Now You See Me, Now You Don't: Change Blindness in Pilot Perception of Weather Symbology* (DOT/FAA/TC-14/16). Atlantic City International Airport, NJ: FAA William Hughes Technical Center.

Ahlstrom, U., Caddigan, E., Schulz, K., Ohneiser, O., Bastholm, R., & Dworsky, M. (2015a). *Initial Assessment of Portable Weather Presentations for General Aviation Pilots* (DOT/FAA/TC-15/42). Atlantic City International Airport, NJ: FAA William Hughes Technical Center.

Ahlstrom, U., Caddigan, E., Schulz, K., Ohneiser, O., Bastholm, R., & Dworsky, M. (2015b). *The Effect of Weather State-change Notifications on General Aviation Pilots' Behavior, Cognitive Engagement, and Weather Situation Awareness* (DOT/FAA/TC-15/64). Atlantic City International Airport, NJ: FAA William Hughes Technical Center.

Ball, F., & Busch, N. A. (2015). Change detection on a hunch: Pre-attentive vision allows 'sensing' of unique feather changes. *Attention, Perception, & Psychophysics, 77*, 2570–2588. https://doi.org/10.3758/s13414-015-0963-9

Becker, M. W., Sundar, R. P., Bello, N. M., Alzahabi, R., Weatherspoon, L., & Bix, L. (2016). Assessing attentional prioritization of front-of-pack nutrition labels using change detection. *Journal of Applied Ergonomics, 54*, 90-99. doi:10.1016/j.apergo.2015.11.014

Costello, M. C., Madden, D. J., Mitroff, S. R., & Whiting, W. L. (2010). Age-Related Decline of Visual Processing Components in Change Detection. *Psychology and Aging, 25*(2), 356–368. doi:10.1037/a0017625

Droll, J. A., Gigone, K., & Hayhoe, M. M. (2007). Learning where to direct gaze during change detection. *Journal of Vision, 7*(14):6, 1–12. doi: 10.1167/7.14.6

Durlach, P. J. (2004). Army digital systems and vulnerability to change blindness. In H. Kwon, N. M. Nasrabadi, W. Lee, P. D. Gader, & J. N. Wilson (Eds.), *Proceedings of the Twenty-Fourth Army Science Conference* (Accession No. ADM001736). Redstone Arsenal, AL: Army Missile Research, Development and Engineering Lab.

FAA (2016). U.S. Civil Airmen Statistics. https://www.faa.gov/data_research/aviation_data_statistics/civil_airmen_statistics/

Federal Aviation Administration. (2010). *Weather technology in the cockpit program capabilities report* (DTFAWA-09-C-00088). Norman, OK: Atmospheric Technology Services Company, LLC.

Hardy, D. J., & Parasuraman, R. (2007). Cognition and flight performance in older pilots. *Journal of Experimental Psychology: Applied, 3*(4), 313-348. doi: 10.1080/13825585.2011.586018

Heathcote, A., Popiel, S. J., & Mewhort, D. J. K. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin, 109*(2), 340-347. http://dx.doi.org/10.1037/0033-2909.109.2.340

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*, 1489–1506. doi: 10.1016/S0042-6989(99)00163-7

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*(11), 1254–1259. doi: 10.1109/34.730558

Jänicke, H., & Chen, M. (2010). A saliency-based quality metric for visualization. In EuroVis'10 *Proceedings of the 12th Eurographics / IEEE - VGTC conference on Visualization*, 1183-1192. doi: 10.1111/j.1467-8659.2009.01667.x

Keshvari, S., van den Berg, R., & Ma, W. J. (2013). No evidence for an item limit in change detection. *PLOS Computational Biology, 9*(2), e1002927. doi:10.1371/journal.pcbi.1002927.

Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: a Tutorial with R, JAGS, and Stan, 2nd edition*. Academic Press/Elsevier, ISBN 9780124058880.

Latorella, K. A., & Chamberlain, J. P. (2002). *Graphical weather information system evaluation: Usability, perceived utility, and preferences from General Aviation pilots* (NASA-2002-01-1521). Hampton, VA: NASA. doi:10.4271/2002-01-1521.

Lee, M. D. (2008). BayesSDT: Software for Bayesian inference with signal detection theory. *Behavior Research Methods, 40*(2), 450-456. doi: https://doi.org/10.3758/BRM.40.2.450

Massidda, D. (2013). Package Retimes: Reaction Time Analysis. Retrieved from (https://cran.r-project.org/web/packages/retimes/retimes.pdf).

Matzke, D., & Wagenmakers, E-J. (2009). Psychological interpretation of the ex-Gaussian and shifted Wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Review, 16*(5), 798-817. doi:10.3758/PBR.16.5.798

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research, 42*, 107–123. https://doi.org/10.1016/S0042-6989(01)00250-4

Peich, M.-C., Husain, M., & Bays, P. M. (2013). Age-related decline of precision and binding in visual working memory.*Psychology and Aging, 28*(3), 729-743. http://dx.doi.org/10.1037/a0033236

Plummer, M., 2003. JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. In: *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*. Vienna, Austria. Retrieved from http://www.r-project.org/conferences/DSC-2003/Proceedings/Plummer.pdf.

Plummer, M., 2011. RJAGS: Bayesian Graphical Models Using MCMC. *R Package Version 3-5* [Computer software]. Retrieved from. http://CRAN.R-project.org/ package¼rjags.

Ratcliff, R., Thapar, A., & McKoon, G. (2001). The effects of aging on reaction time in a signal detection task.*Psychology and Aging, 16*(2), 323-341. http://dx.doi.org/10.1037/0882-7974.16.2.323

R Development Core Team, 2011. R: a Language and Environment for Statistical Computing [Computer software manual]. *R Foundation for Statistical Computing*, Vienna. Retrieved from. http://www.R-project.org.

Rensink, R. A. (2002). Change detection. *Annual Review of Psychology, 53*, 245–277. https://doi.org/10.1146/annurev.psych.53.100901.135125

Sekuler, R., & Sekuler, A. B. (2000). Age-Related Changes, Optical Factors, and Neural Processes. In Alan E. Kazdin (Editor in Chief) *Encyclopedia of Psychology*. American Psychological Association: Oxford University Press, 8, 180-183.

Simons, D. J., & Ambinder, M. S. (2005). Change blindness: Theory and consequences. *Current directions in Psychological Science, 14*(1), 44-48. doi: 10.1111/j.0963-7214.2005.00332.x

Spieler, D. H. (2001). Modelling age-related changes in information processing. *European Journal of Cognitive Psychology, 13*(1/2), 217-234. doi:10.1080/09541440042000287

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137-149. doi: https://doi.org/10.3758/BF03207704

Turatto, M., Angrilli, A., Mazza, v., Umiltà, C., & Driver, J. (2002). Looking without seeing the background change: electrophysiological correlates of change detection versus change blindness. *Cognition, 84*, B1–B10. https://doi.org/10.1016/S0010-0277(02)00016-1

Verma, M., & McOwan, P. W. (2010). A semi-automated approach to balancing of bottom-up salience for predicting change detection performance. *Journal of Vision, 10*(6):3, 1–17, http://www.journalofvision.org/co. doi:10.1167/10.6.3

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record, 58*, 475-482. https://doi.org/10.1007/BF03395630

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision, 4*, 1120-1135. doi:10.1167/4.12.11