NAMRU-D-21-034 Jan-2021

NAVAL MEDICAL RESEARCH UNIT DAYTON

Effects of Cockpit Workload and Motion on Incidence of Spatial Disorientation in Simulated Flight

by Henry P. Williams[1], Dain S. Horning[1], Cortland Etgen[1,2], and Charles R. Powell[1,3]

[1] Naval Medical Research Unit Dayton
[2] Oak Ridge Institute for Science and Education
[3] Parsons Corporation

STATEMENTS

Disclaimer

The views expressed in this technical report are those of the author and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, nor the U.S. Government.

Source of Support

This work was supported by Joint Program Committee-5/Military Operational Medicine Research Program, Aviation Mishap Prevention Working Group, and funded by work unit number H1705.

Human Research/Institutional Review Board (IRB) statement

The study protocol was approved by the Naval Medical Research Unit Dayton Institutional Review Board in compliance with all applicable federal regulations governing the protection of human subjects.

Copyright

Effects of Cockpit Workload and Motion on Incidence of Spatial Disorientation in Simulated Flight

Henry P. Williams[1], Dain S. Horning[1], Cortland Etgen[1,2], and Charles R. Powell[1,3]

[1] Naval Medical Research Unit Dayton

[2] Oak Ridge Institute for Science and Education

[3] Parsons Corporation

This work was supported/funded by work unit number H1705.

The study protocol was approved by the Naval Medical Research Unit Dayton Institutional Review Board in compliance with all applicable Federal regulations governing the protection of human subjects.

Acknowledgements

This work was supported by the Aviation Mishap Prevention Working Group, Joint Program Committee-5, Military Operational Medicine Research Program, and the authors express their sincere appreciation for that support.

Ms. Cassidy Shaver is commended for her superb organizational skills and attention to detail in all aspects of running this study. We recognize CAPT Richard Folga, Jason Whatmore, Adam Bowersock, Gary LaBance, John Kloczkowski, Luke Warner, Matt Lee, Kyle Kemp, and Tony Bailey for their unfailing safety consciousness and professional expertise in operating the very sophisticated and extremely powerful Disorientation Research Device used in this work. We thank Ms. Kiersten Compton for her diligent work in helping to prepare this manuscript. We also thank Dr. Angus Rupert for his expert recommendations on the design and execution of this study.

Abstract

Spatial disorientation (SD) refers to a pilot's misperception of the attitude, position, or motion of his/her aircraft, and it continues to be a deadly threat to military, commercial, and general aviation safety. While increased cockpit workload has been cited as a potential contributing factor to SD, few studies have examined the effect of different types of workload. Fixed- and motion-base simulators have been compared on their effectiveness in flight training in general, but not specifically on their effectiveness in inducing SD. In the current study, 12 pilots flew simulated flights in the Naval Medical Research Unit – Dayton's Disorientation Research Device (DRD). Three different workload conditions were presented. The baseline condition imposed no additional workload, while the other conditions added either a verbal Working Memory Task or a spatial Variable-Following-Distance Task. Pilots flew half of their flights with DRD motion disabled, and half with it enabled. The Variable-Following-Distance Task condition resulted in a statistically significant twofold increase in the number of control reversal errors, and a significant increase in altitude error. Adding motion did significantly increase realism ratings but otherwise had little effect. The results are discussed from multiple resource theory and spatial task interference perspectives.

Introduction

Aviation spatial disorientation (SD) can be described as a pilot's inaccurate perception of the attitude, position, or motion of his/her aircraft relative to the Earth's surface or other points of reference, including other aircraft (Benson, 1989). SD typically occurs under some form of degraded visual environment (DVE) such as flight in instrument conditions or at night (Gibb et al., 2010). If not recognized and resolved quickly, this misperception can lead to controlled flight into the ground, midair collision, or inappropriate control inputs resulting in aircraft stall and departure from controlled flight (Patterson et al., 1997). The prevalence of this problem has been well documented by mishap reports and surveys indicating that nearly all pilots experience some form of SD during their careers (Braithwaite, Durnford, Crowley, et al., 1998; Singh & Navathe, 1994; Tormes & Guedry, 1975). Accident statistics help quantify the size and severity of this deadly threat to aviation safety. Poisson and Miller (2014) reviewed mishap data from the United States Air Force Safety Center's Air Force Safety Automated System. This extensive review covered the 21 year period from fiscal years 1993 through 2013, and it focused upon Class A mishaps, which by definition result in a loss of life, permanent disability, and/or more than $2 million in property damage. The authors found a total of 601 Class A mishaps, and 72 (12%) of these included SD as a causal factor. Tragically there were 101 lives lost in those 72 mishaps. When fatality rates of non-SD and SD-related Class A's were compared, it was found that 16.1% of non-SD mishaps involved a fatality, but distressingly, 61.1% of the SD-related mishaps were fatal.

SD has also been documented as a serious problem for both Army and Naval aviators. Regarding the U.S. Army, Braithwaite et al. (1997) reported that between 1987 and 1995 the Army experienced 970 Class A through C mishaps, of which 291 (30%) involved SD. A U.S. Army Combat Readiness Center/Safety Center review of fiscal years 2002 through 2013 documents that DVE was responsible for 25% of Class A/B mishaps during that time, with DVE mishaps accounting for 46% of total fatalities (Edens & Higginbotham, 2014). Similarly, the U.S. Naval Safety Center indicates that SD was the #1 aeromedical causal factor of Class A mishaps occurring from 1990 – 2008 (Gibb et al., 2011).

In the general aviation (GA) community, SD has been cited as a causal factor in 11% to 16% of all fatal accidents (Collins & Dollar, 1996; Kirkham et al., 1978). In GA accidents that were attributed to SD, the chances of survival were particularly grim. Kirkham et al. (1978) reported that 90% of these were fatal accidents, while Mortimer (1995) found that number to be 92%.

Finally, a report from the Boeing company listed loss of control inflight (LOC-I) and controlled flight into terrain (CFIT) as the top two causes of fatalities in the worldwide commercial jet fleet (of any manufacture) in the period covered by 2009 through 2018 (Boeing, 2019). While SD was not explicitly listed as the cause of these mishaps, LOC-I and CFIT are strongly associated and frequently caused by SD (Véronneau & Evans, 2004; Lawson et al., 2003). The total number of lives lost in the 13 LOC-I and 10 CFIT accidents was 1751.

Clearly SD is a threat to safety in military, commercial, and general aviation. Traditional approaches toward combatting SD have focused on training, ranging from simple demonstrations in a Bárány chair to sophisticated motion-base flight simulators, to training in actual aircraft in simulated instrument conditions. Other methods of reducing SD incidence have concentrated on novel cockpit instrument design. While these approaches certainly have merit, the fact that the SD mishap rate is not improving (Gibb et al., 2011; Newman & Rupert, 2020) suggests that other strategies are needed. One such approach is to gain a better understanding of the root causes of SD by conducting flight simulator research in the lab. Using the knowledge gained from this research, SD models can be developed to specify conditions in which SD is likely to occur, and training can be developed to help pilots avoid those conditions and prevent SD.

Among the variables that may increase the likelihood of pilot SD, high mental workload resulting in task saturation has been identified as a potential causal factor (Stott, 2013). A number of studies have looked at relationships between increased workload and various degrees of spatial orientation. Webb et al. (2012) evaluated how performance on two different secondary cockpit tasks was impacted by SD. One of the secondary tasks was a working memory task in which participants had to remember a string of up to nine single digit numbers. The other task also presented a string of single digit numbers but required subjects to add the two most recent numbers. The primary task was flying a helicopter simulator in various weather conditions and scenarios, some of which were designed to induce SD. The results showed that SD significantly disrupted secondary task performance, but the study was not specifically designed to assess the converse: how does workload imposed by a secondary task affect the probability of SD?

Braithwaite, Durnford, Groh, et al. (1998) came closer to addressing this question. They tested pilots on instrument flight precision and unusual attitude recovery, comparing performance with and without a secondary task, which was a forced choice (High or Low) two-tone audio discrimination task. They also compared standard cockpit instrumentation to a novel display. Generally the results showed that the secondary task negatively impacted IFR flight precision with the standard instrument display, while unusual attitude recovery, and presumably spatial orientation, were not significantly affected by that particular additional task.

Raj et al. (2000) also studied a secondary task's impact on spatial awareness, testing pilots' abilities to maintain a stable hover, with and without the additional task and with and without a vibrotactile stimulus designed to provide veridical drift cues. The secondary task required pilots to add a series of three-digit number pairs. While the impact of the math task on hover performance was not *statistically* significant, hover quality was degraded as measured by the number of drift excursions of more than ten feet from the designated hover point, when the vibrotactile cueing was not provided (A. Raj, personal communication, February 20, 2018). This ten foot limit is a standard operational pass/no pass criterion for hover proficiency.

While the three aforementioned studies examined various aspects of spatial awareness in the presence of certain secondary tasks, none specifically and directly compared how different types of secondary tasks and their associated workload affected the probability of SD. To gain insight into this relationship, Williams et al. (2018) completed a non-motion-base flight simulator study that utilized 24 pilots to evaluate SD under four different workload conditions. For that experiment subjects followed a lead aircraft in trail formation through a series of turns, climbs and descents, first above, and then into clouds. While in a turn in the clouds, the lead disappeared and the subjects were to level their wings and start a gentle climb. There were four workload conditions. The baseline condition presented no additional cockpit tasks. However the other three conditions included either a verbal working memory task, a spatial mental rotation task, or a spatial variable-following-distance task. The latter task required subjects to vary the distance at which they followed the lead aircraft. The mental rotation and variable-following-distance tasks yielded significant increases in unusual attitudes and the verbal working memory task resulted in a statistically significant threefold increase in the number of control reversal errors (CREs). A CRE occurs when a pilot makes an incorrect stick input in a direction opposite to that required for a given flight situation, and a CRE is a good indicator of SD.

CREs can occur when a pilot misinterprets the spatial information presented on the attitude display during transition from visual to instrument flight. A CRE can also occur when pilots are distracted by other cockpit tasks during pitch or roll maneuvers. Published mishap reports indicate CRE can happen to pilots of all experience levels and it has been identified as a causal factor in fatal aircraft mishaps (Bramble, 2008). Since the relationships between mental workload, SD and CRE are not completely understood, the Williams et al. (2018) experiment helped fill that knowledge gap. While that study did incorporate various types of workload drivers and a diverse set of SD measures, there was no motion component within the flight simulation. Unlike actual flight maneuvers that stimulate a pilot's vestibular, tactile, and somatosensory systems with rotational and linear accelerations, the experiment was limited to studying the visual aspects of SD. Since the accelerations experienced in actual flight can be disorienting, the study was missing an important element for understanding SD, that being motion.

Studies of motion vs. non-motion simulators have mostly focused on training effectiveness, not on spatial disorientation per se, and the results are mixed. Kallus et al. (2011) found training performance advantages for simulator motion in spin recovery and properly reacting to the "pitch-up illusion" (a form of the somatogravic illusion), but not for unusual attitude recovery. This study was a simulator-only study where there was no evaluation in an actual aircraft. Meta-analyses (Vaden & Hall, 2005; de Winter et al., 2012) on similar studies generally show some advantage to training with motion but results depend highly on the type of flight task being trained. Koonce (1974) conducted a transfer-of-training study, evaluating actual flight performance, and found that motion improved performance within the simulator, but this benefit did not transfer to the actual aircraft. Jentsch et al. (2011) review a number of studies showing similar results.

For SD related training, there is a fair amount of *subjective/survey* data indicating that motion-base simulators are valuable for demonstrating various forms of SD (see Braithwaite et al., 2004, for a review). However, in terms of effectiveness in inducing SD, *objective* data from studies comparing motion vs. non-motion simulators is lacking. Examining the motion variable is important since adding motion capability can greatly increase simulator complexity and costs for acquisition, operation, and maintenance. Life cycle costs can increase by over $1 million per simulator (Bartel & Foster, 2004), and rental costs can increase by 50% (Vaden & Hall, 2005).

The current study was designed to compare motion vs. non-motion flight simulation regarding ability to induce SD, and also to evaluate three different workload conditions regarding their effects on the probability of SD. The intent was to expand and improve upon the study conducted by Williams et al. (2018) to gain a better understanding of SD.

Method

Participants

One female and eleven male pilots participated in this study with ages ranging from 30 to 62 years, and a mean of 45.5 years. Ten of the twelve held at least a civilian private pilot rating, five were current or former military pilots, and 2 were civilian student pilots who had soloed. Seven held instrument ratings. Flight hours ranged from 40 to 5000, with a mean of 1580.6. Subjects were recruited via word of mouth, posted flyers, and e-mail contact, and they were not compensated.

Apparatus

To create a motion-base flight simulator environment the research was conducted in the Disorientation Research Device (DRD) at the Naval Medical Research Unit – Dayton (NAMRU-D) (Figure 1). The DRD is a six degree-of-freedom device that can replicate angular and linear flight accelerations with up to 3 Gs of sustained force, but a 2 G limit was set for this experiment. The out-the-window (OTW) and instrument panel graphics were generated with X-Plane version 11.41 flight simulation software, and the in-house-developed flight model simulated the T-6 Texan II aircraft. Engine power was controlled with the left hand via a Thrustmaster Warthog throttle and pitch and roll were controlled with the right hand via a Flightlink G-Stick III joystick. Yaw was controlled with adjustable rudder pedals fabricated in-house, and the cockpit was equipped with a seat that was adjustable in height.

Figure 1. Naval Medical Research Unit Dayton’s (NAMRU-Dayton) six degree-of-freedom Disorientation Research Device: the Kraken^{TM} cockpit gondola (Photo courtesy of NAMRU-Dayton).

The cockpit instrument panel was displayed on a 26 inch diagonal ELO 2639L monitor with a resolution of 1366 x 768 pixels. The OTW scene was displayed on a 65 inch diagonal Samsung 4K SUHD TV flat panel display which provided a 83º horizontal x 53° vertical field-of-view. Head and eye movements were tracked and recorded with a Smart Eye 4-camera 60Hz head and eye tracking system. An additional 3-camera video system was used to monitor participants at all times, and an audio system provided two-way voice communications.

The motion control algorithm used for generating the motion cues for the T-6 simulation was a classical washout algorithm extended to the DRD motion space. This approach (Figure 2) combined the use of low-pass (LP) and high-pass (HP) signal filters to generate linear and rotational motion cues while mitigating false cues by washing out any motion to the initial system state. LP filtering was used to create sustained acceleration motion cues by way of tilt-alignment, while HP filtering provided rapid-onset motion cues for linear accelerations and angular rates. The decision to use classical washout was made with particular consideration given to the required speed of computation for real-time control and the large dynamic envelope of the T-6 aircraft.

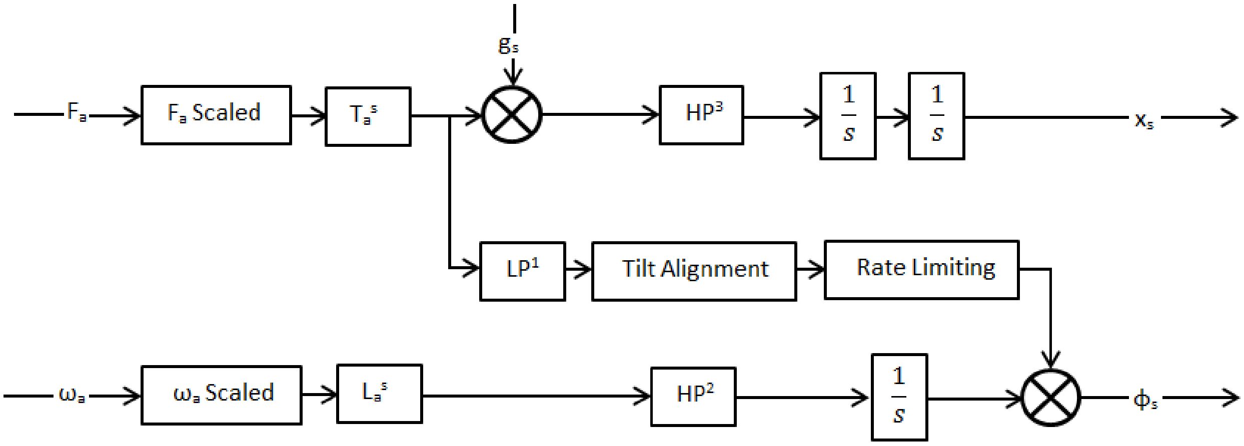

Figure 2. Classical washout motion control algorithm.

Flight physics data were sent from the T-6 flight simulation to the DRD motion control algorithm in real-time at a frequency of 125 Hz. These data included: aircraft-body reference frame linear non-gravitational accelerations (F_a), measured in G's; and aircraft-body reference frame angular rates (ω_a), measured in degrees per second (°/s). Components of F_a were limited to within ±2G using sigmoidal bounding functions while components of ω_a were limited to between ±60°/s. After scaling, both F_a and ω_a were then transformed via their respective rotation matrices (T_a^s and L_a^s) from the aircraft-body reference frame into the DRD reference frame.

For linear acceleration cueing, gravity relative to the DRD reference frame (g_s) was added to F_a to account for all acceleration components in the DRD-frame. This signal was then passed through a 3rd-order linear HP filter (HP^3). This isolated the portion of F_a required to generate linear acceleration cues while simultaneously applying washout characteristics. The filtered signal was then integrated twice to give the corresponding linear displacements (x_s) for cueing linear acceleration. The x- and y-direction components of F_a were passed through a 1st-order linear LP filter (LP^1) before being fed to the tilt-alignment procedure. Tilt-alignment was achieved by equating the low-passed signal with the DRD-capsule orientation matrix, linearized using small-angle approximation in the roll and pitch directions, multiplied by the unit vector in the z-direction. This gives the linearized system of equations:

$$\begin{cases} F_x = \theta_d - \phi_d * \sin(\psi) \\ F_y = \phi_d * \cos(\psi) \end{cases};$$

where ψ is the current capsule yaw angle, and ϕ_d and θ_d are the desired capsule roll and pitch angles, respectively, for the tilt-alignment solutions. Once calculated, ϕ_d and θ_d are then limited to ±10° due to perceptibility of tilt for larger angles. With these desired angles, we calculated the new roll and pitch angular rates by subtracting the current roll and pitch angles (ϕ_n and θ_n) from the desired angles and then dividing by the time step ($\Delta t = 0.008$ s). Next, in an attempt to minimize false rotational cues by keeping the rates below perceivable angular rate thresholds, the angular rate values were limited between ±3°/s using a min-max function. Finally the new angles for tilt were evaluated using the following equations:

$$\phi_{n+1} = \phi_n + \max\left(-3, \min\left(3, \frac{(\phi_d - \phi_n)}{\Delta t}\right)\right) * \Delta \mathrm{t}$$
$$\theta_{n+1} = \theta_n + \max\left(-3, \min\left(3, \frac{(\theta_d - \theta_n)}{\Delta t}\right)\right) * \Delta \mathrm{t}\,.$$

The final phase of this segment was adding ϕ_{n+1} and θ_{n+1} to the angle values corresponding with the aircraft angular rates. The final set of calculations took ω_a (already scaled and rotated using L_a^s) and passed it through a 2nd-order linear HP filter (HP^2). This isolated the portion of ω_a required to generate angular velocity cues while simultaneously applying washout characteristics. The filtered signal was then integrated once, giving the corresponding DRD capsule angular positions. These angular positions were then added to ϕ_{n+1} and θ_{n+1}, giving the final capsule angles (φ_s).

Tuning of the HP and LP filter coefficients was accomplished using extensive feedback from test pilots familiar with the real aircraft. Test pilot feedback was consistently positive for motion cues generated by the relatively docile aircraft maneuvering required for this study. However, it was noted that unusual attitudes as well as aerobatic maneuvering may present false cues due to motion limitations and washout, which appear inconsistent with what one would experience in reality.

Procedure

The study protocol was approved by the NAMRU-D Institutional Review Board. The informed consent document (ICD) was reviewed with participants when they arrived at the lab. The experimenter answered any questions, signed the ICD after the participant signed, and provided them with a copy.

The participant’s flight task was to follow a pre-recorded lead aircraft through a series of turns, climbs, and descents. Each trial began with the aircraft in straight and level flight, at 9,500 feet above mean sea level (MSL). There was a level undercast (cloud deck) at 9000 feet MSL, and it provided a level and well-defined visual horizon as shown in Figure 3. Visual Meteorological Conditions (VMC) existed above the clouds and visibility was excellent and set to 25 statute miles.

Two lead aircraft flights (Flight A and Flight B) were recorded, each a mirror image of the other. In both flights the lead aircraft executed three right (R) and three left (L) turns, each with a 90° heading change and at a 45° angle of bank (AOB). In Flight A, the order of turns was R, L, L, R, L, R, and in Flight B turns were exactly opposite. For both flights, turn 1 was a level turn, turn 2 was climbing, and the remaining turns were descending. At approximately 1 min 30 s into the flight, turn 4 brought the formation into Instrument Meteorological Conditions (IMC) in the clouds, where the formation remained for the rest of the flight. While in the clouds, the lead aircraft remained visible but the horizon was completely obscured. Participants were told that while the flight was in the clouds the lead would disappear at some point, whereupon they should scan their flight instruments, level their wings and start a gentle climb. At approximately 2 min 30 s into the flight the lead disappeared while established in the sixth turn, in the clouds and with no visible horizon. The experimenter ended the trial shortly after the subject stabilized the aircraft in a wings-level gentle climb.

Figure 3. Screen shot from the formation flight scenario used in this experiment, showing a typical view of the lead aircraft while in a right turn (Photo courtesy of NAMRU-Dayton).

Experimental Design

This experiment used a 2 x 3 repeated measures (within subjects) design (Table 1). There were two motion conditions (“No-Motion”, “Motion”) and the three workload conditions (described below). With three workload conditions there were 6 possible presentation orders, and these were counterbalanced across the 12 participants in two blocks, with one block of six participants experiencing the No-Motion condition first, the other experiencing Motion first.

Table 1. Experimental design. NAW = No Added Workload; WMT = Working Memory Task; VFDT = Variable Following Distance Task

		Workload Condition		
		NAW	WMT	VFDT
Motion Condition	No Motion	6 flights	6 flights	6 flights
	Motion	6 flights	6 flights	6 flights

Workload Conditions

Three workload conditions were created. Within each motion condition, each subject flew six flights in each workload condition, three Flight A's and three Flight B's, in a pre-determined randomly alternating order with the constraint that no more than two Flight A's or Flight B's occurred in a row. All pilots were presented with that same predetermined order.

1) No Added Workload (NAW) - For this condition participants simply followed the lead aircraft with no added secondary task. Participants were told to fly in trail formation at the same altitude as the lead. They were told that the spacing between their aircraft and the lead aircraft was not critical, but that they should stay close enough to easily see the lead aircraft's attitude. They were instructed that when the lead disappeared they should look at their attitude indicator, level their wings, and start a gentle climb. The trial ended shortly after the wings-level climb was established. Since there were no additional tasks, this was considered to be the control condition.

2) Working Memory Task (WMT) - This condition was the same as the NAW task above, but with a verbal working memory task (WMT) added. At the beginning of each flight, a set of 7 letters was presented on the OTW display for 10 seconds, and subjects were told to memorize this target letter set. In order to limit phonetic rehearsal the target letters were arranged to avoid forming pronounceable words (e.g., FVPGCQY). After the target set disappeared, a probe letter was shown on the OTW display and presented audibly every four seconds, and the participant pressed one of two buttons on the joystick to indicate whether or not the probe was a target letter. Of the 32 probe letters presented on each WMT flight, on average 10.2 (30.8%) were targets (the number of targets ranged from 9 – 11 per flight). The targets were randomly dispersed among the probes in a pre-recorded order. Twelve different memory set/probe set combinations were recorded, one for each of the WMT flights shown in Table 1. Subjects were told that if they forgot the memory set, they could ask for a "refresh", whereupon the letters would be redisplayed for another 10 s. Response time and percent correct were automatically recorded.

3) Variable Following-Distance Task (VFDT) - This condition was the same as the NAW condition, but with an added task requiring subjects to follow the lead aircraft at various distances. A scale appeared on the simulator's OTW display indicating the requested or "command" following distance (in feet) as well as the actual following distance (see Figure 4). The initial command distance was always 1000 feet with the aircraft set properly at that distance. A new distance was presented in a predetermined random order every 40 s, with a total of 4 distances per flight. Command distances ranged between 400 and 1200 feet at 200 feet intervals. Six different following distance orders were created and repeated for the Motion and No-Motion flights. Following distance error (deviation from command distance) was automatically recorded throughout the flight.

Participants were told to treat the WMT and VFDT as "Critical Mission Tasks". They were instructed to do as well as they could both in flying the simulated aircraft, and on the critical mission tasks. While the added tasks were designed to impose different types of workload, pilot testing was used to adjust the amount of workload in an effort to make the tasks roughly equal in that dimension.

Before the simulator flights began, subjects were given a DRD safety and egress briefing, and a staff member helped strap them into the seat with a five-point safety harness, and adjust the seat and rudder pedals. The eye tracker was calibrated and the experimenter explained the controls, displays, and operating characteristics of the simulator. Subjects then flew a full practice flight in each of the three workload conditions with DRD motion disabled. Research staff members then exited the DRD capsule, secured its door, and exited the DRD device hall.

Figure 4. Screen shot of the Variable Following-Distance Task (VFDT), with 400 feet as the requested following distance, and 345 feet as the actual distance (Photo courtesy of NAMRU-Dayton).

Participants then flew a block of six data collection flights in each workload condition, completing a total of 18 flights, in one of the two motion conditions. At this halfway point they were then given a 10 minute break during which they exited the DRD capsule and were offered a refreshment. They then re-entered the DRD to again complete six flights in each workload condition, but in the other motion condition.

The main dependent variables were number of Control Reversal Errors (CREs) and the percentage of time participants reached or exceeded 65º AOB in the clouds. A CRE was defined as a lateral (roll) stick movement in the direction opposite to that required for a wings level recovery, where the errant stick input:

- had to be 5% or more of the maximum possible stick throw from the center/neutral position, and/or
- resulted in a 5º or more change in aircraft AOB in the wrong direction, steepening the bank angle rather than leveling the wings
- occurred while the pilot was looking at the attitude indicator

For the purposes of this study the CRE period of interest was initiated by the lead aircraft disappearance, when the subject was to look at the attitude indicator, determine their aircraft direction of bank, and recover to a wings level attitude. In these flight conditions, the only reasonable explanation for moving the stick in the wrong direction and committing a CRE is that the pilot has misperceived the direction of bank, either by misinterpreting the attitude indicator or by feeling they are banked in the opposite direction. In either case, the pilot is spatially disoriented.

The 65º AOB threshold was chosen since it is a high bank angle for anything other than acrobatic flight, and it is an extremely high bank angle for flight in the clouds. Also, the FAA requires pilots and passengers to wear parachutes on flights where the intent is to exceed 60° AOB, and the threshold chosen here exceeds the FAA's. Percentage of time, rather than cumulative time, at or beyond 65º AOB was chosen since time in the clouds and trial length could vary slightly depending on pilot performance.

While these criteria are somewhat arbitrary, if exceeded they are good indicators that the pilot is uncertain of the aircraft's attitude. In the clouds it is very unlikely that a pilot would intentionally move the stick in the wrong direction or exceed 65º AOB, so these actions can be taken as signs of SD.

Another dependent variable of interest was altitude error, or altitude difference from the lead aircraft, since subjects were told to fly co-altitude with the lead. While performance here may not be a direct measure of spatial orientation, this task may be one that gets neglected or shed as workload increases and pilots strive to maintain orientation.

WMT and VFDT performance, aircraft flight parameter data, and eye and head tracking information were also collected. The eye tracking information was used to confirm that, after the lead aircraft disappeared, subjects had transitioned to instruments before making any roll inputs to recover wings-level, and that they were not still looking outside for the lead and perhaps unknowingly moving the stick. Verifying that they were looking at their instruments when an errant stick movement occurred helps classify the error as a CRE and not an unintentional movement made while searching outside.

Participants were asked to provide subjective workload ratings for each workload condition. The ten-point Bedford Pilot Workload Rating Scale (Roscoe & Ellis, 1990) was used to obtain these ratings. After flying each workload condition, participants were also asked to use a ten-point scale to provide ratings of any experience of SD. The anchor points on this scale were:

1 = "I did not feel spatially disoriented at all" and, *10 = "I felt severely spatially disoriented"*.

The workload and SD ratings were collected after each block of six flights in each workload condition, and the pilots were asked to provide separate ratings for "above the clouds", and "in the clouds".

After completing their last flight, participants were asked if they knew which block of flights included motion, their choices being *"First", "Second", "Neither", or "Both"*. They were then asked to rate their confidence in their answer on a 10 point scale, with the anchor points:

1 = "Not confident at all" and, *10 = "Completely confident"*

They were then informed which block of flights included motion, and asked *"Compared to flight in an actual aircraft, how realistic did your simulator flights feel today?"*. They were asked to provide separate ratings for the No-Motion and Motion flights, again using a 10 point scale, with the anchor points:

1 = "Not realistic at all" and, *10 – "As realistic as actual flight"*.

Results

Subjective Ratings

Given the subjective nature of the data, non-parametric Aligned Rank Transformed Data (ARTD) 3-way ANOVAs were applied. Briefly, the ARTD procedure entails calculating marginal effects for each main effect and interaction to produce a new response variable for each effect, ranking each new response variable, and then running a full-factorial ANOVA on these ranks. For a detailed explanation see (Wobbrock et al., 2011). The three independent variables were Workload Condition (WL: NAW, WMT, VFDT), Motion Condition (MOT: No Motion, Motion), and Meteorological Condition (MET: VMC, IMC).

Ratings of workload are plotted in Figure 5, Panels a and b. There was no significant effect for MOT on WL ratings, *F*(2, 22) = .13, *p* = .73. The effect for MET was significant, with IMC receiving higher workload ratings than VMC; *F*(1, 11) = 23.64, *p* < .001. There was also a significant effect for WL condition *F*(2, 22) = 18.16, *p* < .0001. Post-hoc pairwise comparisons were conducted using stepwise Holm-Bonferroni corrections to limit familywise error to *p* < .05. These analyses showed that WMT WL ratings were greater than those for NAW (*t*(22) = 5.55, *p* < .0001), and the same pattern held for the VFDT vs. NAW ratings (*t*(22) = 4.81, *p* < .001). WMT and VFDT ratings did not differ significantly from each other. None of the interaction measures approached significance.

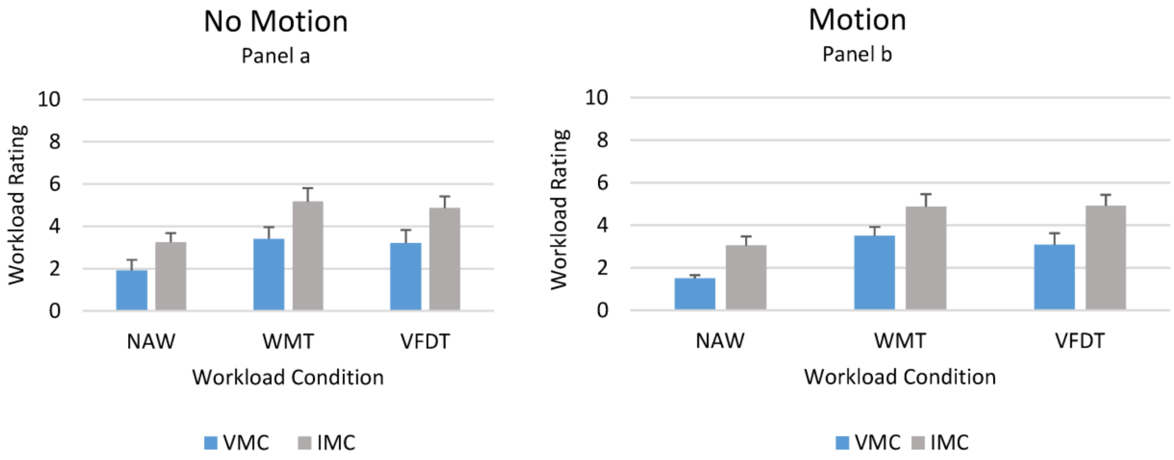

Figure 5. Workload ratings for the No Motion (Panel a) and Motion (Panel b) conditions.

Ratings of SD are plotted in Figures 6, Panels a and b. As with the WL ratings, there was no significant effect for MOT, *F*(1, 11) = 1.48, *p* = .25. The effect for MET was significant, with IMC receiving higher ratings than VMC; *F*(1, 11) = 40.90, *p* < .0001. WL condition did not significantly affect SD ratings (*F*(2, 22) = 1.00, *p* = .39, but the MET x WL interaction was significant; *F*(2,22) = 8.44, *p* < .01. The basis of that interaction appears to be in the relatively larger increase in SD ratings for the VFDT SD ratings across the VMC workload conditions, as compared to the step-like and even increases across the same IMC workload conditions (see Figure 7). Figure 7 collapses ratings across the motion conditions since there was no significant effect there, and also rescales the plot to help visualize the interaction.

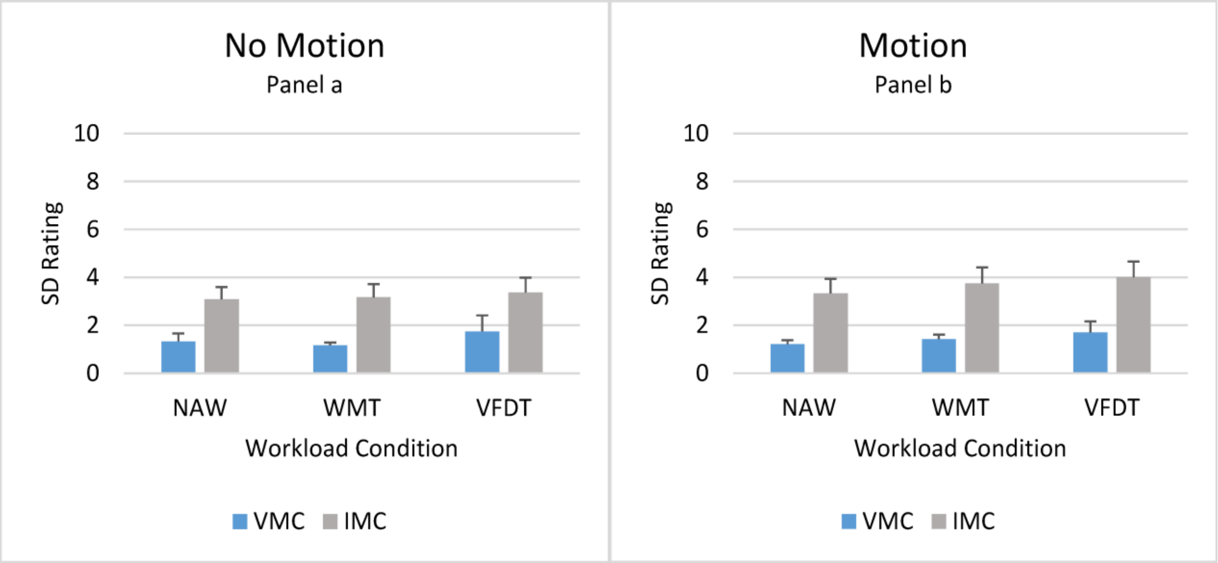

Figure 6. SD ratings for the No Motion (Panel a) and Motion (Panel b) conditions.

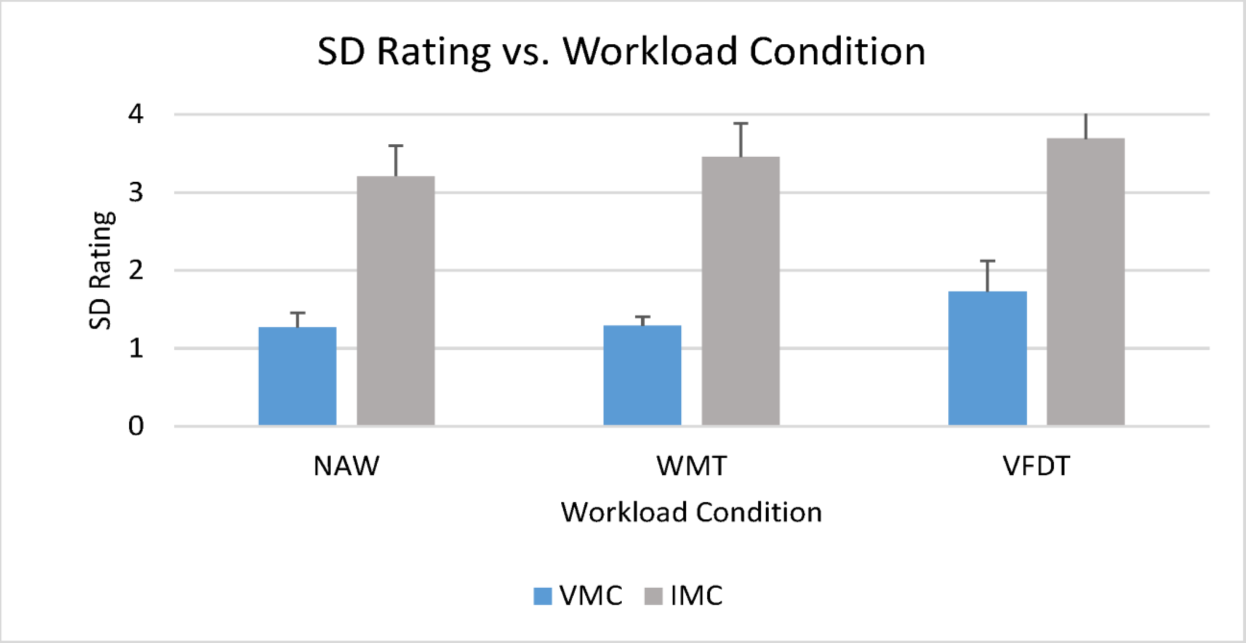

Figure 7. SD ratings collapsed across workload conditions. The significant interaction appears to be due to the relatively larger increase in VFDT SD ratings across the VMC workload conditions, as compared to those increases across the same IMC workload conditions.

Regarding subject's ability to identify which block of flights included motion, nine of the twelve were correct, while the remaining three felt that both blocks of flights (i.e., all of their flights) included motion. Each of those three subjects experienced the motion condition first. Confidence in the motion condition answers spanned the entire scale from "1", a single case, to "10", which was the mode at six cases. The mean confidence rating was 8.40 (*SD* = 2.70).

For realism ratings (1 to 10 scale), the mean rating for the No Motion condition was 4.04 (*SD* = 1.63), while the mean for the Motion condition was 7.63 (*SD* = 1.00). A Wilcoxon Signed-Ranks test indicated that realism ratings were significantly higher in the Motion condition than in the No Motion condition, $Z = 2.94$, $p < .01$ (two-tailed).

CREs

With 12 subjects each flying 36 flights, a total of 432 flights were flown, each with an opportunity for a CRE. The actual total number of CREs was 46, or 10.6% of all trials. Three of the 12 participants committed no CREs, while the remaining nine (75.0%) committed two or more. Figure 8 shows the breakdown of CREs across the workload conditions.

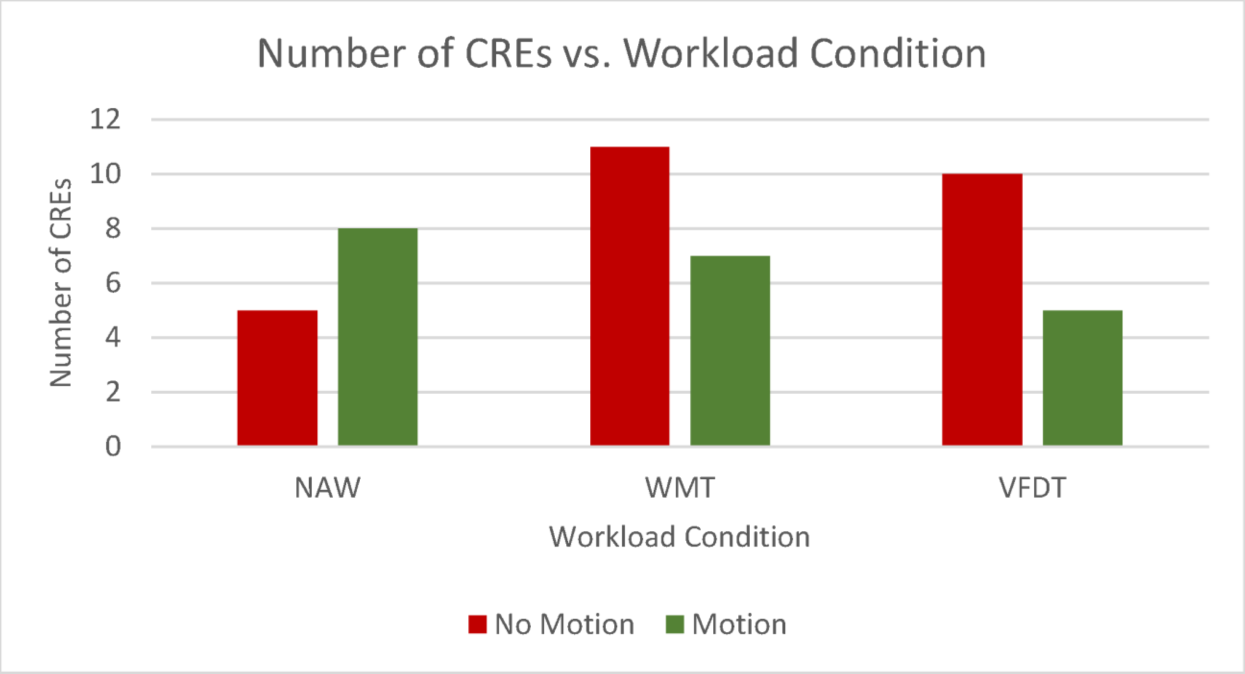

Figure 8. Number of CREs by workload and motion conditions.

There are two characteristics of CRE data that make them somewhat unique: CREs are rare events, and they are dichotomous (i.e., for each flight, a CRE either occurred, or it did not). These properties can make data poor candidates for traditional analyses such as ANOVA, so two less common but more appropriate approaches were considered: Poisson and Negative Binomial Regression. Initial inspection of the CRE data indicated they were somewhat over dispersed (the variances were greater than the means), and since the negative binomial model was formulated to handle this condition, that approach was chosen. The NAW condition was set as the baseline or reference workload condition for parameter estimates. Separate negative binomial regressions were run for the No-Motion and Motion conditions.

The analysis for the No-Motion condition found workload condition to be a significant predictor of CRE committal, $\chi^2(2) = 6.59$, $p = .04$. Parameter estimates revealed that subjects were significantly more likely to commit a CRE in the VFDT condition, $\chi^2(1) = 6.01$, $p = .01$. The exponentiated regression weight for VFDT showed that subjects were 2.0 times as likely to commit a CRE compared to the NAW condition.

Even though the WMT resulted in one more CRE than the VFDT (11 vs. 10, respectively), the number of CREs in the WMT was not statistically greater than that in the NAW condition, $\chi^2(1) = 1.75$, $p = .19$. This paradoxical result is due to the fact that three of the twelve subjects committed more CREs in NAW condition than in the WMT, biasing the numbers against a significant effect for the increased amount of WMT CREs.

The same type of negative binomial regression was conducted for the flights with motion. This analysis found that CRE rates were not significantly affected by workload condition, $\chi^2(2) = .81$, $p = .67$.

A negative binomial regression was also conducted comparing the number of CREs in the No-Motion vs. Motion conditions. This analysis revealed no significant difference, $\chi^2(1) = .67$, $p = .41$.

<u>65º AOB Exceedance</u>

The percentage of time that subjects spent at or beyond 65º AOB in the clouds was calculated for each motion-by-workload condition, and the results are plotted in Figure 9. Examination of the data for normality indicated that a non-parametric analysis was in order, so an ARTD 2 x 3 (Motion x Workload) repeated measures ANOVA was applied. The main effect for workload approached, but did not reach significance, $F(2, 22) = 2.50$, $p = .11$. There was no significant motion effect ($F(1,11) = .07$, $p = .79$), and no significant interaction ($F(2, 22) = 1.21$, $p = .32$).

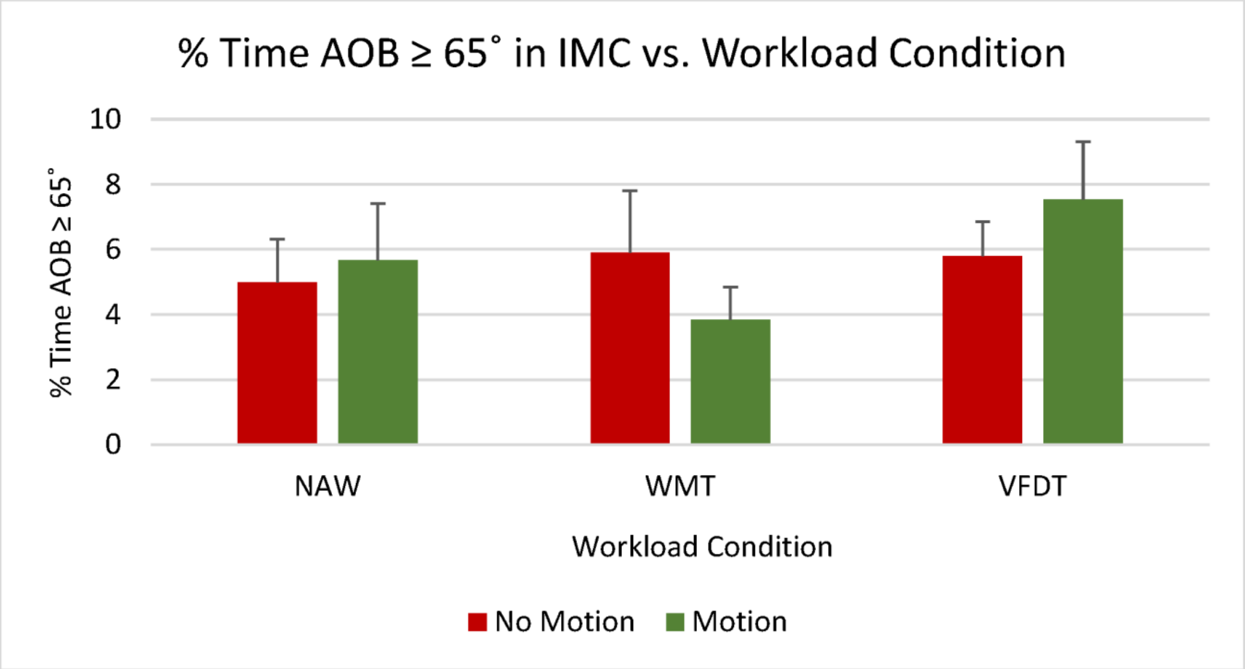

Figure 9. Average percent of time AOB was greater than or equal to 65°. NAW = No Added Workload, WMT = Working Memory Task, and VFDT = Variable Following Distance Task.

Percentage of time was used in the calculations above since time in the clouds and trial length could vary slightly depending on pilot performance. To provide an idea of the average number of seconds per flight spent at or beyond 65º AOB in the clouds, Table 2 converts percentages to seconds.

Table 2. Mean percentage of time and number of seconds per flight spent at or beyond 65° AOB in the clouds averaged across the motion conditions.

	NAW	WMT	VFDT
% Time	5.33	4.87	6.66
Seconds	2.53	2.31	3.16

Altitude Error

Altitude Error was defined as the absolute value of the altitude difference between the lead and subject's aircraft. The data were divided into the two visibility conditions, the two motion conditions, and the three workload conditions. The means are plotted in Figure 10, Panels a and b. Examination of the data for normality indicated a natural log transformation was appropriate for correction, and a 2 x 2 x 3 ANOVA was performed on the transformed data. The main effect for motion was not significant, while that for visibility condition was ($F(1, 11) = 28.48$, $p < .0001$), with lower altitude error during flight above the clouds. The main effect for workload condition $F(2, 22) = 6.92$, $p < .01$ and the Visibility by Workload condition interaction were significant ($F(2, 22) = 4.38$, $p = .03$). Figure 11 collapses ratings across the motion conditions since there was no significant effect there, and this should help visualize the interaction. None of the other interactions were significant.

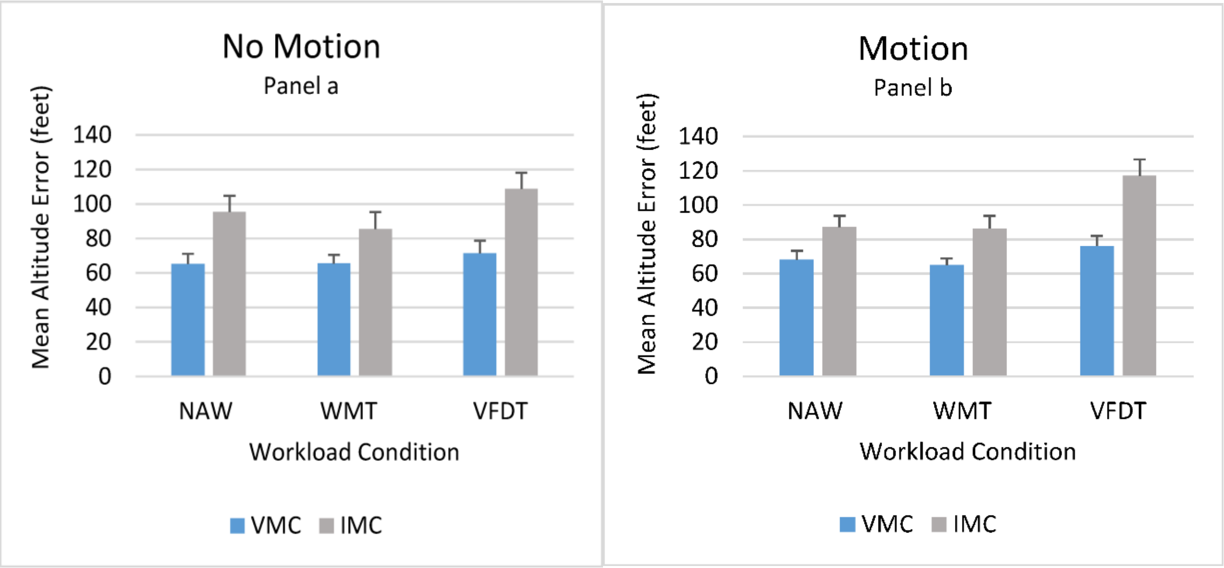

Figure 10. Mean altitude error across the workload conditions for the No Motion (Panel a) and Motion (Panel b) conditions.

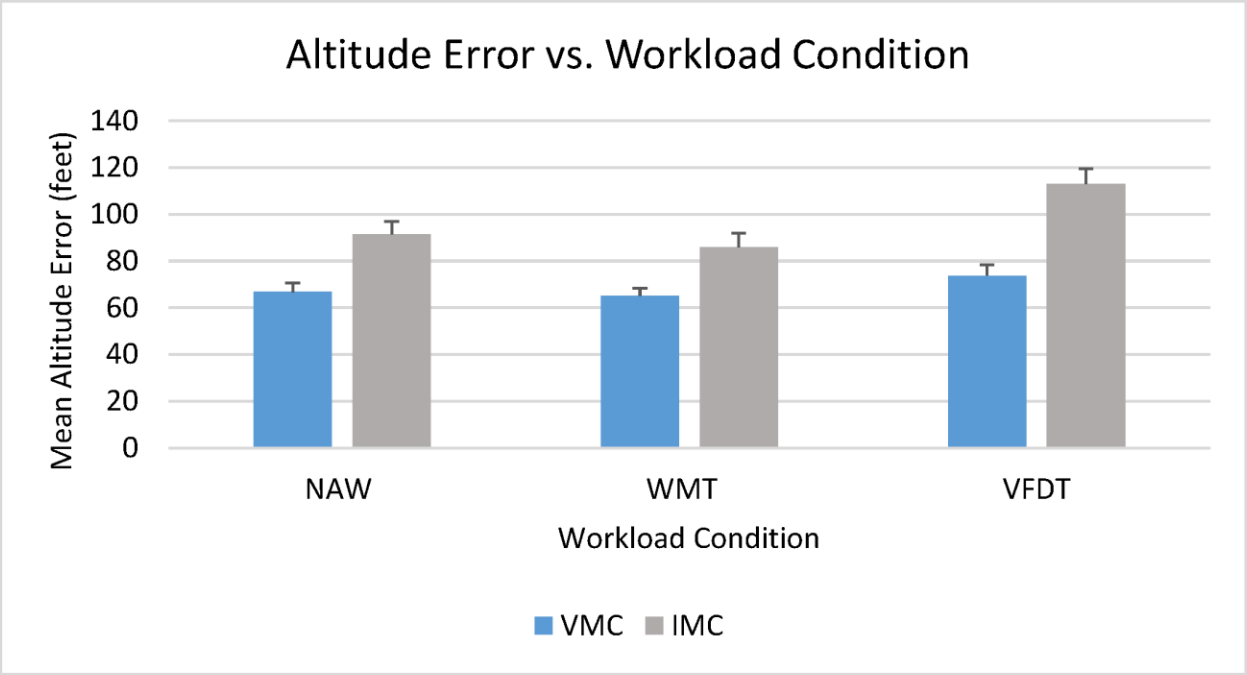

Figure 11. Mean altitude error collapsed across the motion conditions.

Post-hoc pairwise comparisons conducted using stepwise Holm-Bonferroni corrections to limit familywise error to $p < .05$ showed that the VFDT altitude error was significantly greater than both the NAW ($t(11) = 3.24$, $p = .02$) and WMT ($t(11) = 2.74$, $p = .04$) altitude error.

Secondary Task Performance

The means for percent correct on the WMT are plotted in Figure 12. Examination of the data indicated they met the assumptions for normality and a 2 x 2 repeated measures ANOVA was conducted, which found no significant effects for either motion or visibility conditions, with no significant interaction. The overall average percent correct was 89.82% (*SD* = 8.48)

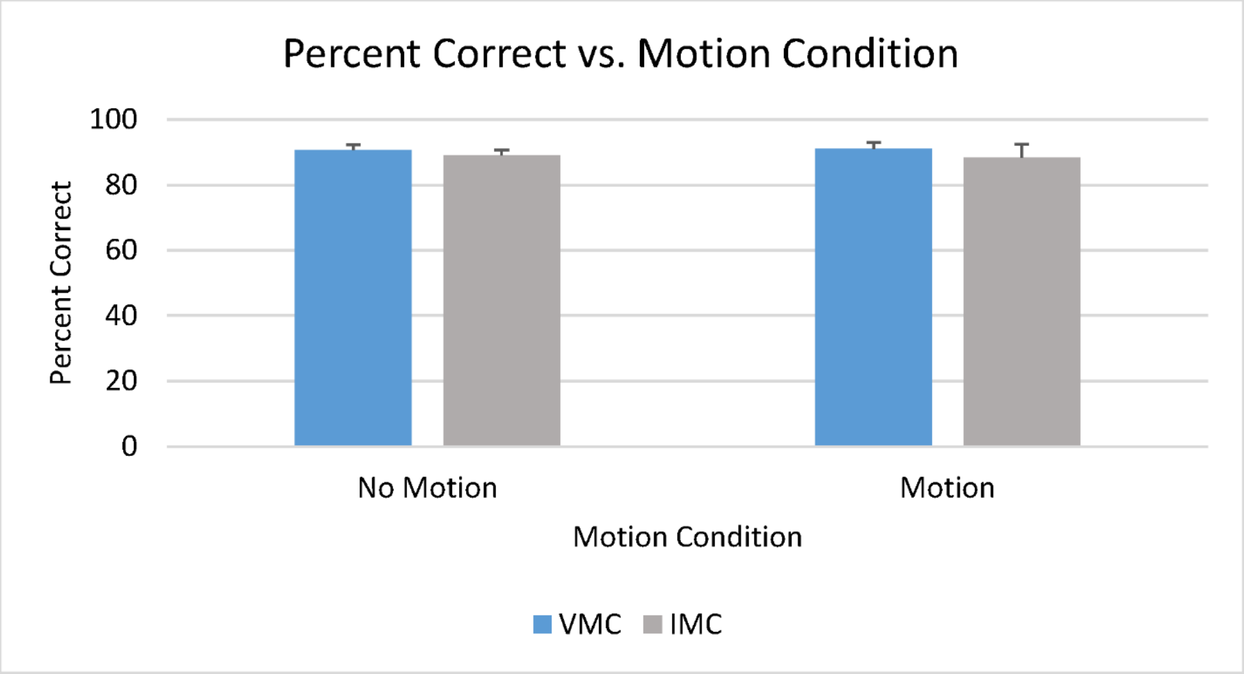

Figure 12. Mean percent correct on the WMT across the motion and visibility conditions.

An analogous ANOVA was indicated and conducted for the reaction time (RT) data, which are plotted in Figure 13. There was no significant effect for visibility condition, and no significant interaction, but RT's were significantly faster in the Motion condition *F*(1, 11) = 7.83, *p* = .02. The mean RT for the Motion condition was 1.478 s (*SD* = .22), while that for No-Motion was 1.597 s (*SD* = .17).

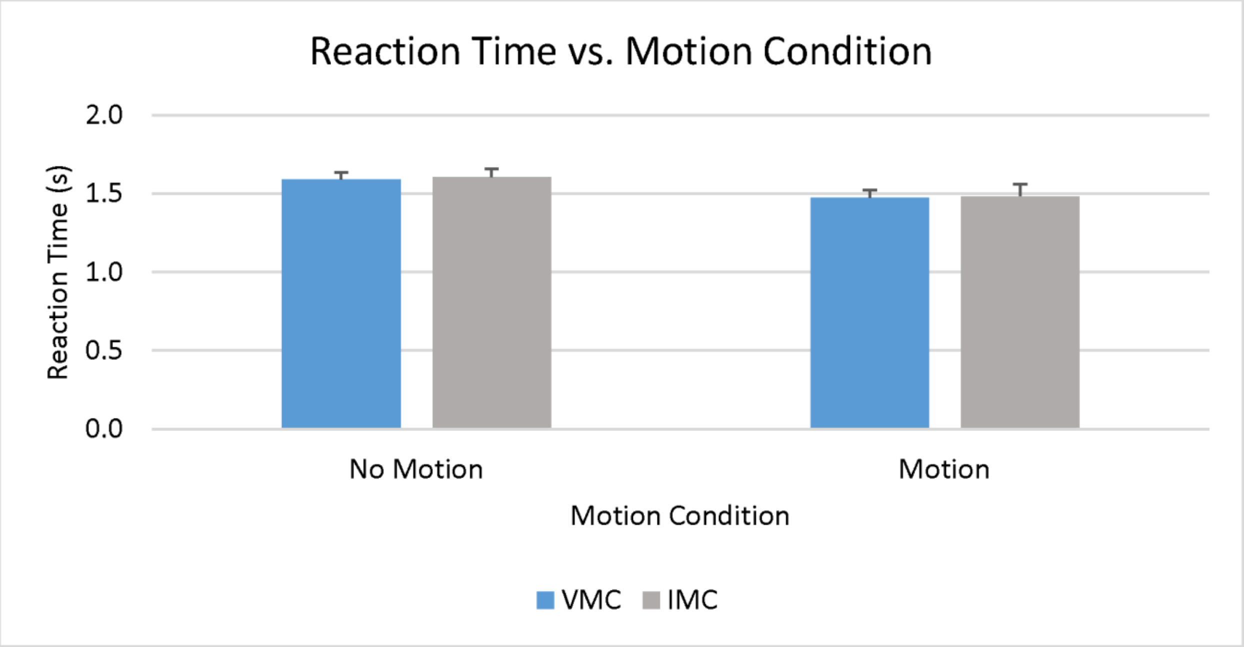

Figure 13. Mean reaction time on the WMT across the motion and visibility conditions.

For the VFDT data, distance error was defined as the absolute value of the difference between the command distance and the actual distance between the two aircraft. Comparing the VMC and IMC portions of the flights for these data would be invalid for three reasons. First, the flights began with zero following distance error and in VMC, creating a low-error bias for this visual condition. Second, there was only one following distance change in IMC, at approximately 28 sec before the lead disappeared. Because all other lead distance changes occurred at 40 sec intervals, pilots had less time to correct the error imposed by the one IMC distance change, creating a high-error bias for IMC. Third, due to an oversight, the pre-programmed average IMC distance change (533.3 ft) was greater than that for VMC (283.3 ft), again creating a high-error bias for the IMC portion of the flights. Therefore following distance error was calculated for the entire flight, collapsing across the VMC and IMC portions. Distance error was then calculated for each motion condition, and it was verified that the data met normality assumptions. A two-tailed paired-sample t-test showed that the mean for no-motion (275 ft, *SD* = 49.84) was not significantly different from that for motion (299 ft, *SD* = 76.39), with *t*(11) = -1.86, *p* = .09.

Discussion

The main goal of this research was to determine if different types of workload and simulator motion affected the incidence of SD, as measured by several different dependent variables. Below we discuss, in turn, the effects of workload and motion, followed by visibility condition effects where applicable.

Workload

It is first important to establish that workload was successfully manipulated by the introduction of the secondary cockpit tasks. The analyses of the subjective workload ratings did indeed indicate this to be the case. The workload ratings for the WMT and VFDT conditions were both significantly higher than the NAW condition (see Figure 5). Also, the workload ratings for the secondary tasks did not differ significantly from each other, indicating that they were generally similar in terms of the amount of workload imposed, as designed via early pilot testing.

Regarding SD ratings, there was no significant main effect of Workload, but there was a significant Workload x Visibility interaction. Figure 7 shows a larger (relatively) increase in SD ratings for the VFDT SD ratings across the VMC workload conditions, as compared to the step-like and even increases across the same IMC workload conditions. It may be that the increased feelings of SD were due to the necessity to focus on the third dimension of following distance, perhaps taxing spatial resources, but it is somewhat surprising that this trend occurred more strongly in VMC where pilots had a reliable visual horizon, versus in the clouds. However it should be kept in mind that although the nature of the significant interaction hints that this may be the case, the trend was not strong enough to result in a significant main effect.

The overall CRE rate in this study was 10.6%, which falls within the range of CRE rates found in previous studies (see Williams et al., 2014; Ercoline et al., 1994; Lincoln et al., 1972). Added workload did have a significant effect on the number of CRE’s in the No-Motion condition only, with the VFDT producing twice the number of CRE’s as the NAW condition. These elevated misperceptions of direction of bank, consistent with the elevated SD ratings in the VFDT discussed above, are indications of SD and again may be due to the requirement to use spatial resources to maintain the proper following distance.

Analyses of the effect of added workload on 65° AOB exceedance revealed a non-significant tendency for steeper bank angles in the VFDT condition with motion, compared to both the NAW and WMT conditions (see Figure 9). This tendency provides some further support for the idea that the VFDT, a spatial task, increased the incidence of SD, perhaps by taxing spatial resources. This explanation is consistent with Wickens’ (1984) multiple resource theory, which posits that humans have different types of cognitive resources, such as verbal and spatial resources. Because these resources are limited, two tasks simultaneously drawing from the same resource pool (e.g., two spatial tasks) are more likely to interfere with each other than two tasks drawing from different resources (e.g., a verbal and a spatial task). The VFDT and maintaining safe bank angles are both spatial tasks that likely interfered with one another, resulting in a tendency for more instances of severe bank angles in the VFDT condition.

For Altitude Error there was a significant main effect of workload, with error being greatest again in the VFDT condition as compared to both the NAW and WMT conditions. The WL x MET interaction was also significant, with Altitude Error being particular high in IMC in the VFDT condition. While Altitude Error is not necessarily a direct indicator of SD, minimizing it is a spatial task, as is the VFDT. The increased error, once more in the VFDT condition, is yet further evidence for spatial resource competition, wherein pilots may have relaxed Altitude Error tolerances in favor of maintaining the proper following distance and retaining spatial orientation.

Motion

In contrast to the workload manipulations, the addition of simulator motion had little effect on the dependent measures used in this study. The subjective workload and SD ratings were not significantly affected by motion. Given that no aggressive maneuvers were required in any of the flights, and that the lead aircraft did not exceed 45° AOB, DRD motion was limited and generally smooth and gentle. These factors likely dampened any tendency for differences in the subjective ratings of the motion and no-motion conditions.

Realism ratings did improve significantly with the addition of DRD motion. The average Motion condition score (7.63) was nearly double that for No Motion (4.04). This result indicates that the motion algorithms were well-designed and suitably matched to the flight scenario. This is encouraging, especially since this was only the second flight simulation study conducted in the DRD.

Neither 65° AOB Exceedance or Altitude Error showed a significant effect of simulator motion. Examination of Figure 9 might seem to suggest an effect of motion, showing that the condition with the greatest AOB exceedance was VFDT with motion. Although speculative, it could be the case that simulator motion in 3-dimensional (3-D) space created some spatial orientation challenges for the pilots, and adding the spatial and 3-D VFDT taxed spatial resources to the point where AOB exceedances increased slightly. However with no statistically significant main effect for motion and no interaction, such an explanation is tenuous, but as we will argue later, given the number of measures that showed the VFDT had at least a tendency to disrupt performance, that explanation may be worth considering.

For Altitude Error, visually inspecting the two panels in Figure 10 shows no indication for a motion effect, consistent with the statistics. Given the fact that minimizing altitude error was essentially a 2-Dimensional visual task (i.e., keeping the lead aircraft visually centered on the OTW display) there is not necessarily a compelling reason to expect that motion would have an effect.

Regarding the VFDT secondary task, there was no significant effect of motion on VFDT error. The mean error for the no-motion condition was 275 ft, while that for the motion condition was somewhat larger at 299 ft, an approximate 9% increase. The two-tailed paired-sample t-test results were $t(11) = -1.86$, $p = .09$, again approaching, but not reaching significance at traditional p value levels. As with the 65° AOB Exceedance results above, we again see some indication that the combination of motion with the spatial VFDT had a tendency, albeit not *statistically* significant, to degrade pilot performance.

For the WMT secondary task, there was no significant effect of motion on percent correct, and Figure 12 visually supports this conclusion. Overall, subjects posted a respectable 89.9% correct on this verbal memory task, indicating that they took the task seriously, and that it also was challenging enough to avoid a ceiling effect. There was a significant motion effect found in the WMT RTs, which were significantly faster in the Motion condition, showing a small (119 msec) but statistically significant increase in speed with no evidence of tradeoff in accuracy. The WMT was a non-spatial, verbal memory task, so motion would not be predicted to interfere with it as much as motion might interfere with a spatial task like the VFDT. However, it is unclear why WMT RT's would *improve* with motion, unless motion, in the absence of an additional spatial task, actually provided useful orientation information to the pilots, making the flying task easier and freeing up general attentional resources that could be applied to the WMT.

Except for the Realism Ratings and WMT RT results, which were not direct measures of SD, there was no statistically significant effect of motion. The accelerations and motions that pilots experience in actual flight can certainly be disorienting (Cheung, 2004), but there was no definitive indication that motion increased the probability of SD in this study. The most likely explanation is that the flight profile was not well suited to detect a motion effect. The profile required no aggressive or especially disorienting maneuvers. Therefore DRD motion was limited and benign, resulting in no significant increase in the likelihood of SD. Being a null result, the general lack of an effect for motion must not be interpreted to mean that motion is unimportant in SD studies; it just was not a critical variable in *this* study.

Meteorological Condition

Because the first portion of all flights occurred above the clouds in VMC, and the latter portion occurred in the clouds in solid IMC, the data were also analyzed according to visibility conditions. Workload ratings significantly increased in IMC (see Figure 5) as others have reported (see Braithwaite, Durnford, & Groh et al., 1998) and as would be expected, given that pilots no longer had a visual horizon for attitude reference. Accordingly, SD ratings increased as well in IMC, and Figure 7 shows that SD ratings more than doubled consistently in IMC across all workload conditions. This result emphasizes the importance of a visual horizon as a primary spatial cue for maintaining spatial orientation (see Patterson et al., 1997), since the only difference between VMC and IMC in this study was that the horizon was present in the former, but not the latter.

Altitude error was significantly higher in IMC than in VMC, and this result can be explained by the presence of the horizon in VMC. If a pilot's sight picture keeps a lead aircraft visually on the horizon, versus above or below it, he/she will be flying at the same altitude as the lead. This important visual spatial cue was absent in the clouds, making it more difficult to perceive and correct altitude errors.

Performance on the WMT, a verbal memory task, was not significantly affected by visibility condition in terms of either accuracy or speed. The formation flight task itself was a spatial task, and maintaining proper flight attitude was more challenging in IMC than in VMC, as indicated by SD and Workload ratings, as well as Altitude Error. However, IMC did not seem to impact WMT performance. This null result may be due the nature of the tasks. The WMT was verbal and the flight task was spatial, so the two tasks did not have to compete for the same cognitive resources.

Correct Perception of Motion Condition

Nine of the twelve subjects correctly identified which block of flights included motion, and three thought that all flights included motion. The confidence ratings ranged across the entire scale from 1 to 10, with a relatively high mean rating of 8.40. Interestingly the three who thought all flights included motion had their motion flights first, perhaps priming them to "perceive" motion in their second block as well. Regardless, this result speaks to the power of the visual system in its ability to induce illusions of motion in a static environment.

Comparison to the Previous NAMRU-D SD Study

The current study had many methodological elements in common with the study conducted by Williams et al. (2018), an important exception being that the earlier work did not include a motion condition. Here we compare the results of the two studies, using the No-Motion condition results from the current study.

Workload and SD ratings used the same scale for each study, and the results were very consistent in terms of the pattern and magnitude of the ratings. Workload ratings increased significantly when the secondary tasks were added, and SD ratings increased significantly in IMC.

For the earlier study, the criteria for a CRE was an errant stick input 15% or more of the maximum stick throw from the center/neutral position, and/or a 5° change in AOB in the wrong direction. These values were arbitrarily set. After visual inspection of the current study's data, it was decided to set the stick input threshold to 5% to allow the analysis of a larger set of data. In the previous study the overall CRE rate was 5.6% vs. 10.6% for the current study, a difference certainly influenced by the threshold change. More interestingly, in the previous study the WMT task yielded a significant increase in CREs, while the VFDT produced the significant increase in this study. Although the WMT here did show a slightly larger increase, the nature of the data and negative binomial sensitivities biased that result away from statistical significance. These mixed results are difficult to interpret but it is accurate to say that 1) in each study, adding workload produced a significant increase in CREs, but the type of workload producing that effect differed between studies, and 2) in both studies, adding either the WMT or VFDT showed a tendency for an overall increase in CREs, but that tendency was not always statistically significant.

In the 2018 study, the overall average percent time spent at or beyond 65° AOB was 2.64%, compared to 5.56% in the current study. A possible explanation for that difference is that 75% of the pilots in the previous study had an instrument rating, whereas 58% had one in the present study. Pilots with an instrument rating should be expected to have more effective instrument scan patterns. Also, pilots in the earlier study had a bit more flight time (M = 1708 hrs) than those in the current study (M = 1580 hrs).

Continuing with 65° AOB exceedance, in the 2018 study the VFDT showed a significant increase in steeper bank angles compared to the WMT condition. In the current study there was a non-significant trend in that same direction, but only in the Motion condition. These results are difficult to reconcile, except to say that across the two studies there was a consistent tendency for the VFDT to disrupt performance on the spatial task of maintaining safe bank angles.

Altitude Error in the 2018 study was greater in VMC than in IMC, but exactly the opposite was true for the current study. There is no ready explanation for this difference. However, added workload did similarly affect Altitude Error across the two studies. In the 2018 study the VFDT (and the WMT) did significantly increase Altitude Error over the NAW condition, and in the current study the VFDT increased Altitude Error over the NAW (and WMT) condition. In other words, in both studies the VFDT did significantly increase Altitude Error over at least the NAW condition.

Performance on the secondary tasks common to each study was notably consistent. Mean percent correct on the WMT was 88.4% in the first study, and 89.9% in the second. Repeatability of results for mean RTs was even more remarkable at 1.592 s vs. 1.597 s. Finally the VFDT also showed excellent consistency, with an average following distance error of 285 ft in the first study, and 275 ft in the second. In addition to the repeatability of Workload Ratings across the two studies, the fact that these performance numbers varied so little in two studies, conducted in two different simulators two years apart indicates that these tasks are good candidates for inducing additional workload, and indicates that the procedures and methodology used in both studies were sound.

Summary

In this study the inclusion of DRD motion did not significantly increase measures of SD, but adding the VFDT did. Overall, there were five indications that the VFDT, a spatial task, increased the likelihood of SD, or at least taxed spatial resources that are important in maintaining orientation. First, the number of CRE's increased significantly in the VFDT condition. Second, Altitude Error increased significantly in this condition, an indication that the availability of sufficient spatial resources for performing both tasks was strained. Third, for SD ratings there was a significant Workload x Visibility condition interaction which appeared to be due to elevated SD ratings in the VFDT in VMC. One way to view this result is that *even* with a visible horizon, the VFDT tended to increase feelings of SD. Fourth, for 65° AOB Exceedance in the clouds, there was a tendency for increased bank angles in the VFDT versus the NAW condition, but the main effect for Workload was not statistically significant ($F(2, 22) = 2.50$, $p = .11$). However, examination of Figure 9 shows that the combination of the VFDT plus Motion resulted in the most time spent at extreme bank angles in the clouds, which is a dangerous position in which to be. Fifth, again for the VFDT paired with Motion, there was an increase in following distance error that approached, but did not reach significance ($t(11) = -1.86$, $p = .09$). Given the other indications that the VFDT was disorienting, we argue that these last two measures, though not statistically significant by traditional standards, should be given some consideration.

Based on the results of this study and the findings of Williams et al. (2018), several recommendations can be made. First, pilots should be made aware that added workload in the cockpit can increase the probability of SD and entry into unsafe attitudes, such as high AOB in the clouds. Extra attention should be given to maintaining a safe attitude when workload increases and visibility decreases. Second, it should be known that the added workload need not be spatial in nature to negatively affect spatial orientation. While it was the spatial VFDT in this study that led to increased measures of SD, both the VFDT and the non-spatial WMT provoked increased SD in the Williams et al. (2018) study. Finally, the envelope of motion in future DRD studies should be expanded to include realistic flight maneuvers that are known to bring on SD. Doing so can allow the testing of SD countermeasures, such as improved display design (e.g., helmet mounted display symbology) that will help reduce this deadly threat to aviation safety.

References

Bartel, C. & Foster, D. (2004). Life Cycle Motion Base Cost Comparison: Electric vs. Hydraulic. *SimTecT 2004: Simulation - Better Than Reality*.

Benson, A. J. (1989). Spatial disorientation - general aspects. In J. Ernsting, A. Nicholson, & D. Rainford (Eds.) *Aviation Medicine* (pp. 419-436). Oxford: Butterworth-Heinemann.

Boeing (2019). Fatalities by CICTT aviation occurrence categories: 2018 statistical summary, http://www.boeing.com/news/techissues/pdf/statsum.pdf

Braithwaite, M. G., DeRoche, S. L., Alvarez, E. A., & Reese, M. A. (1997). *Proceedings of the First Triservice Conference on Rotary-Wing Spatial Disorientation: Spatial Disorientation in the Operational Rotary-Wing Environment*. U.S. Army Aeromedical Research Lab, Fort Rucker, AL.

Braithwaite, M. G., Durnford, S. J., Crowley, J. S., Rosado, N. R., & Albano, J. P. (1998). Spatial disorientation in US Army rotary-wing operations. *Aviation, Space, and Environmental Medicine, 69*(11), 1031-1037.

Braithwaite, M. G., Durnford, S. J., Groh, S. L., Jones, H. D., Higdon, A. A., Estrada, A., & Alvarez, E. A. (1998). Flight simulator evaluation of a novel flight instrument display to minimize the risks of spatial disorientation. *Aviation, Space, and Environmental Medicine*, *69*(8), 733-742.

Braithwaite, M. G., Ercoline, W. R., & Brown, L. (2004). Spatial disorientation instruction, demonstration, and training. In F. Previc, & W. Ercoline (Eds.), *Progress in astronautics and aeronautics: Spatial disorientation in aviation* (Vol. 203, pp. 323-377). American Institute of Aeronautics and Astronautics.

Bramble, W. J. (2008, October). *Spatial disorientation accidents in large commercial airplanes: Case studies and countermeasures*. Paper presented at the 61st annual International Air Safety Seminar (IASS) meeting, Honolulu, HI.

Cheung, B. (2004). Nonvisual illusions in flight. In F. Previc, & W. Ercoline (Eds.), *Progress in astronautics and aeronautics: Spatial disorientation in aviation* (Vol. 203, pp. 243-281). American Institute of Aeronautics and Astronautics.

Collins, W., & Dollar, C. (1996). *Fatal general aviation accidents involving spatial disorientation: 1976-1992* (Report No. DOT/FAA/AM-96/21). Oklahoma City, OK: Federal Aviation Administration-Office of Aviation Medicine.

De Winter, J. C., Dodou, D., & Mulder, M. (2012). Training effectiveness of whole body flight simulator motion: A comprehensive meta-analysis. *The International Journal of Aviation Psychology*, *22*(2), 164-183.

Edens, T. J., & Higginbotham, M. D. (2014, March). Degraded visual environments: a leading factor in aviation accidents. *Army Aviation Magazine, 31*, 22-23.

Ercoline, W. R., Weintstein, L. F., DeVilbiss, C. A., & Bitton, D. F. (1994, October). Attitude information (orange peel) on the head-up display for recovery from unusual attitudes. *Proceedings of the SAFE Association 32nd Annual Symposium*, Reno, NV.

Gibb, R.W., Ercoline, W. R., & Scharff, L. (2011). Spatial disorientation: decades of pilot fatalities. *Aviation, Space, and Environmental Medicine, 82*(7), 717-724.

Gibb, R.W., Gray, R., & Scharff, L. (2010). *Aviation visual perception: research, misperception, & mishaps*. Ashgate Publishing Ltd.

Jentsch, F., Curtis, M., & Salas, E. (2011). *Simulation in aviation training (Critical essays on human factors in aviation).* Routledge.

Kallus, K. W., Tropper, K., & Boucsein, W. (2011) The importance of motion cues in spatial disorientation training in VFR-pilots. *The International Journal of Aviation Psychology, 21*(2), 135-152. https://doi.org/10.5701080/10508414.2011.556458

Kirkham, W. R., Collins, W. E., Grape, P. M., Simpson, J. M., & Wallace, T. F. (1978). Spatial disorientation in general aviation accidents. *Aviation, Space, and Environmental Medicine, 49*(9), 1080-1086.

Koonce, J. M. (1974). *Effects of ground-based aircraft simulator motion conditions upon prediction of pilot proficiency: Part I*. (Report No. AFOSR-TR-74-1292).

Lawson, B. D., Smith, S. A., Kass, S. J., Kennedy, R. S., & Muth, E. R. (2003). Vestibular stimuli may degrade situation awareness even when spatial disorientation is not experienced. *Proceedings of the RTO HFM Symposium on Spatial Disorientation in Military Vehicles* (pp. 1-21, RTO-MP-086). La Coruna, Spain.

Lincoln, R., Palmer, E., & Wempe, T. (1972). *Effect of pictorial display configuration on the frequency of control reversals during aircraft landing approaches* (Report No. NASA-TM-X-62, 191). NASA.

Mortimer, R. G. (1995). General aviation airplane accidents involving spatial disorientation. In *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting* (pp. 25-29). Santa Monica, CA: Human Factors and Ergonomics Society.

Newman, R. L., & Rupert, A. H. (2020). The magnitude of the spatial disorientation problem in transport airplanes. *Aerospace Medicine and Human Performance, 91*(2), 65-70.

Patterson, F. R., Cacioppo, A. J., Gallimore, J. J., Hinman, G. E., & Nalepka, J. P. (1997). Aviation spatial orientation in relationship to head position and attitude interpretation. *Aviation, Space, and Environmental Medicine, 68*(6), 463-471.

Poisson, R.J., & Miller, M. E. (2014). Spatial disorientation mishap trends in the U.S. Air Force 1993-2013. *Aviation, Space, and Environmental Medicine, 85*(9), 919-924.

Raj, A. K., Kass, S. J., & Perry, J. F. (2000). Vibrotactile displays for improving spatial awareness. In *Proceedings of the International Ergonomics Association XIV Triennial Congress/44th Annual Meeting of the Human Factors and Ergonomics Society* 44(1), 181-184.

Roscoe, A. H. & Ellis, G. A. (1990). *A subjective rating scale for assessing pilot workload in flight: a decade of practical use* (Royal Aerospace Establishment Report No. 90019). Royal Aerospace Establishment.

Singh, B., & Navathe, P. D. (1994). Indian Air Force and world spatial disorientation accidents: a comparison. *Aviation, Space, and Environmental Medicine, 65*(3), 254-256.

Stott, J. R. R. (2013). Orientation and disorientation in aviation. *Extreme Physiology & Medicine, 2*(1), 2.

Tormes, F. R., & Guedry, F. E. (1975). Disorientation phenomena in naval helicopter pilots. *Aviation, Space, and Environmental Medicine*, *46*, 387-393.

Vaden, E. A., & Hall, S. (2005). The effect of simulator platform motion on pilot training transfer: A meta-analysis. *The International Journal of Aviation Psychology*, *15*(4), 375-393.

Véronneau, S. J., & Evans, R. H. (2004). Spatial disorientation mishap classification, data, and investigation. Spatial Disorientation in Aviation. In F.H. Previc & W.R.Ercoline (Eds.). *Spatial Disorientation in Aviation: Progress in Aeronautics and Astronautics*, (Vol. 203, pp. 197-241). AIAA.

Webb, C. M., Estrada, A., & Kelley, A. M. (2012). The effects of spatial disorientation on cognitive processing. *International Journal of Aviation Psychology*, *22*(3), 224-231.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & R. Davies (Eds.), *Varieties of attention* (pp. 63–101). Academic Press.

Williams, H. P., Horning, D. S., Lawson, B. D., Powell, C. R., & Patterson, F. R. (2018). *Effects of various types of cockpit workload on incidence of spatial disorientation in simulated flight*. (Report No. NAMRU-D 19-06). NAMRU-D.

Williams, H. P., Littman, E. M., Folga, R. V., & Patterson, F. R. (2014). *Field of view evaluation for flight simulators used in spatial disorientation training* (Report No. NAMRU-D 14-41). NAMRU-D.

Wobbrock, J.O., Findlater, L., Gergle, D. and Higgins, J.J. (2011). *The Aligned Rank Transform for nonparametric factorial analyses using only ANOVA procedures*. Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '11). Vancouver, British Columbia (May 7-12, 2011). ACM Press, pp. 143-146.