NAMRU-D-19-29 Dec-2018

NAVAL MEDICAL RESEARCH UNIT DAYTON

Examining Perceptual Mechanisms in the Black Hole Illusion

by F. Eric Robinson, Henry P. Williams, Dain Horning, and Adam T. Biggs

Naval Medical Research Unit Dayton

NOTICES

Disclaimer
The views expressed in this article reflect the results of research conducted by the author and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, nor the United States Government.

Source of Support
This work was funded by an In-house Laboratory Independent Research Program, administered by the Office of Naval Research.

Human Research/Institutional Review Board (IRB) statement
The study protocol was approved by the Wright-Patterson Air Force Base Institutional Review Board in compliance with all applicable federal regulations governing the protection of human subjects.

Abstract

The Black Hole Illusion (BHI) continues to persist despite decades of research into the visual cues that guide landing approaches. Hypotheses attempting to explain the effect tend to focus on individual cues rather than the broader spatial strategies that guide their use. This report details our efforts to resolve apparent inconsistencies in the literature regarding the visual cues that affect nighttime approaches, as well as evaluate the potential for a novel hypothesis (the Line Bias Effect) to account for BHI wherein pilots may misestimate the location of the horizon based on the edges of the runway under BHI conditions. We present findings from a flight simulator study in which qualified pilots flew nighttime approaches under different combinations of starting distance from the runway, starting altitude, and runway length. We conclude that longer runways lead to lower approaches, and that pilots may misperceive the apparent size and/or shape of the runway. Further, we detail the results of a series of computer-based perceptual studies in which participants estimated the intersection points of various lines. These studies indicate that participants had difficulty in judging line angles, lending credence to the notion that pilots may misjudge the horizon based on the runway edge lines at night. We discuss the findings from all of our studies in the context of a broader spatial strategy to examine the factors that may cause pilots to experience BHI effects.

Contents

Background

According to a worldwide study of airline accidents, the approach and landing phase of aviation accounts for 52% of all accidents/fatalities despite covering only 4% of total flight time (Boeing, 2006). These accidents are also three times more likely to occur at night than during daytime (Khatwa & Helmreich, 1998). Deficient visual cues at night sometimes lead to spatial disorientation, which can be defined as an erroneous sense of one's position relative to the earth's surface (Gillingham, 1992). Visual judgments can have a large effect on behavior, despite the fact that people are generally poor at making them (Andrade, 2011). Specifically, people are very poor at making comparative judgments of glideslope, especially under nighttime conditions (Murray, Allison, & Palmisano, 2009). One particular misperception of glideslope - the black hole illusion (BHI) - is regularly cited as a leading cause of spatial disorientation (Sipes & Lessard, 2000), if not the leading cause of in-flight visual-spatial problems (Matthews, Previc, & Bunting, 2002).

BHI occurs during nighttime approaches with minimal visual cues (no visible terrain features, horizon, etc.; Perrone, 1983). Under such conditions, pilots tend to overestimate their altitude, leading them to fly to a low approach path and in extreme cases impact terrain or obstacles short of the runway. Unfortunately, BHI is poorly understood from a theoretical perspective. A myriad of individual visual cues have been proposed that may influence BHI, and there is broad consensus that multiple cues and strategies likely regulate visual descents. However, there is little agreement about the relative weighting of those cues and no overall description of the spatial strategy which serves to guide how such cues are used. The goals of the studies presented here were to help clarify possible environmental contributors to BHI and to investigate the cognitive mechanisms involved in low approaches during nighttime landings. By emphasizing the cognitive contributors underlying BHI, we hope to eventually provide a mechanistic explanation for BHI effects and facilitate interventions. We begin with an overview of the visual scene and the cues available during a visual approach with good visibility. Next we describe several hypotheses proposed to explain BHI effects, along with our own hypothesis. We conclude the introduction with a brief overview of the four studies presented in this report. Each study is then described in more detail individually.

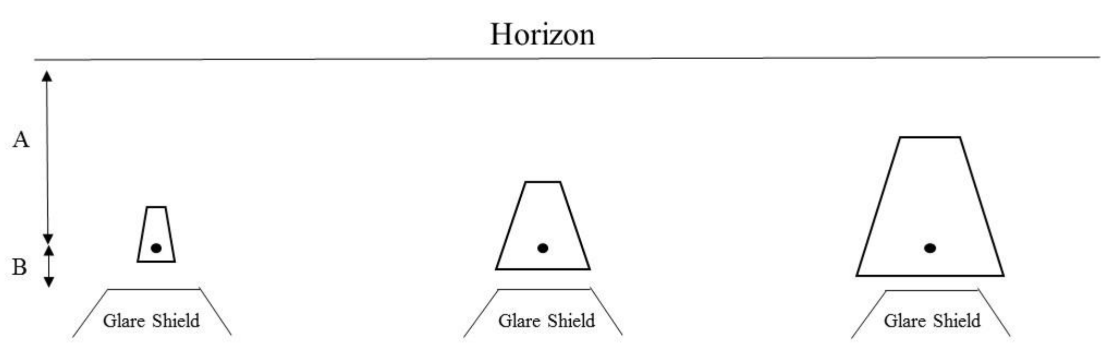
Visual Cues in a Standard Approach

Numerous visual cues have been identified that contribute to pilots' ability to fly a visual approach. The horizon, aim point, runway geometry, optic flow, and the aircraft itself (particularly the glare shield) all provide useful information to the pilot when judging glide path. Interrelationships among these cues provide information that allows pilots to judge glideslope and maintain a constant approach path. Given a typical daylight approach with good visibility, pilots make spatial judgments by utilizing the aim point (the location on the runway where the pilot intends to touch down) as the primary visual cue. The glare shield, the runway itself, and the horizon serve as secondary cues that provide information based on their relationship to the aim point (Patterson, Williams, Folga, & Arnold, 2015).

Figure 1 illustrates some of the visual cues pilots may use to guide their approach. For instance, the locations of the aim point, horizon, and glare shield remain stationary in the visual field if the pilot maintains a stable approach path. This visual relationship between these three cues leads to two indicators of glide path: absolute H angle refers to the visual angle between the aim point and the horizon, and relative H angle refers to the visual angle between the aim point and the glare shield. The horizon remains stable in the visual field because it appears to be at eye level (Lintern & Liu, 1991). The aim point remains stable because it is the point to which the aircraft is flying. Maintaining a constant absolute or relative H angle therefore ensures a stable glide path; manipulating either can affect glideslope performance (Galanis, Jennings, & Beckett, 1998; Jacobs, Morice, Camachon, & Montagne, 2018; Murray, Allison, & Palmisano, 2009).

In addition, the aim point provides a stable source point from which optic flow is generated. Optic flow generates a sense of movement through space; disruptions to optic flow can affect the ability to interpret distance (Nicholson & Stewart, 2013). The form or aspect ratio of the runway can also serve as a reliable cue during a landing approach. This ratio is the ratio of apparent runway width to runway length. During the approach, the near and far runway thresholds appear to move down towards the glare shield and up towards the horizon, respectively, whereas the edge lines appear to splay out over time. Maintaining a constant aspect ratio during the approach results in a constant glideslope (Galanis, Jennings, & Beckett, 1998; Murray, Allison, & Palmisano, 2009). In sum, the interaction of many cues provides the pilot with the necessary information to fly a visual approach under conditions of good visibility. As long as the pilot maintains stable H angles, the correct form ratio of the runway, and the proper rate of optical expansion, the aircraft should follow a proper glideslope and rate of descent.

Figure 1. During a typical daylight approach, the absolute H angle (line A) and relative H angle (line B) remain constant in the visual field while the near and far runway thresholds expand towards the glare shield and horizon.



Visual Cues in BHI

Hypotheses to explain BHI tend to focus on the lack of cues or the inability to interpret cues without contextual features under nighttime conditions. For example, Thompson (2010) describes a hypothesis wherein pilots try to maintain a constant visual angle of the runway in the visual field. This hypothesis essentially states that pilots maintain the visual “height” of the runway, possibly as an attempt to substitute for H angle when the horizon isn’t visible. This notion allows for testable hypotheses. For instance, approach paths calculated based on this strategy broadly aligned with past simulator studies. Thompson also predicts that short runways should see worse effects. Gibb (2007) reviewed several hypotheses as well:

1. A long and narrow runway may give a false perception of height and distance without contextual visual cues. All else being equal, a runway will appear longer in the visual field when viewed from a higher altitude. Difficulty in relating a 2D retinal image to the 3D world reduces the pilot’s ability to use perceptual constancy as a cue under BHI conditions.
2. Featureless terrain may impair the pilot’s ability to use objects’ relative retinal size as a cue.
3. Prior knowledge of a long runway may conflict with an apparently small visual angle, leading to an overestimation of runway length and angle of descent. This leads to an overestimation of visual angle based on the retinal image and knowledge of the runway’s actual length-in-depth. The runway will appear longer, similar to hypothesis 1.
4. Lack of terrain cues may make the runway appear to “float” in space, making judgments of depth, distance, and altitude difficult (i.e., autokinesis). A lack of cues reduces the ability to use relative retinal comparisons for size, shape, and depth constancy.

5. Featureless terrain may remove distance cues, altering perceptions of runway size, shape, and/or depth. This is similar to hypothesis 4.
6. Lack of visual cues impairs a pilot's ability to judge the slope of the runway and terrain. Distance, depth, and orientation cannot be perceived under BHI conditions, meaning that optical slant is the sole cue regarding the actual slant of the runway. Optical slant is relative to the surface, which cannot be perceived under BHI conditions. Slant may affect how long the runway appears (see hypothesis 1).
7. Approach lighting systems may appear to increase the length of the runway, leading it to appear longer and narrower than it actually is (related to hypothesis 1).
8. Lack of visual cues may cause overestimation of runway slant. Objects appearing together are perceived at the same distance in the absence of other visual cues, favoring the foreshortened frontal plane and resulting in exaggerated perceptions of slant.
9. Pilots may use the wrong cues to estimate glide path and runway slant during BHI conditions. Perrone (1982; 1983) asserts that short, wide runways should produce more reliable judgments than long, narrow runways.

The hypotheses described by Gibb can be grouped based on whether they emphasize the appearance of the runway itself or the impact of impoverished contextual cues around the runway. Common among many of these hypotheses, however, is an emphasis on retinal imagery and comparisons among the images to explain the mechanism of BHI effects. Though not the sole proposed mechanism, these explanations emphasize the perceived retinal image caused by the (lack of) visual cues and how that may impact pilots' ability to judge altitude. This makes intuitive sense because as Gibb, Schvaneveldt, and Gray (2008) point out, the only visual cues available to estimate altitude and distance in BHI conditions are the size and shape of the retinal image of the runway. However, many hypotheses tend to treat the information from the retinal image in isolation and neglect the influence of how that image is processed as part of a broader spatial strategy.

The "Line Bias" Hypothesis

Rather than emphasize the physical appearance of the visual cues on the retina in isolation, we emphasize the visual scene as a whole and how pilots may process that information as part of a broader spatial strategy to judge glide path. Summarized another way, pilots may be attempting to use contextual and peripheral cues as part of a broader strategy even when these visual cues are not present, which inherently leads to poor spatial judgment due to the appearance of the runway.

We test this hypothesis by examining the horizon more specifically as a visual cue and how its absence may influence pilots' glide paths. We propose that pilots use the same spatial strategy regardless of day or nighttime visual conditions, and that in the absence of a visible horizon pilots generate an implicit horizon based on the runway edge lines. Eye tracking evidence suggests that pilots fixate on cues such as the runway and horizon at similar rates in both day and night conditions (Kim, Palmisano, Ash, & Allison, 2010), implying similar strategies govern both day and night approaches. Perspective and compression gradients provide sufficient information to judge the location of the horizon implicitly (Lintern & Liu, 1991), and eye tracking data is consistent with the possibility that pilots use an implicit horizon (Kim, Palmisano, Ash, & Allison, 2010).

We hypothesize that pilots use the runway edge lines and horizon together as a reference point during a daytime visual approach, with the horizon serving as a stable reference point for interpreting the splay of the edge lines. Without this reference point under poor visual conditions, we propose that pilots may experience what we have termed a "line bias" illusion. Under the "line bias" illusion, pilots may overestimate the splay of the edge lines and make an inaccurate estimate of the horizon's location in the visual field (and thus misperceive absolute H angle). Further, estimates of the implicit horizon (and thus perceptions of H angle) may change over time as the appearance of the runway changes during the course of the approach. In the context of a strategy where the H angle should be stable and consistent, such a misperception would lead the pilot to alter their glideslope.

The Current Studies

This report describes a series of four studies designed to clarify some of the environmental hypotheses reviewed above as well as the proposed line bias illusion. Consistent with Galanis and colleagues (1998), we focus our efforts on manipulating variables relevant to runway length, the perception of the horizon, and apparent slant of runway edges. Study 1 uses a series of simulated nighttime approaches to comparatively test some of the hypotheses described earlier. Study 2 examines the plausibility of the line bias illusion using a basic perceptual judgment task. Studies 3 and 4 examine potential mechanisms behind the line bias effect based on the perception of the inclination of simulated runway edge lines.

Study 1

Overview

Study 1 aimed to evaluate a subset of existing hypotheses in order to identify the most relevant contributors to BHI. In particular, we focused on hypotheses related to runway length and the presence of visual contextual cues. Though we cannot directly test each individual hypothesis against each other with a single experiment, we can ascertain the relative strengths of the hypotheses using fairly simple manipulations.

The constant visual angle hypothesis described by Thompson (2010), as well as hypotheses 1, 3, 7, and 9 reviewed by Gibb emphasize the effect of runway aspect ratio (especially runway length) and the impact this may have on pilots' perception of altitude. Thompson asserts that a short runway should cause greater BHI, whereas the other hypotheses predict that longer runways should lead to more severe BHI. Kim and colleagues (2010) found no effect of runway length on glide slope, but other studies have shown high-aspect ratio (i.e., long, thin) runways lead to greater risk of BHI (e.g., Gibb, Schvaneveldt, & Gray, 2008; Lintern & Koonce, 1991; Lintern & Walker, 1991). These hypotheses can be tested together by manipulating runway length. Depending on the direction of effect (if any), we can rule out one or more of the length-based hypotheses.

Similarly, hypotheses 2, 4, and 5 reviewed by Gibb (2007) suggest that the lack of distance cues in featureless terrain increases the risk of BHI, but the literature has been equivocal on this hypothesis as well. Gibb, Schvaneveldt, & Gray (2008) tested the effect of terrain features directly and found that the presence of terrain features does not improve glideslope performance. However, Kim et al. (2010) found that pilot gaze fixations toward the ground responded to pilot experience and lighting conditions, suggesting that ground-based contextual cues are relevant to glideslope performance. Lintern and colleagues have investigated the impact of ground features on simulated landings and found that approaches tended to be higher with greater contextual detail and lower with less detail (Lintern &Koonce, 1991; Lintern & Walker, 1991). Lintern & Koonce (1991) concluded that a lack of scene detail may permit a low approach rather than cause one. In other words, lack of detail may not cause a pilot to fly too low, but can prevent a pilot from realizing the error.

We indirectly test hypotheses based on contextual features by evaluating their logical implications. Collectively, hypotheses focused on ground-based contextual features assert that pilots cannot accurately judge distance and runway size/shape without visual context provided by the terrain. Therefore, pilots should be unable to perceive subtle changes in runway-based cues over relatively short distance intervals during the approach and should be unable to adjust glideslope in response to starting at subtly different points relative to the runway. We test this prediction by manipulating the starting position of the aircraft. If pilots are unable to interpret the runway without contextual features, we should expect to observe different glide paths for different starting positions because pilots would not be able to fly to a consistent glideslope. If the absence of contextual cues leads the runway to "float" in space (Hypothesis 4 in Gibb), error should be more or less randomly distributed. If the absence of contextual cues prevents pilots from properly interpreting other visual cues (e.g., the shape of the runway; Hypothesis 5 in Gibb), error should be stable if perceptions are systematically biased and random if they are not. Depending on the consistency and error of glideslopes across the various starting positions, we can make inferences about the relative merits of the hypotheses described in Gibb (2007).

Each of the current hypotheses described previously is grounded in sound reasoning based on the workings of the human visual system, but further testing is needed to resolve apparent discrepancies in the empirical literature and better define the potential effects of a given visual cue. Study 1 aims to help resolve conflicting predictions and findings in the literature, as well as clarify the contribution of various cues to BHI effects. Certified pilots flew a series of approaches to runways of varying lengths under BHI conditions, starting at different combinations of altitude and distance from the runway. Eye tracking was also utilized to better describe the spatial strategy guiding the pilots' approaches.

Method

Participants. A total of 19 male volunteers ranging in age from 35 to 62 (mean age 52.6) participated in this study. Participants were military and civilian personnel from on and around

Wright-Patterson AFB, OH who had soloed in an aircraft. Participants had an average of 3,864.5 flight hours (min: 220; max: 23,000). The maximum value of 23,000 hours represented nearly 3x the flight hours of the next highest person. Without the extreme outlier of 23,000 hours, the average number of flight hours dropped to 2,801.4. Participants reported normal or corrected to normal vision and no history of simulator-induced sickness. Two participants reported being left handed, and two separate participants reported being left eye dominant. Three participants did not report eye dominance.

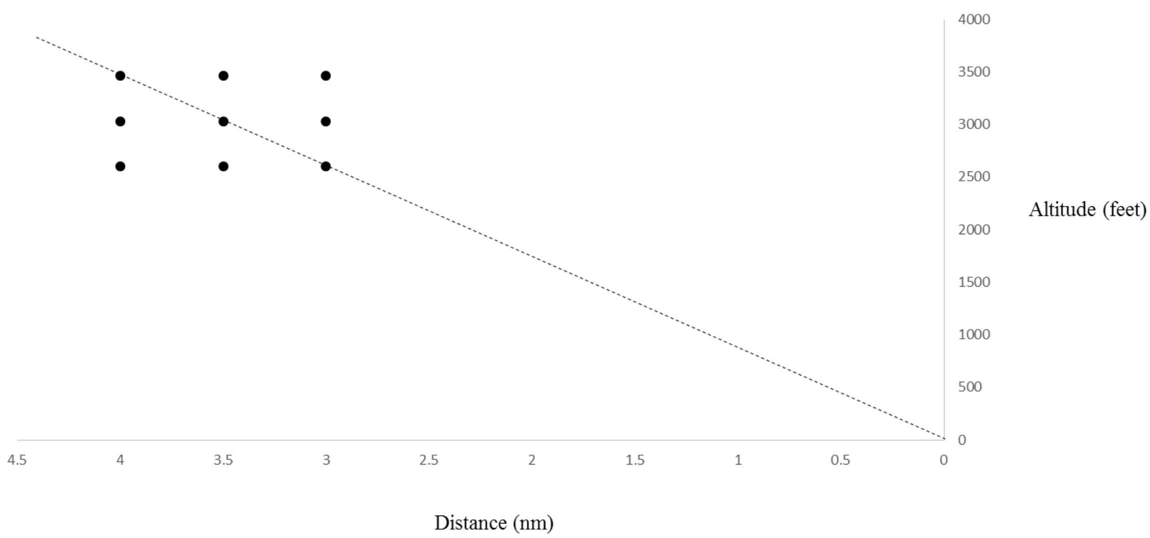
Apparatus. Participants flew simulated landing approaches using a fixed-base flight simulator emulating a T6 Texan (a small, single engine propeller aircraft) operated via X-Plane software. The flight instruments were displayed on a 26 inch diagonal ELO monitor, while the outside-the-cockpit view was displayed on a 60 inch diagonal Samsung LED High Definition TV, providing an 87º wide by 49º high field of view. A FitPC3Pro drove the outside the window scene graphics. Participants sat in an open cockpit on a SPARCO seat that was adjustable in height. Control inputs were made via a Thrustmaster Cougar joystick and Thrustmaster Warthog throttle.

Participants' head and eye movements were tracked and recorded using an ISCAN AA-ETL-600 Head and Eye Tracking System. This system uses a camera mounted to a frame on the participants' head to determine fixation point within ± 2º.

Task. Participants flew simulated landing approaches into the airfield at Naval Air Station Fallon, NV (runway dimensions: 200 feet wide x 14,000 feet long). The aircraft started paused in midair at one of nine possible combinations of distance from the runway (3, 3.5, or 4 nm), altitude (2,598; 3,031; and 3,464 feet) and runway length (8,000, 10,000 or 12,000 feet).

The participants were informed that these values would change from flight to flight, but were not informed of the specific distances, altitudes, or runway lengths. The participants' task was to assume control of the aircraft and fly as close as possible to a 3^{0} approach. One-third of the starting points were above glide path, one-third were below glide path, and one-third were on glide path (Figure 2). Thus, pilots were forced to make a decision regarding their position relative to the proper glide path and could not simply rely on a given power setting and airspeed to perform a proper approach. Mild turbulence was also present. The instrument panel display was modified such that the vertical speed indicator and glideslope indicators were unavailable. In addition to runway length manipulations, the airfield was modified such that approach lighting and Precision Approach Path Indicators (PAPIs) were unavailable. Pilots were thus forced to rely on their own perception, judgment, and piloting skills to maintain a standard 3^{0} glide slope.

Figure 2. Starting locations for the flight simulator task. The dotted line represents a 3^{0} glide path.



Design. This study utilized a 3 (starting distance) x 3 (starting altitude) x 3 (runway length) within subjects design, resulting in 27 separate flights. Condition presentation order was randomly determined for each participant, as the high number of trials and relatively low number of participants made counterbalancing impractical.

Experimental procedure. Participants reported to the laboratory at the Naval Medical Research Unit Dayton (NAMRU-D) for one session. Upon arrival, participants were given the opportunity to read the informed consent document and ask any questions. Participants then completed a demographic questionnaire to determine flight hours and confirm flight qualifications.

After signing the informed consent document and completing the demographic questionnaire, participants were led to the flight simulator. The simulator was adjusted to a comfortable seat height and rudder pedal distance for each participant. The participant then completed two practice approaches during daylight conditions. The first was an autopilot approach that began eight nm out from the airfield. The airfield was displayed at its full length of 14,000 feet. This approach allowed the participant to see what a nominal airspeed, power setting, and sight picture for a 3º approach should be in the simulated aircraft. Following the autopilot flight, participants flew the same daylight approach without the autopilot to allow them to get a sense of the simulation's handling characteristics and to act as a baseline measure of the pilots' ability to fly approaches without the PAPI lights, vertical speed indicator, or other glide slope guidance. Participants did not receive feedback regarding this flight, or any of the subsequent flights.

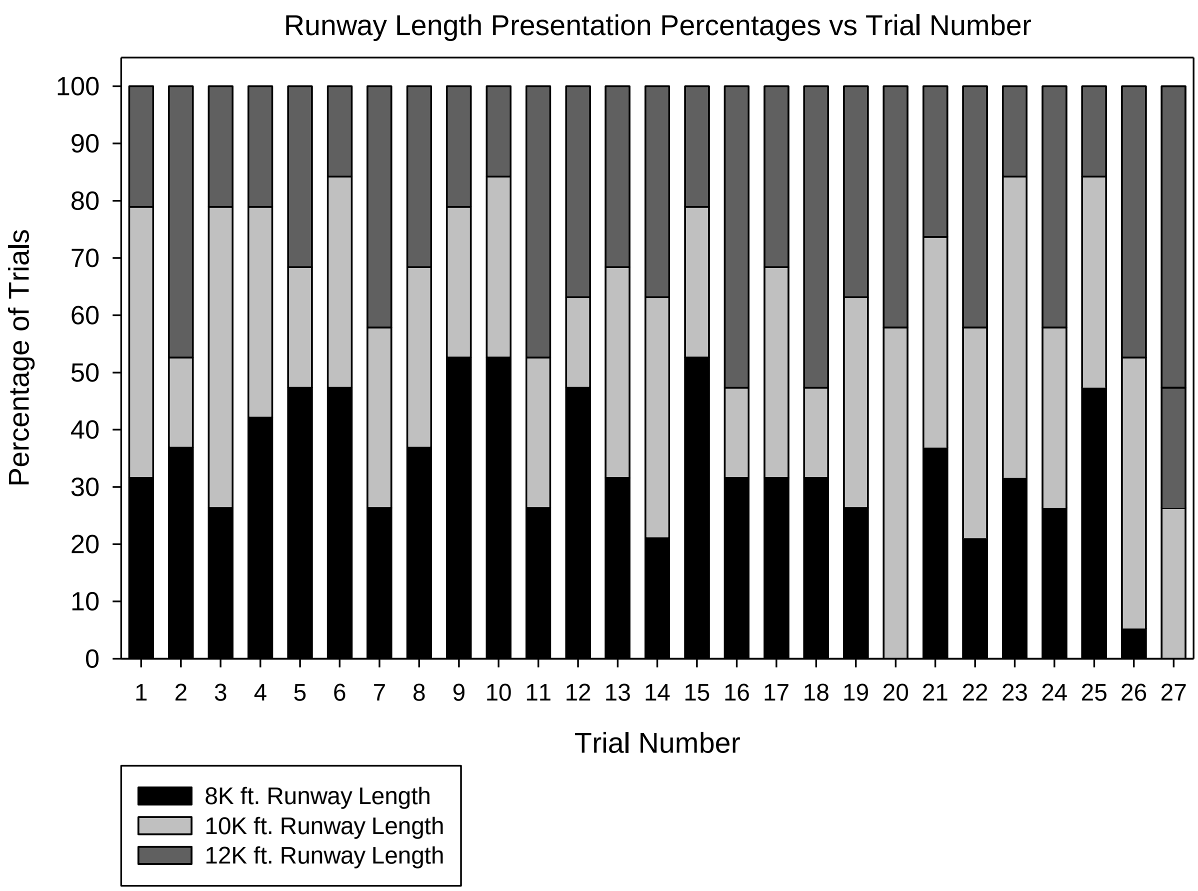
Upon completion of the two daylight flights, participants completed the series of 27 nighttime approaches under black hole conditions with the sky obscured and no visible stars, no visible horizon, and no cultural lighting. Participants flew in a darkened room with nighttime instrument illumination. Breaks were offered every nine trials or as requested. Each nighttime approach lasted approximately two minutes, depending on starting distance. After the 27 flights were completed, the participants were debriefed, thanked, and dismissed.

Analyses and Results

Confirming presentation order. Because we did not actively counterbalance condition presentation order, we first confirmed that our conditions were indeed randomly distributed throughout the trials in order to rule out any potential order effects. Chi square tests of independence were performed for each parameter (starting distance, starting altitude, and runway length) to confirm that the level of each parameter was not related to trial number. Chi square tests for starting distance (χ^2 (52) = 42.63, $p = 0.82$) and starting altitude (χ^2 (52) = 42.95, $p = 0.81$) indicated that starting levels of these variables were indeed randomly distributed across the 27 trials. We therefore are not concerned with order effects in the overall starting position of the aircraft across the 27 trials.

Chi square tests for runway length indicated that condition order for this variable may not have been independent of trial number (χ^2 (52) = 69.79, $p = 0.05$). We therefore examined the distribution of variables within the early, middle, and late trials to better determine whether starting runway length was distributed unevenly across the duration of the study more globally, or only on a trial-to-trial basis. Chi square tests grouping trials into thirds (i.e., trials 1-9, 10-18, and 19-27) indicated that runway length was indeed distributed unevenly across the study as a whole (χ^2 (4) = 11.82, $p = 0.02$). We therefore examined the frequency distribution of runway length across the 27 trials to better visualize the extent of any possible order effects (Figure 3). Examination of the frequency distribution indicated trial-to-trial fluctuations in the presentation of each runway length, but no obvious trend that would immediately raise concerns about learning or order effects. Each runway length is represented throughout the course of the 27 trials. In spite of the results of the chi square tests, we therefore believe it is unlikely that differences in the frequency of runway length across all 27 trials were an issue to the extent that participants would show learning effects or other bias that may influence study results.

Figure 3. Frequency of each runway length per trial expressed as a percentage.



Flight simulation data. Visual inspection of individual subject plots indicated that 0.5 nm provided enough distance to allow pilots to intercept the proper glide path from any starting location. In order to maintain consistency between flight profiles and remove any error due purely to starting location, all analyses were therefore performed on flight data from 2.5 to zero nm away from the runway. Upper and lower tolerances for the target altitude were calculated based on a 3º glide slope ± 0.5°, consistent with previous research (Jacobs, Morice, Camachon, & Cyril, 2018; Lintern & Liu, 1991). The lower bound of the target altitude was calculated as the distance to runway multiplied by the tangent of 2.5°, and the upper bound of the target altitude was calculated as the distance to runway multiplied by the tangent of 3.5°. Target altitude was calculated in real-time during data capture in LabVIEW (National Instruments) according to target altitude = distance from runway*tan(3°). All subsequent calculations and processing were performed using a custom MATLAB (Mathworks) script. Flight data was sampled at a rate of 10 Hz. Examination of the data indicated that Subject 15 was an extreme outlier (> 3 SD) in one condition and a moderate outlier in two other conditions. Subject 15 was therefore excluded from the analyses described below.

The outcome measure for Study 1 was calculated as a weighted sum of all altitude error (distance from the target altitude on a 3º glide slope) between 2.5 nm out and when the flight was stopped immediately prior to touchdown:

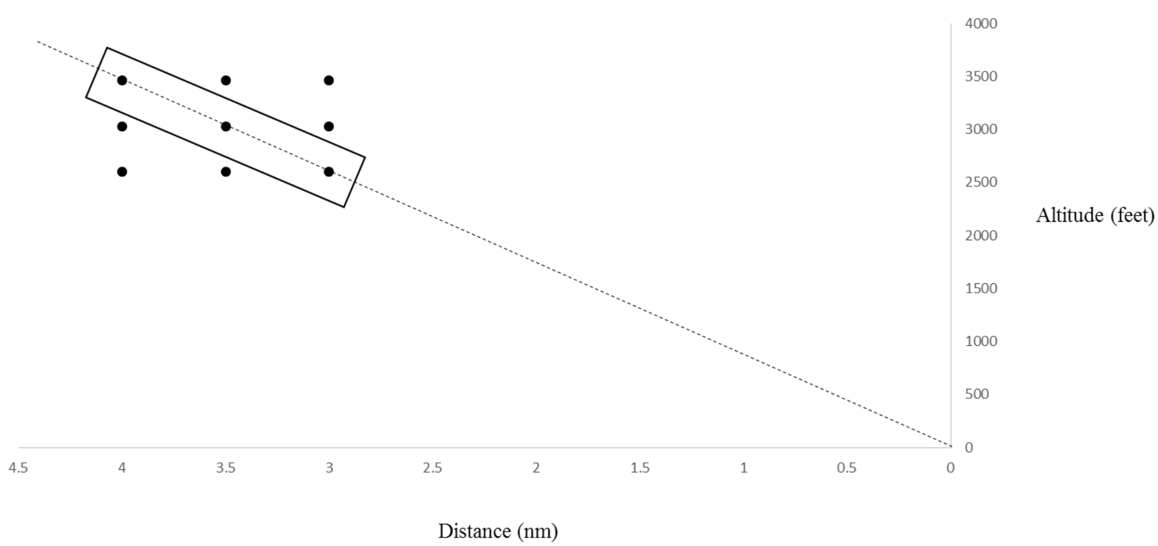
$$Error = \sum_{i} \frac{(Altitude_i - Target\ Altitude_i)}{Distance\ from\ Runway_i}$$

Negative values therefore represent error below the glide path, with increasingly negative values representing more severe BHI.

Our design confounded starting distance and starting error (closer distances were likely to start above glide path and farther distances were likely to start below glide path). We therefore ran separate ANOVAs examining the effects of starting distance and starting altitude without the confounding conditions included, along with an ANOVA examining all starting points in the design.

Unconfounded analysis of the effects of starting distance. We first examined the effects of starting distance using a 3 (starting distance) x 3 (runway length) repeated measures ANOVA, including only the starting positions of the aircraft that fell along the 3º glide slope (Figure 4). Table 1 shows the results of this ANOVA. We found a significant main effect of runway length, but no main effect of starting distance. The interaction was also not significant. Pairwise comparisons (uncorrected) indicated that error increased significantly with each increase in runway length (Figure 5; Table 2).

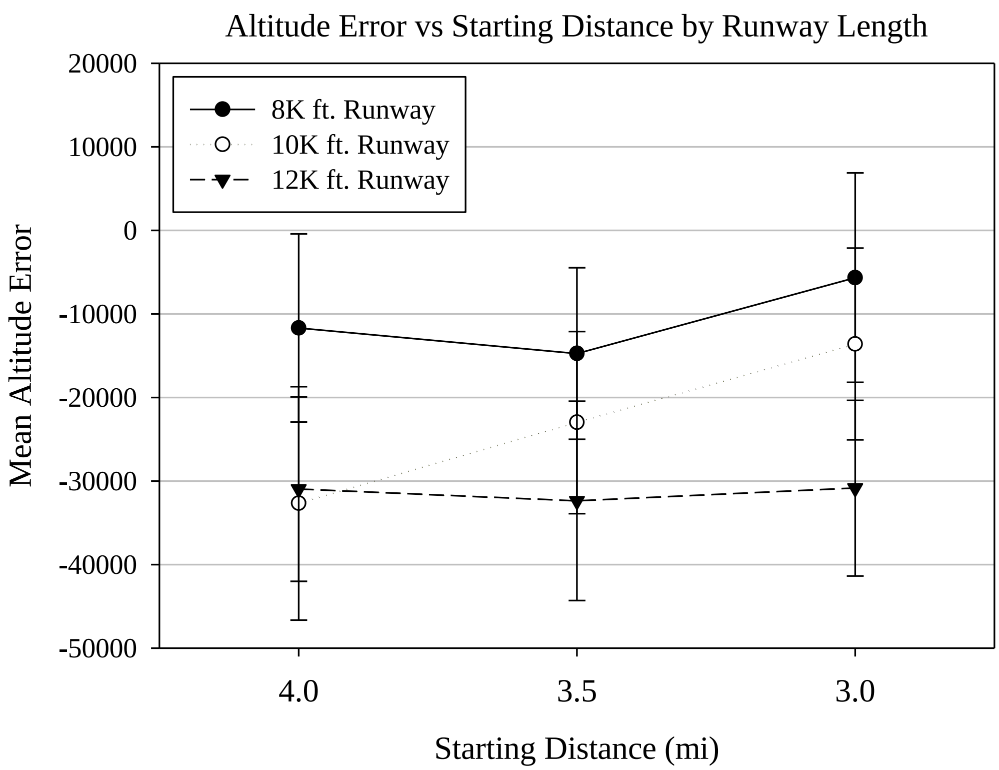
Figure 4. Starting locations analyzed for the unconfounded distance analysis. Points within the box are all along the glide path, thus starting error remains constant across starting points.



Variable	F(df)	p	η_p^2
Starting distance	1.56 (2, 34)	0.23	0.08
Runway length	9.98 (2, 34)	< 0.01	0.37
Distance x runway length	0.55 (4, 68)	0.70	0.03

Table 1. ANOVA results examining the effect of starting distance and runway length.

Figure 5. Main effect of runway length on altitude error in the unconfounded starting distance ANOVA.

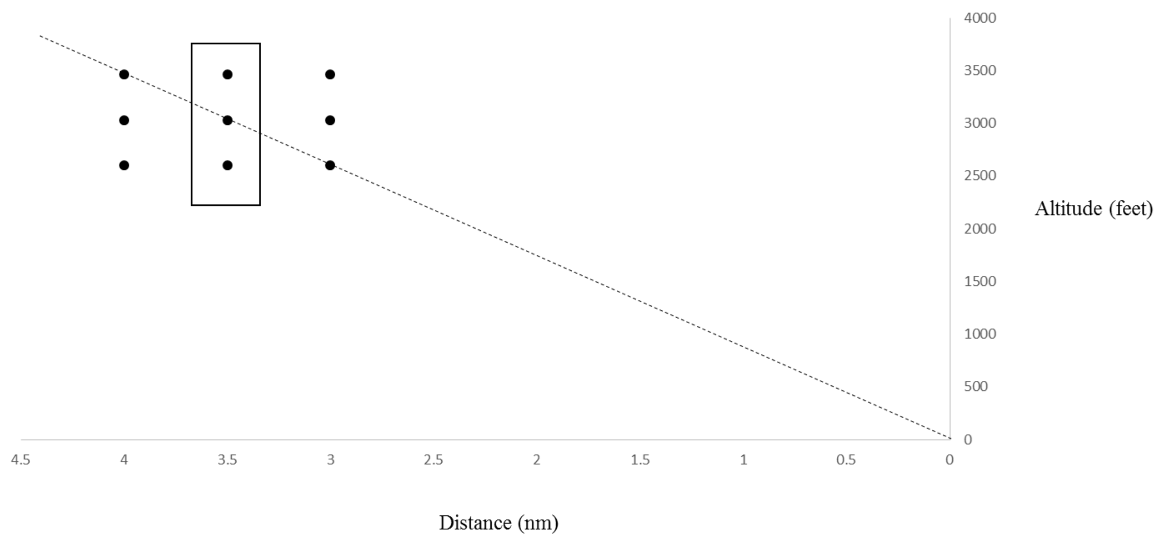


Comparison	Mean Diff.	Std. Error	p
8k - 10k	12,399.43	5,371.92	0.034
8k - 12k	20,718.47	4,741.96	< 0.001
10k - 12k	8,319.04	3,747.25	0.040

Table 2. Pairwise comparisons for the effect of runway length in the unconfounded starting distance ANOVA.

Unconfounded analysis of the effects of starting altitude. We next examined the effect of starting altitude, using only the starting locations at the middle starting distance (i.e., the highest altitude was above glide path, the middle altitude was on glide path, and the lowest altitude was below glide path; Figure 6). A 3 (starting altitude) x 3 (runway length) repeated measures ANOVA indicated a significant main effect of starting altitude. The main effect of runway length only approached significance in this analysis. The interaction was again non-significant (Table 3; Figure 7). Uncorrected pairwise tests indicated that starting on or below the glide path caused greater BHI error than starting above the glide path (Table 4). Despite the non-significant omnibus test, pairwise comparisons likewise indicated that error was greater for the 12,000 foot runway than the 8,000 foot runway (Table 5).

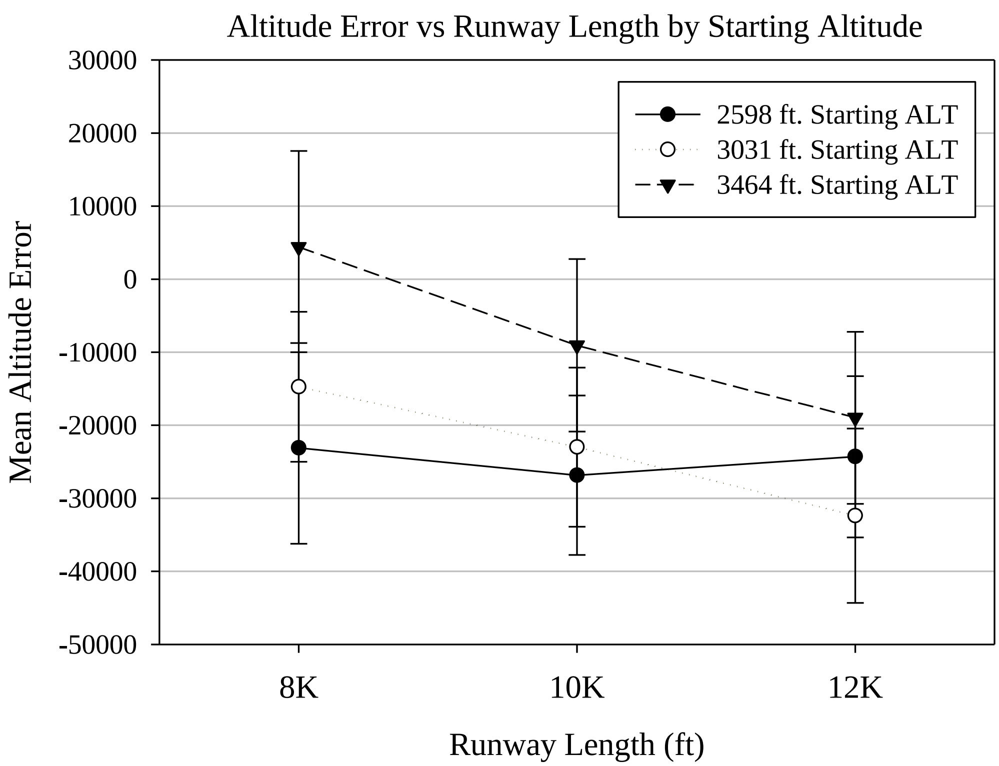
Figure 6. Starting locations for the unconfounded altitude analysis. Starting points within the box are all at the same distance and altitude error above and below the glide path are equal.



Variable	F(df)‡	p	η_p^2
Starting altitude	5.30 (1.41, 24.04)	0.02	0.24
Runway length	2.83 (2, 34)	0.07	0.14
Altitude x runway length	0.99 (4, 68)	0.42	0.06

Table 3. ANOVA results examining the effect of starting altitude and runway length.
‡Greenhouse-Geisser correction used to adjust for violation of sphericity.

Figure 7. Main effect of starting altitude.



Comparison	Mean Diff.	Std. Error	p
Low – On path	-1,385.83	5,439.70	0.802
Low - High	-16,882.54	4,087.77	0.001
On path - High	-15,496.71	7,289.85	0.048

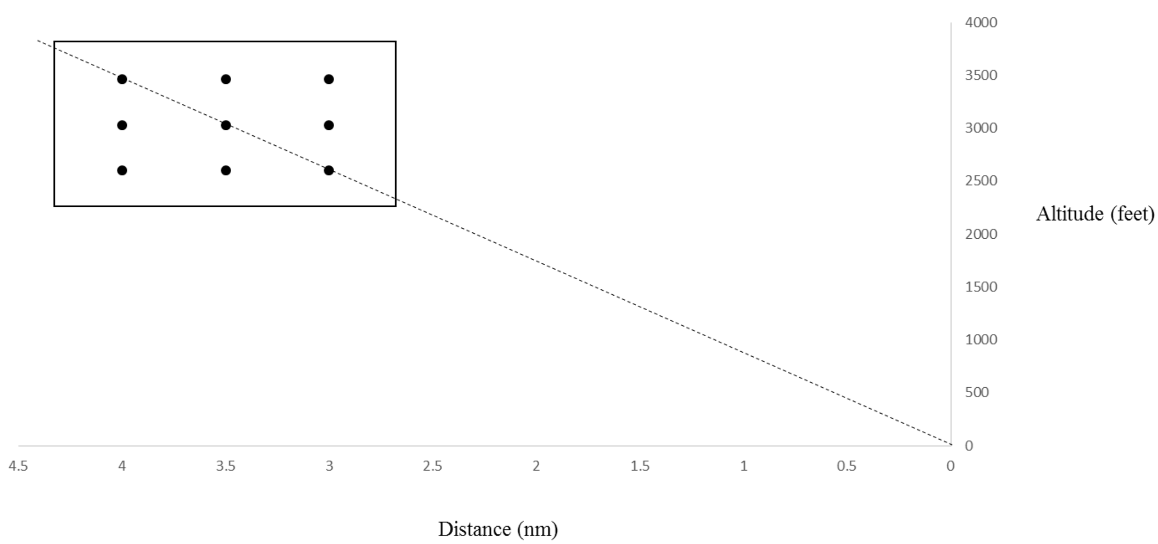
Table 4. Pairwise comparisons for the effect of starting altitude in the unconfounded starting altitude ANOVA.

Comparison	Mean Diff.	Std. Error	*p*
8k - 10k	8,487.08	5,973.97	0.173
8k - 12k	14,079.36	6,522.13	0.045
10k - 12k	5,592.28	5,331.03	0.309

Table 5. Pairwise comparisons for the effect of runway length in the unconfounded starting altitude ANOVA.

Full ANOVA with all starting points included. As an exploratory analysis, we also analyzed total error using a 3 (starting altitude) x 3 (runway length) x 3 (starting distance) repeated measures ANOVA (Figure 8). This analysis also serves as a stronger test of the effect of runway length because it utilizes the full data set (increasing power), and the effect of runway length is isolated from the confounding of starting distance and starting altitude. We found significant main effects for starting altitude, starting distance, and runway length, but no significant interactions (Table 6). Negative error (indicating a low flight path) increased when pilots started at low altitudes, started far from the runway, and with longer runways (Figures 9 and 10).

Figure 8. Full ANOVA using all nine starting locations.



Variable	F(df)	p	η_p^2
Starting distance	14.02 (1.45, 26.02)‡	< 0.01	0.44
Starting altitude	12.29 (2, 36)	< 0.01	0.41
Runway length	13.53 (2, 36)	< 0.01	0.43
Distance x altitude	0.92 (4, 72)	0.46	0.05
Distance x runway length	0.23 (4, 72)	0.92	0.01
Altitude x runway length	0.87 (4, 72)	0.49	0.05
Distance x altitude x runway length	1.33 (8, 144)	0.23	0.07

Table 6. ANOVA results for Study 1. ‡Greenhouse-Geisser correction used to adjust for violation of sphericity.

Figure 9. Main effects of starting distance and runway length.

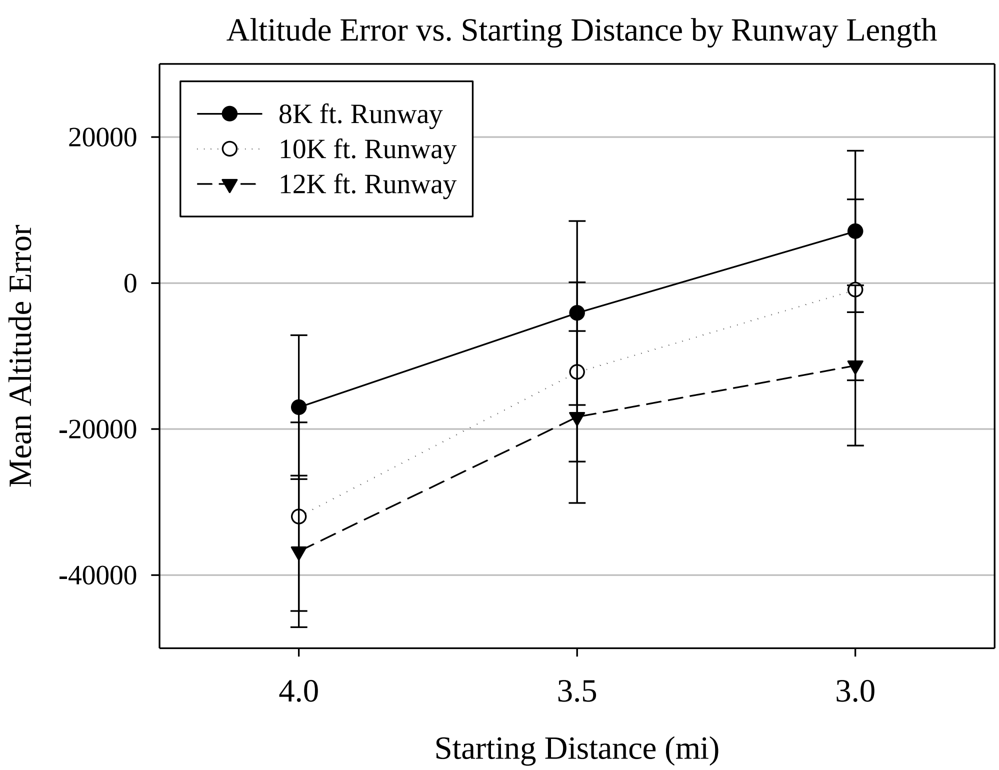
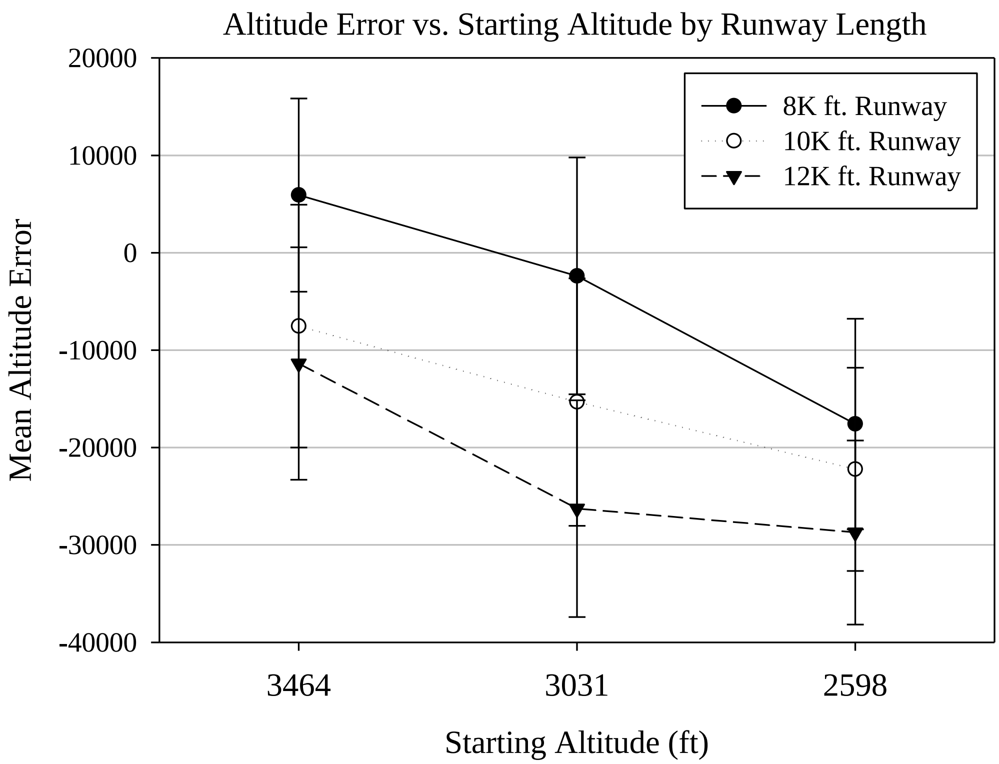


Figure 10. Main effects of starting altitude and runway length.



Eye tracking data. The eye tracking data was collected to evaluate where participants focused attention when looking outside the cockpit. Such information was collected in the hope that it would guide interpretation of the results of Studies 2-4 and evaluate the overall plausibility of the line bias hypothesis. Unfortunately, the eye tracker proved insufficient to support a quantitative analysis for a variety of reasons. First, the resolution of the tracker (± 2º) was not fine enough to distinguish which segment of the runway the participants focused on for the majority of the approach. Second, the data stream collected by our IOS did not include the coordinates of the eye tracker, such that quantitative analysis was not feasible. Finally, even if such data had been captured, the changing locations of the various parts of the runway in the visual field over the course of the approach coupled with the varied length of time consumed by each approach would have made it prohibitively difficult to quantitatively evaluate which part of the runway was at a given location on the screen at any given time to match with the eye tracking data. Therefore, we conducted a qualitative analysis based on video of the eye tracker recorded during the flights.

A total of 119 videos were reviewed to evaluate trends in eye movements over the various approaches. This represents a sample of 23% of the total number of videos available. All participants were included in this sample of 119 videos. However, some participants' data was insufficient for any evaluation (e.g., the eye tracker malfunctioned and failed to record the fixation point, calibration issues or glare from eyeglasses caused the cursor to be unreliable or off the runway, etc.). Ten participants had sufficient data to make reasonable judgments regarding their eye positions during the approach, leading to a final sample size of 104 videos. These 104 videos from 10 participants were used to generate the eye tracking results described here.

Of the 10 participants included in the eye tracking analysis, all demonstrated a tendency to focus on the closer half of the runway towards the near threshold/aim point. However, seven participants showed possible evidence of glancing to the further half of the runway towards the far threshold and even beyond the runway close to the horizon. These glances tended to be infrequent and brief, and also appeared to occur more often near the end of the approach just prior to touchdown when the near threshold was very low in the visual field. However, such glances were also noted during early portions of at least some approaches for five participants.

Discussion

Study 1 examined the effect of various starting altitudes, starting distances from the runway, and runway lengths on pilots' ability to follow a 3^{o} glide slope under BHI conditions. We found that longer runways and lower starting altitudes were associated with greater flight path error. Despite a significant main effect during the 3 x 3 x 3 ANOVA, we do not believe that starting distance from the runway affected glide path error due to the confounded nature of this analysis (discussed below). We also found some evidence in videos of the eye tracking data that suggests pilots may refer to the horizon even if it is not visible. We discuss these findings in the context of existing hypotheses to explain BHI, followed by some limitations of the study.

Implications for the effects of runway length. The ANOVA examining the effect of starting distance and the full ANOVA examining the effect of all variables together indicated that longer runways were associated with greater error. Likewise, pairwise comparisons in the starting altitude ANOVA revealed greater error for the 12,000 foot runway than the 8,000 foot runway. Given the greater power of the full 3 x 3 x 3 ANOVA, along with the significant main effect and trend in the smaller ANOVAs, we believe this finding to be sufficiently convincing to discuss it as valid here. This finding supports hypotheses 1, 3, 7, and 9 described by Gibb (2007), which collectively indicate that longer runways (or perceptions of longer runways) in the absence of other contextual cues may cause BHI effects by causing the runway to appear as it would be seen when the pilot is high and far away. Although we physically manipulated runway length, a similar effect would likely occur if the runway merely *appears* longer to the pilot. In contrast, the constant visual angle hypothesis described by Thompson (2010) does not appear to be supported.

ANOVAs unaffected by confounding of starting distance and altitude indicated that starting distance from the runway did not impact levels of error on the flight sim task. We did find a significant main effect of starting distance in the 3 x 3 x 3 ANOVA, but we discount this finding due to the aforementioned confounding of distance and starting altitude. Given the significant effect of starting altitude in the unconfounded ANOVA, we believe the effect of starting distance in the larger (confounded) ANOVA is driven by the altitude effect. Two of three starting points at greater distances started below the prescribed glide path, and two of three points at shorter distances started above it. This pattern matches the observed altitude effects, in which lower altitudes were associated with greater error (lower flight paths). Among the three starting altitudes, the lower two altitudes were statistically similar, whereas the highest altitude demonstrated less flight error.

Implications for the effects of terrain cues. Our manipulation of starting distance and altitude is most directly relevant to Hypotheses 4, and 5 in Gibb (2007), which state that the lack of contextual objects and terrain features causes the runway to appear to "float" in space and inhibits the ability to determine the aircraft's position relative to the ground or runway (hypothesis 4), and that lack of distance cues such as global and local features prevents the proper perception of size/shape/depth constancy of the runway (hypothesis 5). We did not directly manipulate the presence of contextual features, so our interpretation of these effects should be taken with caution. However, we feel that some insight can be drawn based on our findings.

Pilots in our study started 3.0, 3.5, or 4.0 nm out from a runway of variable length, in the absence of any contextual visual information. Visual differences between starting conditions were therefore very subtle, and indeed multiple subjects noted that they didn't perceive any changes from one trial to the next. If the lack of contextual features prevents pilots from accurately perceiving the runway, it is logical to assume that error would accumulate over time as pilots would begin to deviate from glide path and find it difficult to notice and correct these errors over the distances used in our study. It is therefore reasonable to expect glideslope to vary across starting position in our study because pilots would not be able to fly a consistent glide path based on the visual cues available. The "floating runway" described in hypothesis 4 implies error should be randomly distributed. Particularly in the case of starting distance, error should have begun accumulating almost immediately, leading to more error at greater starting distances from the runway. In contrast, the "misinterpretation" described in hypothesis 5 allows for the possibility of either systematic bias or random error.

Our finding that starting distance did not affect total error over the last 2.5 nm of the approach indicates that pilots' final flight path was similar regardless of how far they had to fly. Pilots' glide paths over the last 2.5 nm of the approach were also consistent whether starting at the lowest or middle starting altitude; only the highest starting altitude was different[1]. The lower two starting altitudes led to significantly greater flight simulator error than the highest starting altitude in both ANOVAs, whereas the two lower altitudes were not significantly different from one another in either of the ANOVAs. Pilots seem to follow a stable low glide path under BHI conditions. If they start below the prescribed glide path, they will stay there. More interestingly, the lack of difference between the lower altitudes in the unconfounded ANOVA indicates that even if the pilot starts on the 3º glide path they will tend not to maintain it, and instead adopt a lower flight path. Kim et al. (2010) found similar effects of starting altitude.

Pilots' ability to fly a consistent (even if incorrect) approach across the various starting locations suggests pilots are able to use one or more visual cues to fly a stable approach even without contextual features. This discounts Hypothesis 4 as reviewed by Gibb (2007). Rather than random error, pilots appear to demonstrate a systematic bias in the interpretation of visual cues under BHI conditions. Despite this consistency, the inaccuracy of the approaches observed

[1] We speculate that the highest altitude was different because pilots tried to avoid making steep descents. The highest starting altitude would have required a fairly steep descent to correct the glidepath at the closer starting distances. It is possible that the pilots had not finished correcting by the time they entered our 2.5 nm analysis window.

in this study indicates that contextual features may help in the interpretation of the visual scene, even if the contextual features themselves are not the primary cue. Hypothesis 5 as described by Gibb (2007) therefore remains plausible, consistent with the conclusions of Lintern & Koonce (1991). Pilots appear to systematically misinterpret the size or shape of the runway under BHI conditions.

Eye tracking data. Finally, the eye tracking data indicated that pilots tended to focus on the runway, particularly the near half. However, at least some participants appeared to glance toward the far threshold of the runway and even beyond that to the horizon. The eye tracking results should be considered with extreme caution, but they are consistent with the results described by Kim and colleagues (2010).

Limitations and future research. Study 1 has several limitations, particularly in the study design. We set up the starting points in a box bracketing a 3^{o} glide slope. This introduced a confound between error valence and starting distance such that starting points that were most above glide path were also closest to the runway, and starting points that were most below glide path were always farthest. We felt that nearly any design we selected would have introduced some type of confound and our approach was the most readily interpretable. However, other configurations would certainly have been reasonable and justifiable. For example, we could have started participants on lines parallel to the glide path in order to maintain a consistent distance above or below the prescribed path. Starting participants on lines parallel to the glide path would keep distance from glide path the same, but introduces 9 levels of altitude or 9 levels of distance. In both cases the distance from the safe ($\pm 0.5^{o}$) cone changes with each starting point anyway. We also could have started pilots either on glide path or on the edges of the $\pm 0.5^{o}$ cone of safe approach. Starting them either on glide path or on the cone leads to 27 separate starting locations without shared values of altitude or distance (creating a 27-way ANOVA), and also alters the starting distance off of glide path depending on the distance from the runway. Future research examining these alternatives would be helpful in describing the relationship between glide slope performance and such factors as absolute error, relative error, etc. over the course of the approach. However, those were not the primary questions for this study.

The next limitation of our study is that we did not directly manipulate the presence of the horizon or ground-based contextual features. We therefore cannot directly describe the effects of ground-based contextual cues on the glideslope, nor can we disentangle the relative contribution of the horizon vs. contextual cues in pilots' ability to interpret the remaining visual cues. Our conclusions regarding the implications of our findings for hypotheses related to ground features should therefore be considered speculative.

A third limitation of this study is that multiple factors that are common in a real-life approach were absent from our simulation. For example, pilots were not briefed on runway length, and we restricted the instruments and landing aids that were available. This lack of information made the task more challenging and potentially introduced error that otherwise would not have occurred. Many factors beyond the visual stimuli contribute to an approach, and those were absent here. However, we felt that removing those additional factors would allow us to more effectively test the visual contributors to BHI.

Future research investigating the effect of these additional factors (particularly cognitive factors) would be potentially quite revealing. For example, one pilot during debriefing stated that it is safer to come in too high than too low, but a high approach requires a longer runway to land safely. This leads to an interesting possibility for future research to test one of our findings. We found that all else being equal, longer runways promote black hole effects and lead to increased error. However, our pilots did not know the length of the runway. Knowledge of a short runway (and thus knowledge that there is little available space) may prompt a pilot to fly lower to ensure adequate landing distance. Conversely, knowledge of a long runway may help mitigate BHI effects because pilots know they have more space with which to work. Thus, the relationship between BHI error and runway length may partly depend on cognitive factors. Indeed, Galanis, Jennings, and Beckett (1998) proposed that pilots may weight cues differently depending on pilot familiarity with the runway.

A final limitation of Study 1 is the relatively poor quality of the eye tracking data. Even in the videos that were deemed usable for analysis, the fixation point may have been off to the side of the runway. The lack of any quantitative analysis and the relatively poor data quality overall precludes any firm conclusions based on the eye tracking data. Our eye tracking results should be considered speculative at best until we are able to replicate our results using a more precise eye tracking system that will better support quantitative analysis. Future research should validate our findings using more precise measurements, and examine whether similar glances occur during daytime and night approaches to help clarify whether pilots attempt to use a similar visual strategy between daylight and BHI conditions.

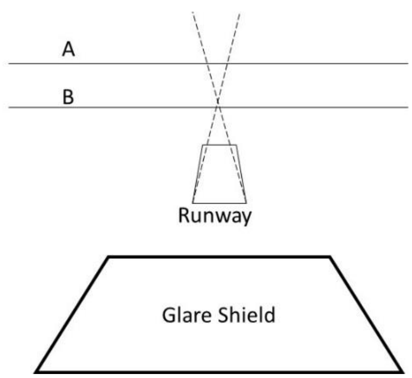
Having demonstrated some of the relevant visual cues that impact BHI, we now turn our attention to potential mechanisms by which these cues influence glide path. In particular, we explore the visual strategies that may allow pilots to fly a consistent (but low) glide path in the absence of visual contextual information. The following studies describe a series of computer-based tasks designed to investigate the perceptual mechanisms underlying BHI. Studies 2-4 used a different group of participants than Study 1. However, the same participants completed Studies 2-4, with the order of study presentation counterbalanced across participants. Therefore, although the studies described below are labeled 2-4, they were conducted in parallel using the same set of volunteers.

Study 2

Overview

The horizon is a critical cue for allowing visually-guided approaches. However, this cue is absent in BHI conditions. If pilots attempt to use the same strategies at night as during the day, we must account for how the horizon is used at night, even if it may not be visible. Even without the true horizon as a visual cue, a pilot may still attempt to use the horizon as a cue during a nighttime approach. To do so, the pilot infers an implicit horizon by estimating its location through available information. We hypothesize that in the absence of a visible horizon, pilots may be biased when estimating the implicit horizon based on the vanishing point from the parallel runway lines (Figure 11).

Figure 11. We hypothesize that in the absence of a true horizon (line A) the perceived intersection of the runway lines serves as an estimate of the horizon. To the extent that this estimate is biased (line B), the pilot may fly an inaccurate glide path due to the effect on perceived H angles.



Gibson (1966) proposed that invariance in the world, relationships among objects in the visual field, and the relative motion of such objects permit navigation in a 3D world despite a 2D retinal image. One optical invariant in the world is that parallel lines appear to converge at the horizon (i.e., the vanishing point). For this reason, the perceived splay of the lines serves as an altitude cue for the observer (Flach et al., 1997), and the horizon can be specified by the convergence point of parallel runway edge lines (e.g., Lintern & Liu, 1991).

An implicit horizon can be very powerful, both as its own cue and as a basis for determining H angles. H angle and an implicit horizon provide the least uncertainty during visual approaches, and are likely to be favored by pilots flying visual approaches (Galanis, Jennings, & Beckett, 1998). The presence of cues specifying an implicit horizon can affect flight path even if a horizon is visible. Lintern and Liu (1991) observed that artificially raising a visible horizon in the visual field during simulated approaches led to lower flight paths, whether by raising the horizon only, or by raising the entire visual scene. This effect was reduced in the presence of cues specifying a veridical implicit horizon. Left unexamined, however, was performance based on an implicit horizon and no visible horizon. To the extent that pilots can only rely on an implicit horizon, misperceptions may lead to bias in the flight path.

People are poor at judging perspective angles of converging lines (Erkelens 2015a; Erkelens 2015b). Further, the visual acuity of foveal vision subtends approximately 2º. This is sufficient to maintain the entire runway in foveal vision for most of the approach, but increases the uncertainty of the H angle because some components are in the periphery (Galanis, Jennings, & Beckett, 1998). Pilots tend not to directly reference the horizon, but may use it via peripheral vision (Kim, Palmisano, Ash, & Allison, 2010). The peripheral nature of the horizon may increase the likelihood that an implicit horizon based on runway cues will be inaccurately perceived. If the proposed line bias illusion is present, we hypothesize that the estimated horizon will fall short of the actual horizon. If pilots use their initial (false) estimate of the horizon to set their H angle, the glide path may change based on changes in the estimate of the implicit horizon over time. Specifically, we hypothesize that estimates of the horizon will be most biased at greater distances from the runway when the edge lines are smaller, their relative angle is steeper, and the gap between the end of the lines and the true horizon is largest. This bias will likely be reduced as the pilot approaches the runway and cues based on runway shape become less uncertain (Galanis, Jennings, & Beckett, 1998). As the estimate becomes more accurate throughout the approach, absolute H angle based on the implicit horizon will appear to increase (Figure 12a). Pilots may correct this apparent change by lowering the aircraft (Lintern & Liu, 1991; Figure 12b).

Figure 12a. As pilots approach the runway, biased estimates of the implicit horizon (dotted lines) relative to the true horizon (solid line – not actually visible in a BHI scenario) may change as the runway changes in the visual field. If the pilot uses the initial implicit horizon at the far left of the figure to set their H angle, the absolute H angle will appear to increase as they get closer to touchdown. Pilots may fly lower in an attempt to keep this cue stable during the approach.

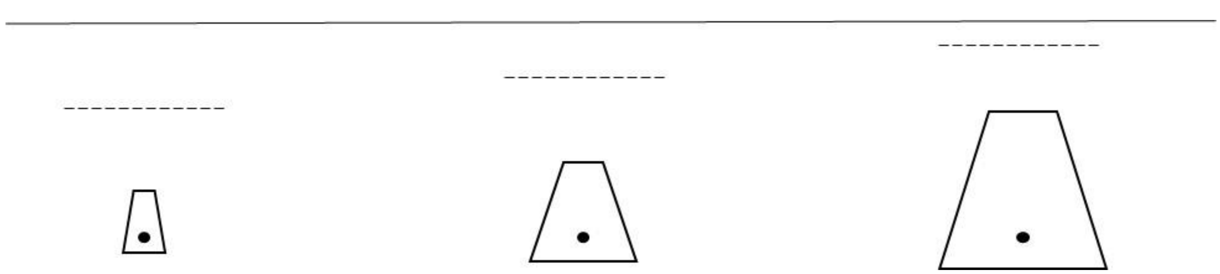
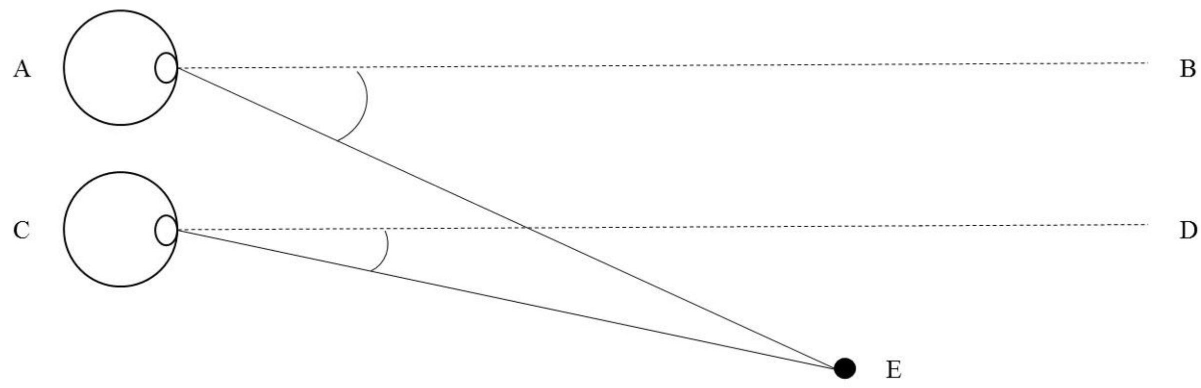


Figure 12b. The absolute H angle corresponds to the visual angle between the horizon and aim point. The true horizon appears at eye level in the visual field (Lintern & Liu, 1991). Absolute H angle therefore remains constant when flying a stable approach path but varies if pilots begin to deviate from the glide path. In order to compensate for an apparently large and/or increasing absolute H angle (angle BAE), pilots may fly lower to reduce it (angle DCE).



In addition to the loss of the horizon during BHI conditions, the nighttime configuration of the runway edge lines may affect how the edge lines are perceived compared to a daytime approach. At night, the edges of the runway are identified via lights spaced at regular intervals rather than a solid line. This produces a dotted line effect, which limits the data points available for visual sampling compared to the solid line that would be visible during the day. Further, visual sampling of runway lights is uneven due to the angle and altitude of the glide path (i.e., lights further away appear to be closer together). Uneven sampling, coupled with the lack of a horizon to serve as an indication of the vanishing point, may produce biases in the perceived angle or intersection point of the runway edge lines.

We tested the possibility that pilots may underestimate the location of the horizon in the visual field using a computerized task to evaluate how well people judge the intersection point of two dotted line segments. We hypothesized that participants would underestimate the intersection point (i.e., place it below the true intersection point on the screen), and that this effect would vary with the angle at which the lines were presented. We further hypothesized that the effect would be exacerbated when the dots were unevenly spaced (as they would appear to be when looking at runway edge lights receding into the distance while on approach to landing).

Method

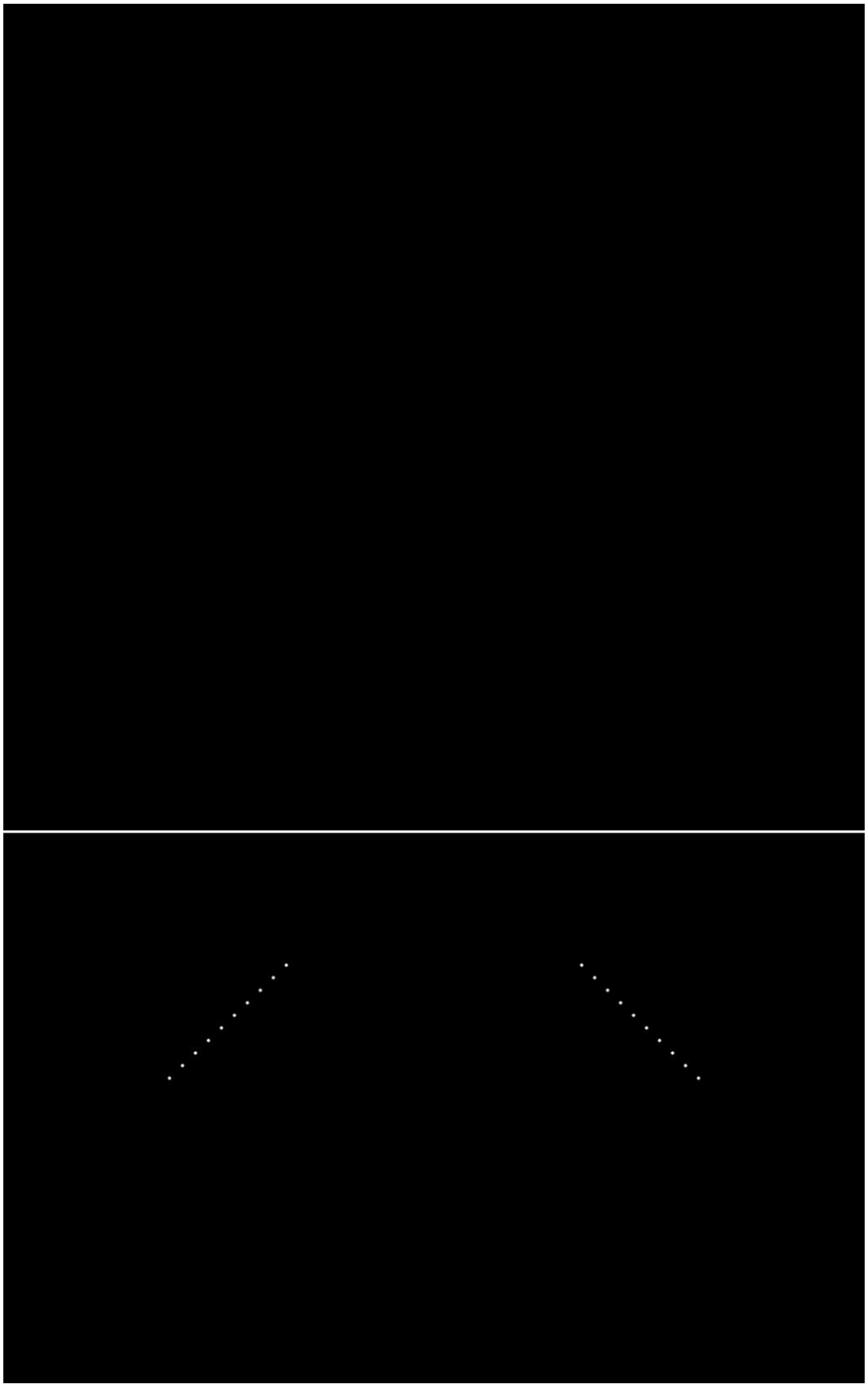
Participants. A total of 36 volunteers (18 males and 18 females ranging in age from 19 to 44) participated in this study. Participants were military and civilian personnel from on and around Wright-Patterson AFB. Participants reported normal or corrected to normal vision and the ability to use a standard computer keyboard easily. Two participants reported having flight experience (1,711.9 and 1,350 flight hours). Three participants reported being left handed, and 10 reported being left eye dominant (one of whom was also left handed).

Apparatus. Participants completed a computerized task instantiated on a Dell® Precision T1700 desktop computer with an Intel® XEON processor running Windows 7®. A 56 cm diagonal Hyundai LCD monitor with 1050 x 1680 resolution was used to display the task. The monitor was turned such that it was in portrait orientation. Participants used an Amazon Basics® keyboard and mouse for response inputs.

Task. Participants completed a custom task developed for this study using the Godot game development engine. The participants were instructed to estimate the point at which two dotted line segments intersected one another on the screen. Participants were shown two dotted line segments each consisting of 10 white dots against a black background, along with a white horizontal solid line extending the width of the monitor (Figure 13). The two dotted lines were presented at the same angle relative to the vertical, but mirrored. The segments were angled towards one another, but did not physically intersect. The participants' task was to use the up and down arrow keys on the keyboard to move the solid horizontal line to the point at which they estimated the dotted line segments would intersect were they to extend far enough.

The outcome measure for Study 2 was error, calculated as the number of pixels between the true intersection point of the stimulus lines and where the participant placed the horizontal line. Negative values indicated that the participant placed the response line too low on the screen, whereas positive values indicated that the participant placed the response line too high. The starting location of the solid horizontal line varied randomly in order to avoid any effects of distance travelled or direction of approach relative to the intersection point. When the participant was satisfied with their response, they pressed the space bar to move to the next trial. Participants did not receive any feedback regarding their accuracy.

Figure 13. Screenshot of the task for Study 2. Participants used the up and down arrow keys to move the solid line to the point at which the two dotted line segments intersected.



Design. This study followed a 4 (line angle) x 2 (dot spacing) within subjects design. The levels of each variable were as follows:

Line angle. 10º off vertical, 20º off vertical, 30º off vertical, and 40º off vertical, rotated about the uppermost dot of the line segment.

Dot spacing. Evenly spaced or unevenly spaced. Dotted line segments with unevenly spaced dots were the same overall length as the evenly spaced dotted line segments, and contained the same number of dots. The unevenly spaced lines were always oriented such that the closer-together dots were towards the top of the monitor as they would appear on a nighttime landing approach. Spacing was determined by a multiplicative algorithm, the code for which can be found in Appendix 1. This algorithm created the desired basic effect, but did not replicate the exact spacing that would be seen during an approach.

Trials were blocked based on dot spacing, with the angle of the dotted line segments varying within blocks (Table 7). Each line angle was presented 15 times, leading to two blocks of 60 trials each, plus five additional practice trials presented at the beginning of each block. Practice trials used different line angles than the angles listed above. Trials were presented in random order within each block. The order of blocks was alternated between participants.

Evenly spaced dots	**Unevenly spaced dots**
Block 1: 4 angles; 15 trials per angle (60 total trials)	Block 2: 4 angles; 15 trials per angle (60 total trials)

Table 7. Blocking scheme for Study 2

Experimental procedure. Participants completed a separate consent form for each task described in Studies 2-4. After consent, participants were seated in front of the computer with the center of the screen at approximately eye level from a viewing distance of 75cm. The testing room was illuminated to a level of 750 lux using fluorescent lighting of the type commonly found in office environments. Participants were instructed on the task, and told that they should be both fast and accurate in their estimates of the intersection point. The participants then completed the blocks of the task. Each block began with five practice trials using line angles different than those listed above. Participants were offered breaks between blocks. If the participant had no more tasks to complete after finishing all the blocks within a task, the participant was thanked, debriefed, and dismissed. If more tasks remained, the participant was offered a break before beginning the next task.

Analyses and Results

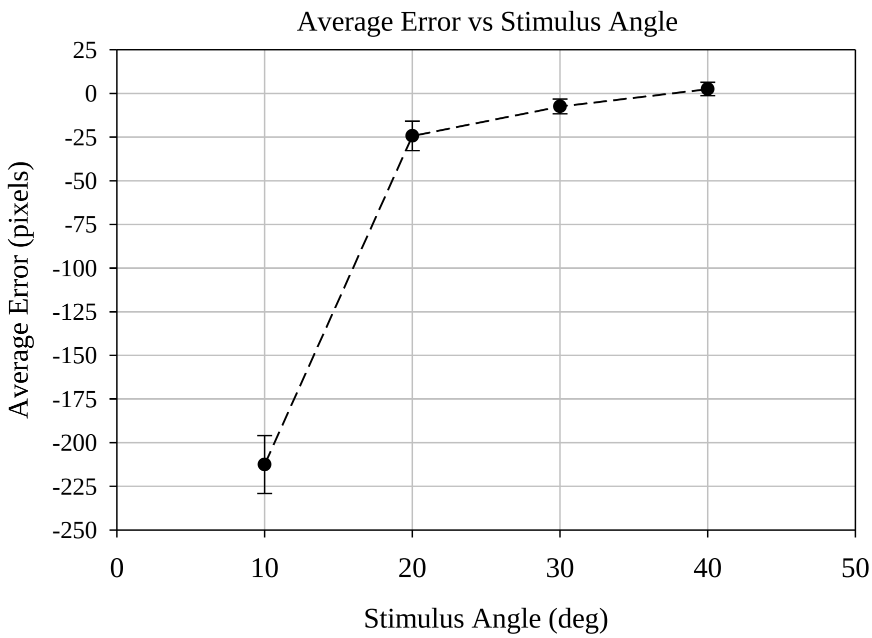
Study 2 was analyzed using a 2 (dot spacing) x 4 (line angle) repeated measures ANOVA. Results indicated a main effect of line angle, but no effect of dot spacing and no significant interactions (Table 8).

Variable	SS	df	MS	*F*	*p*	ηp squared
Intercept	1022987	1	1022987	152.31	< 0.001	0.36
Line Angle*	2185907	3	728636	108.48	< 0.001	0.54
Dot Spacing	1425	1	1425	0.21	0.65	< 0.01
Angle x Spacing	329	3	110	0.02	1.00	< 0.01
Error	1826886	272	6716			

*Table 8. ANOVA results for Study 2. Variables with a * are significant at p < 0.05.*

Post hoc testing conducted using individual t-tests (Bonferroni corrected $\alpha = .0125$) for each line angle revealed that estimation error for stimulus line angles of 10º and 20º were each significantly different from zero (t (69) = -12.83, p < 0.001; t (69) = -2.88, p = 0.01, respectively). Participants placed the “horizon” too low, particularly at smaller angles off the vertical (Figure 14).

Figure 14. Mean error for each stimulus line angle. Line angle is on the X axis. Error (in pixels) is on the Y axis. Negative values indicate that the response was lower than the true intersection point. Error bars represent standard error.



Discussion

This study demonstrated that people tend to underestimate the convergence point of two dotted line segments, particularly at steep angles. Participants placed the convergence point extremely low in the 10º line condition, and to a lesser extent in the 20º line condition. This finding implies that pilots may have similar difficulty estimating where the runway edge lines converge at the horizon. To the extent that pilots imagine the horizon to be lower than it actually is, they may set their H angle to be too small. Such a line bias illusion would lead to reduced altitude and a BHI approach if pilots attempt to maintain this shortened H angle for the duration of the approach.

Though we believe the results of Study 2 support the possibility that pilots may have difficulty estimating the location of the horizon, our stimuli differed from the visual stimuli experienced during an approach in several key ways that prevent us from making a stronger assertion. First, the line segments seen by our participants were much larger than those seen by pilots during approach, particularly compared to what pilots would see at greater distances from the runway. However, it seems reasonable to expect that larger lines would be easier to estimate. We therefore believe that this limitation does not bias us towards finding effects that would not be observed in a real approach. Second, the spacing between dots was greater than pilots would see on a lighted runway edge line. Pilots would see a nearly solid line when far from the runway. The lack of an effect of dot spacing on participants' responses somewhat eases our concern about the implications of this limitation for actual approaches, but we cannot rule out the possibility that tighter dot spacing would lead to different effects. Third, the line segments remained a constant length across stimulus angles, but the runway edge lines simultaneously appear to change length and angle as the aircraft descends during a real approach. However, the runway edge lines appear shortest at the greatest distances from the runway, corresponding to smaller angles. Smaller angles were associated with the greatest error in our study. Therefore, actual runway lines would likely be more difficult to judge than the stimulus lines in our study, not easier (see our point about line lengths immediately above).

We do note one confound in our study design that is worth more detailed discussion. The convergence point of two line segments is a product of both the line angle and the distance between the end points of the lines. Smaller stimulus line angles led to a convergence point further from the end of the line segments, introducing a confound between stimulus angle and the distance over which the estimate must be made. Adjusting for both line angle and distance between the lines would have either introduced additional variance in the data or lead to new confounds (e.g., we would have had to change line segment length or the distance between the line segments across trials). Rotating about the topmost point in the line segment maintained a constant width between the lines and ensured that any changes in convergence point were driven solely by differences in angle, even if the distance confound could not be completely eliminated.

However, this confound is not entirely unrealistic as the gap between the horizon and the far runway threshold is greatest at smaller angles of the runway edge lines and least at larger angles of the runway edge lines due to the expansion of the runway in the visual field as the aircraft's distance from the runway decreases. This optical expansion would also increase the apparent width of the runway, and future research should incorporate such visual changes to see how they impact the ability to estimate the convergence point.

Overall, the results of Study 2 support the plausibility of the hypothesized line bias illusion (although our hypothesis about dot spacing was not supported). Participants placed the estimated convergence point of two stimulus lines too low relative to the actual convergence point. Such a tendency would correspond to underestimating the location of the horizon at the convergence point of runway edge lines, potentially causing pilots to fly lower to compensate for misperceived absolute H angles during the approach. Studies 3 and 4 investigate this illusion further by exploring possible causal mechanisms.

Study 3

Overview

Study 2 established that people have difficulty estimating the convergence points of two lines, supporting the plausibility of the hypothesis that pilots may underestimate the location of the horizon if they are forced to estimate its location based on the vanishing point of the runway edge lines. Study 3 investigated possible causal mechanisms behind these findings, grounded in a set of known tendencies for people to misinterpret the slope of lines as well as their projections into space (e.g., the Poggendorff illusion). In particular, error in reading the position of a point on a sloping line follows two general rules: 1) the direction of the error is in the same direction as the slope of the line, and 2) error increases at farther distances between the point and the calibrated edge against which it is read (Poulton, 1985). These tendencies can cause people to misinterpret graphs, and is most common when reference cues are lacking. The effect can be counteracted by providing reference points nearer to the point at which the estimate is made (Poulton, 1985). Consistent with the notion that the lack of contextual cues affects perceptions of runway size and shape (see Study 1), we believe the lack of contextual cues under BHI conditions may make it difficult to judge the relative angle of the runway edge lines, leading to biased estimates of the horizon and perhaps making it more difficult to utilize runway form ratio as a cue. Specifically, we hypothesize that people will "flatten" sloping lines toward the horizontal when forced to extend them into space. Such an effect would lower an implicit horizon estimated based on the intersection point of two lines.

We hypothesize that in the absence of contextual cues to serve as reference points (such as the horizon, terrain, or buildings), pilots may have difficulty judging the relative angles of the runway edge lines such that the lines would appear "flatter" (i.e., more horizontal). Lines that are judged to be more horizontal would be perceived to have a lower convergence point, thereby lowering the estimated location of the horizon in the visual field. We tested this possibility using a simple computer task in which participants judged the apparent point of intersection between a single line segment presented at various angles and a stable response line. As in Study 2, we investigated the possible contribution of perceived dot spacing. In addition, we explored whether any effect would be due to the apparent compression of lights at the far end of the runway or the apparent wider spacing of lights at the near end of the runway by having participants respond at both the top and bottom of the computer screen.

Method

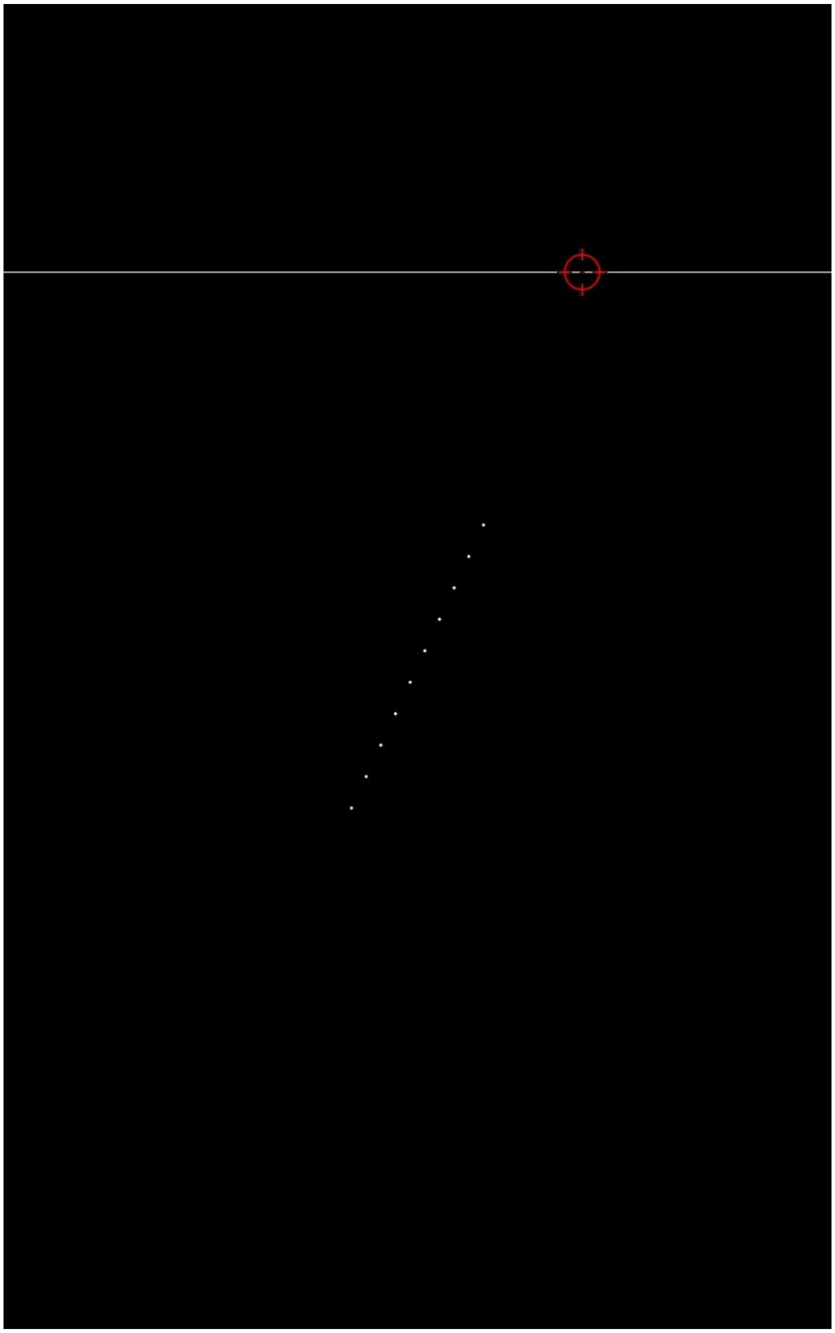
Participants. The same participants used in Study 2 contributed data for Study 3.

Apparatus. The apparatus for Study 3 was the same as for Study 2.

Task. Participants completed a custom task developed using the Godot game development engine. The participants were instructed to estimate the point at which a dotted line intersected a solid line on the screen. Participants were shown a dotted line segment consisting of 10 white dots in the center of a black background, along with a white horizontal solid line extending the width of the monitor (Figure 15). The two lines did not intersect one another. The participants' task was to use the left and right arrow keys on the keyboard to move a red cursor along the solid line to the point at which they estimated the dotted line segment would intersect the horizontal solid line if the line segment extended far enough. When the participant was satisfied with their response, they pressed the space bar to move to the next trial. The starting location of the cursor varied randomly in order to avoid any effects of distance travelled or direction of approach relative to the intersection point. Participants did not receive any feedback regarding their accuracy.

We analyzed error in the participants' responses in degrees, derived by calculating the angle of the line that would have been formed by the point of the participants' response and the center of the stimulus line and comparing this value to the true angle of the stimulus line. Positive values represent overestimations indicated by a response too close to the edge of the screen (i.e., the angle of the line indicated by the participant would be farther off the vertical than the stimulus line). Negative values represent underestimations indicated by a response too close to the center of the screen (i.e., the angle of the line indicated by the participant would be closer to vertical than the stimulus line).

Figure 15. Screenshot of the task for Study 3. This figure portrays the evenly spaced condition, with the response line at the top of the screen. Participants used the left and right arrow keys to move the cursor to the point at which they estimated the dotted stimulus line would intersect the solid response line.



Design. This study followed a 5 (line angle) x 2 (dot spacing) x 2 (response location) x 2 (direction of line slant) within subjects design. The levels of each variable were as follows:

Line angle. 0º (vertical), 10º off vertical, 20º off vertical, 30º off vertical, and 40º off vertical, rotated about the center of the line segment. The 0º condition served as a check to allow us to identify any systematic bias in participants' ability to judge the intersection point.

Dot spacing. Evenly spaced or unevenly spaced. Dotted line segments with unevenly spaced dots were the same overall length as the evenly spaced dotted line segments, and contained the same number of dots. The unevenly spaced lines were always oriented such that the closer-together dots were towards the top of the monitor. Spacing was determined using the same algorithm described for Study 2.

Response line location. Top of the screen or bottom of the screen. This manipulation allowed us to better determine whether any possible effect of dot spacing was driven by the closer spacing of the dots at the top of the screen or the wider spacing of the dots at the bottom of the screen.

Direction of line slant. Left or right slant. Because runway edge lines consist of two parallel lines but our task only used one line, we used this manipulation to ensure that both sides of the runway would be equally represented in our task.

Trials were blocked based on response location and dot spacing, with the angle of the dotted line and direction of the dotted line slant varying within blocks (Table 9). Each combination of line angle and slant was presented 10 times, leading to four blocks of 90 trials each (the 0º line only had one slant), plus five additional practice trials presented at the beginning of each block. Practice trials used different line angles than the angles listed above. Trials were presented in random order within each block. The order of blocks was counterbalanced using a balanced Latin Square design.

	Evenly spaced dots	**Unevenly spaced dots**
Respond at the top	Block 1: 5 angles; Slanted left or right (90 total trials)	Block 2: 5 angles; Slanted left or right (90 total trials)
Respond at the bottom	Block 3: 5 angles; Slanted left or right (90 total trials)	Block 4: 5 angles; Slanted left or right (90 total trials)

Table 9. Blocking scheme for Study 3

Experimental procedure. Study 3 followed the same procedure as Study 2.

Analyses and Results

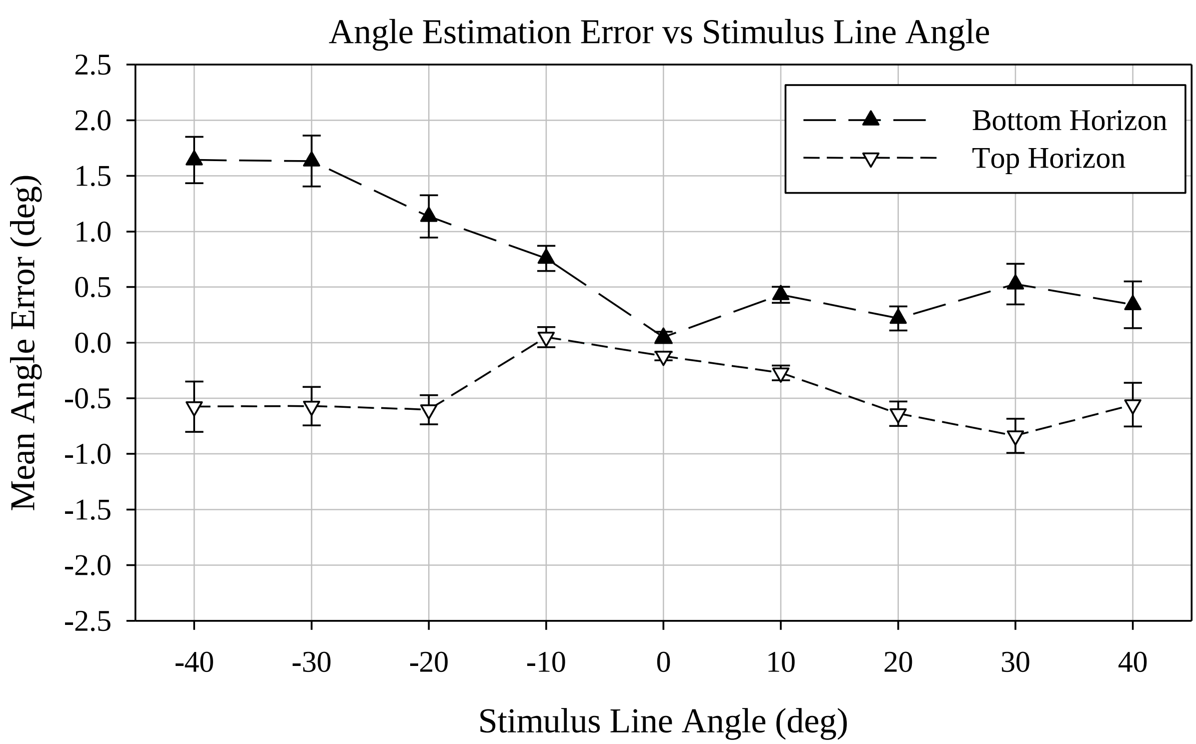
The data for Study 3 were analyzed using a 9 (stimulus line angle) x 2 (dot spacing) x 2 (response location) repeated measures ANOVA. Results indicated significant interactions between stimulus line angle and response location as well as between dot spacing and response location (Table 10).

Variable	SS	df	MS	F	p	η_p squared
Intercept	27.06	1	27.06	15.80	< 0.001	0.01
Line Angle*	100.85	8	12.61	7.36	< 0.001	0.04
Dot Spacing	0.85	1	0.85	0.50	0.48	< 0.01
Response Location*	471.70	1	471.69	275.42	< 0.001	0.18
Angle x Spacing	5.30	8	0.66	0.39	0.93	< 0.01
Angle x Location*	148.64	8	18.58	10.85	< 0.001	0.06
Spacing x Location*	7.63	1	7.63	4.46	0.03	< 0.01
Angle x Spac x Loc	5.41	8	0.68	0.39	0.92	< 0.01
Error	2171.61	1268	1.71			

*Table 10. ANOVA results for Study 3. Variables with a * are significant at $p < 0.05$.*

An examination of the plots for the interactions between line angle/response location and dot spacing/response location reveals the nature of these effects. Figure 16 shows the interaction between line angle and response location. The figure indicates that participants tended to place the intersection between the stimulus line and the response line too close to the edge of the screen when responding at the bottom of the screen and that this error tended to increase with greater angles off the vertical, particularly for lines slanted to the left. Conversely, participants tended to place the intersection point too close to the center of the screen when responding at the top of the screen but the error was relatively constant across line angles.

Figure 16. Interaction between stimulus line angle (X axis) and response location for Study 3. The error (in degrees) of the participants' responses is on the Y axis.



Post-hoc tests using Tukey's HSD confirmed these impressions. When responding at the bottom of the screen, responses at the -40º and -30º line angles were significantly worse than responses at -10º and vertical. Responses for the -20º line were not significantly different from the -40º, -30º, or -10º lines, but were significantly worse than for the vertical lines. Responses for positively valued stimulus angles at the top bottom of the screen did not show any significant differences. Likewise, responses at the top of the screen did not show any significant differences across stimulus line angle within the left or right slanted lines. See Tables 11 and 12 for the full set of *p* values for the post-hoc tests for responses at the top and bottom of the screen, respectively.

Angle	-40	-30	-20	-10	0	10	20	30	40
-40									
-30	1.00								
-20	1.00	1.00							
-10	0.27	0.29	0.20						
0	0.83	0.84	0.75	1.00					
10	1.00	1.00	0.99	0.99	1.00				
20	1.00	1.00	1.00	0.14	0.63	0.97			
30	1.00	1.00	1.00	**0.01**	0.09	0.46	1.00		
40	1.00	1.00	1.00	0.32	0.87	1.00	1.00	1.00	

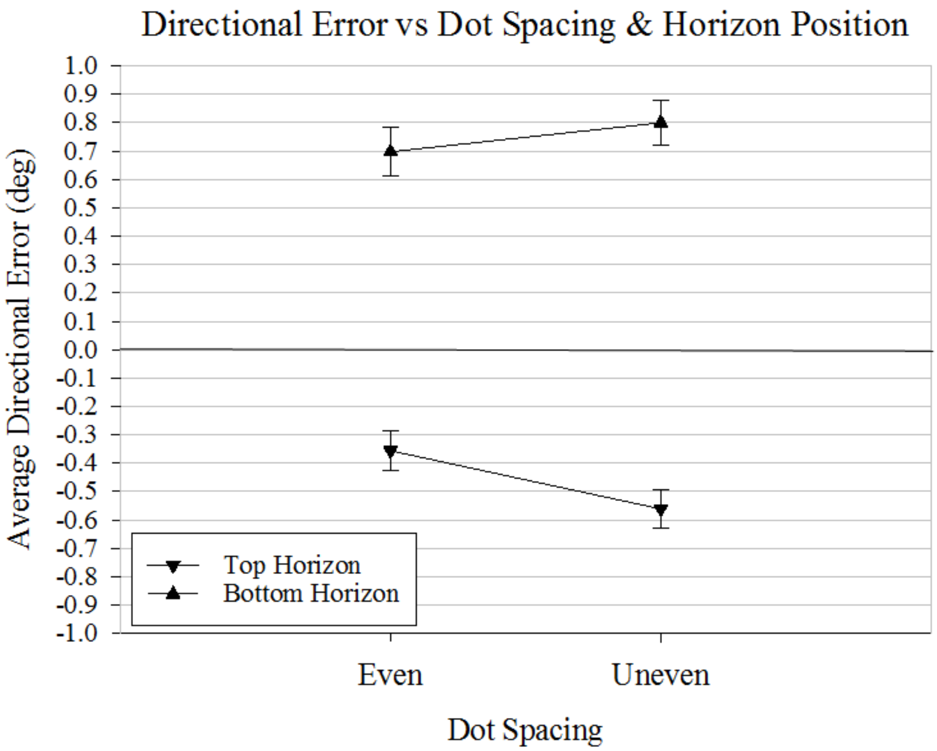
Table 11. Tukey's HSD results (p values) for the responses at the top of the computer screen at different stimulus line angles for Study 3. For all tests, MS = 1.71 and df = 1268.

Angle	-40	-30	-20	-10	0	10	20	30	40
-40									
-30	1.00								
-20	0.67	0.70							
-10	**0.01**	**0.01**	0.96						
0	**< 0.001**	**< 0.001**	**< 0.001**	0.10					
10	**< 0.001**	**< 0.001**	0.11	0.99	0.96				
20	**< 0.001**	**< 0.001**	**< 0.01**	0.55	1.00	1.00			
30	**< 0.001**	**< 0.001**	0.32	1.00	0.76	1.00	1.00		
40	**< 0.001**	**< 0.001**	**0.03**	0.91	1.00	1.00	1.00	1.00	

Table 12. Tukey's HSD results (p values) for the responses at the bottom of the computer screen at different stimulus line angles for Study 3. For all tests, MS = 1.71 and df = 1268.

Figure 17 shows the interaction between dot spacing and response location. Examination of the figure reveals that participants tended to place the cursor too close to the edge of the screen when responding at the bottom of the screen, but placed it too close to the center of the screen when responding at the top of the screen. This effect was exacerbated when the dots were unevenly spaced.

Figure 17. Interaction between dot spacing and response location. The Y axis displays error in degrees. Negative values indicate that participants placed their response too close to the center of the screen, whereas positive values indicate that participants placed their response too close to the edge of the computer screen.



Discussion

Study 3 offered evidence that participants had difficulty judging the angles of line segments, and hinted at the possibility that participants may have perceived the lines as more horizontal in some cases, consistent with Poggendorff-type illusions. Participants tended to indicate that dotted line segments would intersect the target line closer to the edge of the computer screen than the true intersection point when responding at the bottom of the screen, particularly when the line was slanted to the left. Under these conditions, the net effect of these tendencies was such that the participants' estimations of the dotted line segments' paths appeared more horizontal than the actual line segments. Such an effect is consistent with the results of

Study 2, in which participants placed the estimated horizon too low in the visual field. If participants judge the runway edge lines to be more horizontal than they actually are, the convergence angle of a pair of runway edge lines would increase, leading to a lower estimate of the implicit horizon. As described in Study 2, this underestimate may lead to a miscalibrated H angle and a low approach as pilots attempt to keep this cue constant.

However, this effect was not displayed consistently across all experimental conditions. In particular, responses above the line segments tended to place the intersection point too close to the center of the screen and responses above and below right-slanted lines tended to cancel each other out such that the net effect was a lateral shift in the line rather than a change in apparent slant. We do not have a ready explanation for the asymmetry in the left- vs. right-slanted lines, but offer hypotheses regarding the other effects below.

Participants' responses were far more accurate when responding at the top of the screen compared to the bottom. This result is puzzling considering the lack of dot spacing effects (i.e., the difference in spacing between dots near the top of the screen compared to dots near the bottom of the screen did not appear to influence results). One possible explanation is that people are simply better at judging lines in the world when they are "up" than when they are "down." People tend to favor the tops of objects when interacting with them. For example, when turning a dial clockwise, people tend to describe it as turning right (the direction the top is traveling relative to the user) rather than left (the direction the bottom is traveling). This cognitive bias towards the top of objects may result in a higher skill level when making judgments about items near the top of the visual scene.

The error in participants' estimates tended to increase at greater line angles, particularly for line segments slanted to the left. Given participants' accuracy in judging the intersection point of the vertical stimulus, people seem to be reasonably skilled at judging lines without tilt. It could be the case that people can easily make binary vertical-not vertical judgments, but have a harder time making more subtle distinctions between angles. This difficulty may increase at greater angles, allowing bias to influence the participants' judgments. Alternatively, the vertical line intersection corresponded to the center of the response line. Participants may have simply been good at dividing the response line in half, but did not know what proportion of the line to divide in the other conditions. However, participants were not told that the line segment was completely vertical, so this explanation still requires at least some judgment of verticality from the participant.

An alternative explanation for the increase in error at greater line angles is a possible confound between stimulus angle and the distance over which the estimate occurred. The stimulus line segment was rotated about the center and was set at a fixed length, meaning that the distance between the end of the line segment and the intersection point on the response line increased at greater line angles. Participants' error may have increased at greater angles simply because they had to estimate over a greater distance. This possibility is examined in Study 4. We note that the tendency for error to increase at greater line angles is inconsistent with the results of Study 2 (in which error was worse for smaller line angles). This apparent discrepancy is addressed in greater detail at the end of this report.

Study 4

Overview

Study 3 established that participants had difficulty judging the intersection point of angled line segments, and in some conditions may overestimate the apparent slant of such segments. However, the horizontal response line used in Study 3 introduced a confound such that the distance between the end of the line segment and intersection point increased at greater line angles. Therefore, participants in Study 3 may have been less accurate at these angles not because their perception was altered, but because they could not judge the intersection as accurately over the longer distances. Study 4 tested that possibility by replacing the horizontal response line used in Study 3 with an arc that was equidistant from the end point of the dotted line segment regardless of the stimulus angle.

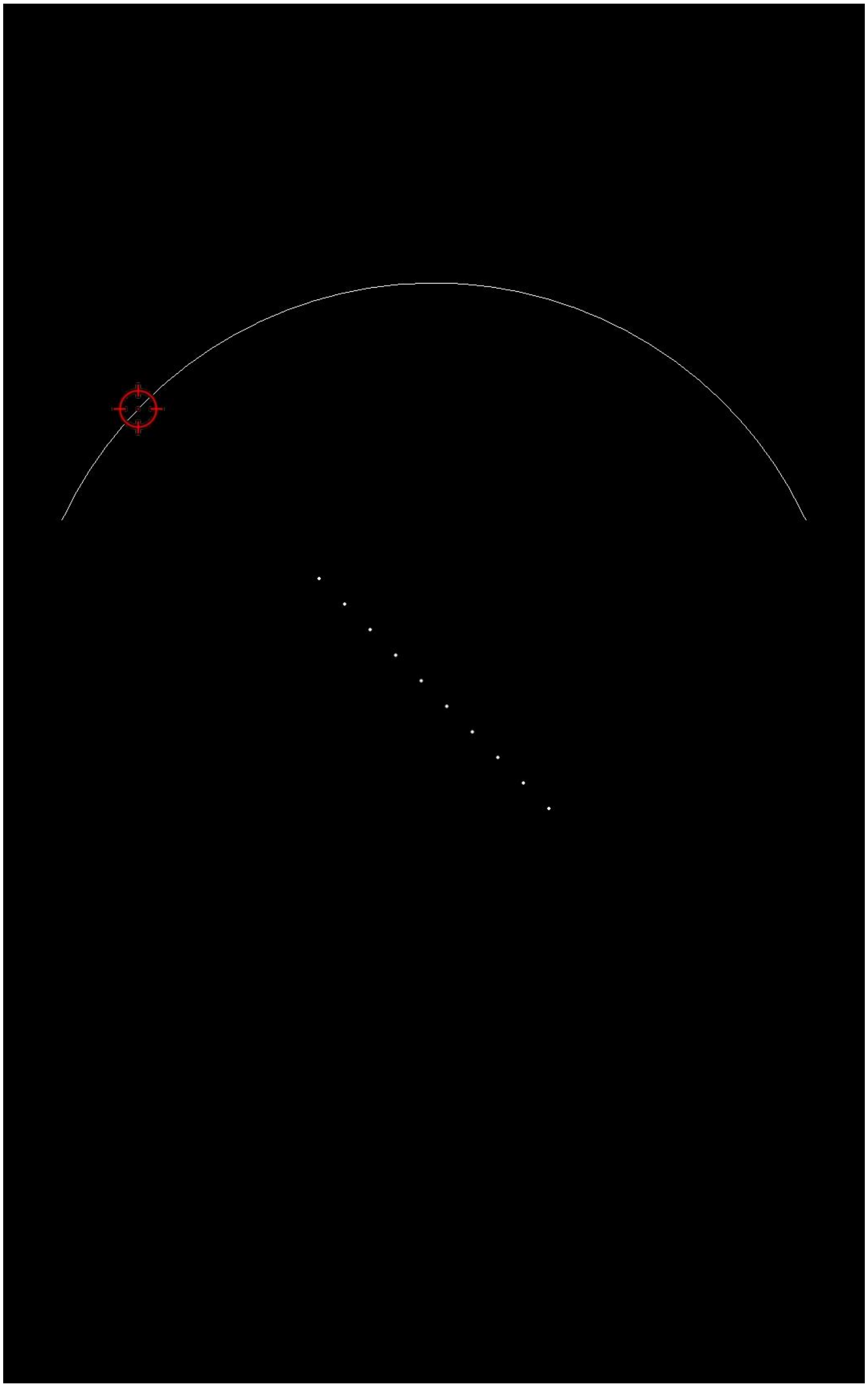
Method

Participants. The participants for Study 4 were the same as for Studies 2 and 3.

Apparatus. The apparatus for Study 4 was the same as for Studies 2 and 3.

Task. Participants completed a custom task developed using the Godot game development engine. The participants were instructed to estimate the point at which a dotted line intersected a solid arc on the screen. Participants were shown a dotted line segment consisting of 10 white dots in the center of a black background, along with a solid white arc. The dotted line segment and the arc did not intersect one another. The participants' task was to use the left and right arrow keys on the keyboard to move a red cursor along the arc to the point at which they estimated the dotted line segment would intersect the arc if the line segment extended far enough (Figure 18). When the participant was satisfied with their response, they pressed the space bar to move to the next trial. The starting location of the cursor varied randomly in order to avoid any effects of distance travelled or direction of approach relative to the intersection point. Participants did not receive any feedback regarding their accuracy. Our dependent measure was the difference in the angle of the line formed by the participant's response and the true angle of the stimulus line, calculated in the same manner as for Study 3.

Figure 18. *Screenshot of the task for Study 4. This figure portrays the evenly spaced condition, with the response arc at the top of the screen. Participants used the left and right arrow keys to move the cursor to the point at which they estimated the dotted stimulus line would intersect the solid response line.*



Design. Study 4 followed the same design as Study 3.

Experimental procedure. Study 4 followed the same procedure as Studies 2 and 3.

Analyses and Results

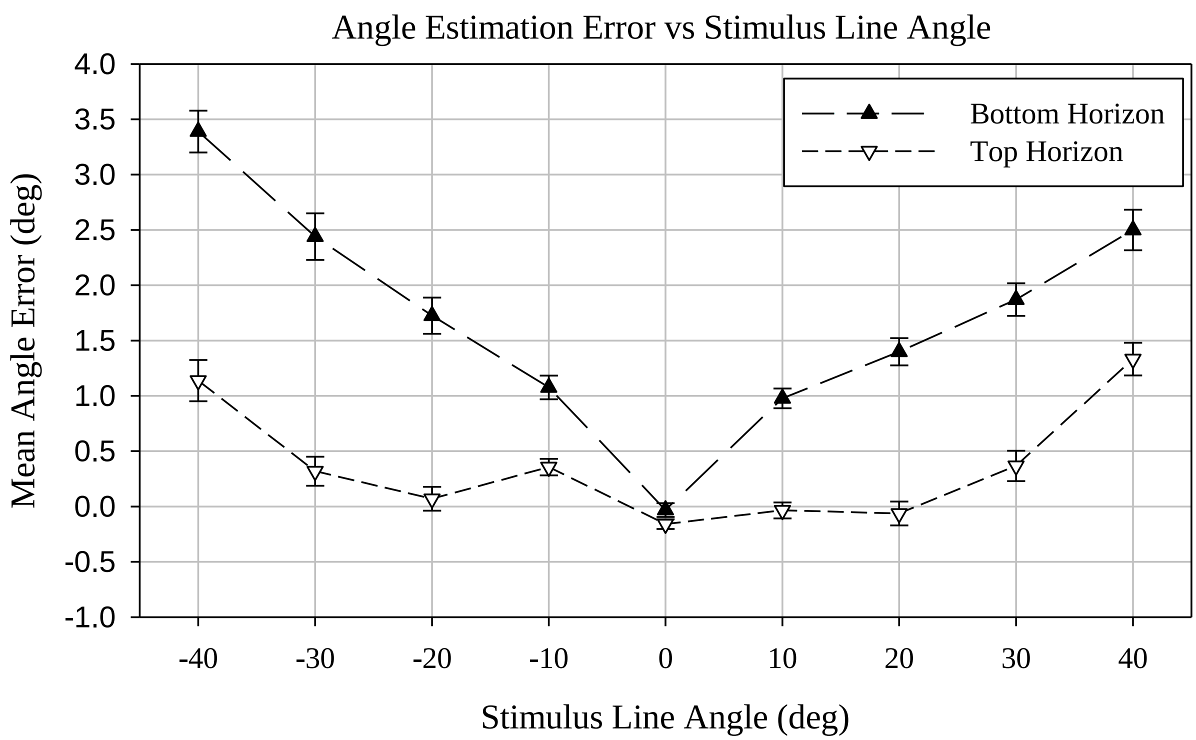
As before, the data for Study 4 were analyzed using a 9 (stimulus line angle) x 2 (dot spacing) x 2 (response location) repeated measures ANOVA. Results indicated a significant interaction between stimulus line angle and response location (Table 13).

Variable	SS	df	MS	F	p	η_p squared
Intercept	1355.28	1	1355.28	1054.51	< 0.001	0.46
Line Angle*	596.81	8	74.60	58.05	< 0.001	0.27
Dot Spacing	0.05	1	0.05	0.04	0.84	< 0.01
Response Location*	561.68	1	561.68	437.03	< 0.001	0.26
Angle x Spacing	2.40	8	0.30	0.23	0.98	< 0.01
Angle x Location*	125.06	8	15.63	12.16	< 0.001	0.07
Spacing x Location	2.02	1	2.02	1.57	0.21	< 0.01
Angle x Spac x Loc	1.13	8	0.14	0.11	1.00	< 0.01
Error	1583.40	1232	1.29			

*Table 13. ANOVA results for Study 4. Variables with a * are significant at $p < 0.05$.*

An examination of the plot for the interaction between line angle and response location offers clarity regarding the nature of this effect (Figure 19). The figure indicates that participants tended to place the intersection between the stimulus line and the response line too close to the edge of the screen, particularly at greater stimulus line angles. This tendency was more pronounced when responding at the bottom of the screen.

Figure 19. Interaction between stimulus line angle (X axis) and response location for Study 4. The error (in degrees) of the participants' responses is on the Y axis.



Post-hoc tests using Tukey's HSD confirmed these impressions. Table 14 shows the *p* values for each comparison when responding at the top of the screen. When responding at the top of the screen, errors were greater at ±40º than at all other angles. The other differences were not significant.

Angle	-40	-30	-20	-10	0	10	20	30	40
-40									
-30	**<0.01**								
-20	< **0.001**	1.00							
-10	**0.01**	1.00	0.99						
0	< **0.001**	0.53	1.00	0.39					
10	< **0.001**	0.93	1.00	0.85	1.00				
20	< **0.001**	0.87	1.00	0.76	1.00	1.00			
30	**0.01**	1.00	0.99	1.00	0.35	0.81	0.72		
40	1.00	< **0.001**	< **0.001**	< **0.001**	< **0.001**	< **0.001**	< **0.001**	< **0.001**	

Table 14. Tukey's HSD results (p values) for the responses at the top of the computer screen at different stimulus line angles for Study 4. For all tests, MS = 1.29 and df = 1232.

Table 15 shows the *p* values for each comparison when responding at the bottom of the screen. When responding at the bottom of the screen, error generally increased as the stimulus line moved further from the vertical, though this trend was slightly more pronounced for lines slanted to the left than to the right. Errors increased symmetrically across both directions of slant (i.e., error at positive and negative line angles was not statistically different for a given absolute value of angle), with the exception of ±40º.

Angle	-40	-30	-20	-10	0	10	20	30	40
-40									
-30	< **0.001**								
-20	< **0.001**	**0.02**							
-10	< **0.001**	< **0.001**	0.07						
0	< **0.001**	< **0.001**	< **0.001**	< **0.001**					
10	< **0.001**	< **0.001**	**0.01**	1.00	< **0.001**				
20	< **0.001**	< **0.001**	0.97	0.97	< **0.001**	0.76			
30	< **0.001**	0.23	1.00	< **0.01**	< **0.001**	< **0.001**	0.56		
40	< **0.01**	1.00	**0.01**	< **0.001**	< **0.001**	< **0.001**	< **0.001**	0.09	

Table 15. Tukey's HSD results (p values) for the responses at the bottom of the computer screen at different stimulus line angles for Study 4. For all tests, MS = 1.29 and df = 1232.

Discussion

Study 4 eliminated a confound in Study 3 and provided stronger evidence in favor of the hypothesis that people may perceive angled line segments as more horizontal than they actually are. In contrast to the asymmetrical results of Study 3, the results of Study 4 indicate that participants tended to overestimate the stimulus line angle regardless of whether the stimulus line was slanted to the left or to the right. Further, the results indicate increasing error with increasing line angles at both the top and bottom response locations. This pattern of results indicates that participants placed the perceived intersection point too close to the edge of the computer monitor at both the top and bottom response locations, with the net effect of flattening the stimulus line segment across all stimulus conditions.

The results of Study 4 are again consistent with the results of Study 2, offering further support for the possibility that a Poggendorff-like effect may cause pilots to underestimate the location of the horizon in the visual field during nighttime approaches. In addition, Study 4 eliminated the distance confound discussed in Study 3. By maintaining a constant distance between the end of the line segment and the intersection point, we ensured that any observed effects were indeed a product of the stimulus condition. Study 4 increases our confidence in asserting that participants tended to "flatten" the lines, and that such misperceptions could explain the proposed line bias illusion.

We are unsure why the results of Study 4 (without the distance confound) are more pronounced than those of Study 3 (which should have been more difficult). One possibility is that by removing a flat reference cue, the curved response line increased uncertainty and made it harder to judge the angle of the stimulus line relative to the response line. The apparent tilt of the stimulus lines may have therefore been exaggerated, leading to more biased responses. If a similar lack of reference points occurs during BHI conditions, Study 4 may actually be more representative of real world conditions than Study 3 if the pilot is unable to use the glare shield as a flat reference point. Future research should examine how close reference points must be to the stimulus line and whether the glare shield could serve as an anchor to help judge runway edge lines.

General Discussion

The studies described here have collectively helped to clarify potential causes of BHI. Prior hypotheses in the literature on BHI rely primarily on information available to the retina and how pilots use that information to judge altitude. The current findings place retinal information in the context of a broader spatial strategy, implying that pilots attempt to utilize other information beyond what is directly available to the visual senses and that cognitive processes play a role in BHI in addition to purely visual processes. We believe similar processes guide both day and night approaches, but critical cues become biased at night, reducing pilots' ability to fly an accurate glideslope. We first summarize the findings of each of the studies described above, as well as how they collectively support a single explanation for BHI effects. We next describe an apparent inconsistency in our findings. We also discuss the possible influence of other spatial cues, using line splay as an example. Finally, we discuss the limitations of the study and directions for future research, and concluding thoughts.

Summary of Findings

Study 1 narrowed the hypotheses reviewed by Gibb (2007) and others, and Studies 2-4 explored the line bias illusion as a novel causal mechanism for BHI. Study 1 supported the idea that longer runways can increase Black Hole error, as well as the notion that the lack of contextual features may impair pilot judgments of runway shape or size. Eye tracking data indicated that pilots tended to focus most of their attention on the lower half of the runway towards the aim point, although some pilots appeared to glance towards the far threshold/horizon.

Study 2 demonstrated that the line bias illusion is a plausible explanation for BHI. Participants tended to place the intersection point of two converging lines too low relative to the true intersection point, implying that pilots may do the same with a horizon located at the vanishing point of runway edge lines. Further, this estimate shifted depending on the angle of the stimulus lines. Studies 3 and 4 provided complimentary information to support the findings of Study 2 by demonstrating that people have difficulty extending line segments to their intersection point with another line. Study 3 used straight response lines, which improves face validity in the context of a flat glare shield and horizon. Study 4 used curved response arcs, which allowed more precise measurement of participant perception by eliminating the distance confound discussed previously. Studies 3 and 4 each demonstrated that participants had difficulty judging the intersection point of the stimulus lines, and Study 4 in particular provided support for the notion that participants may have perceived the stimulus lines as flatter than they actually were. Together, these studies reinforce the findings of Study 2 and offer a possible causal mechanism behind the line bias illusion.

A Possible Explanation for BHI Based on the Four Studies

The line bias illusion appears to affect pilots' ability to utilize the horizon and H angles as they normally would in a daylight approach. Study 1 provided support for the hypothesis that pilots may not accurately perceive the size and shape of the runway under BHI conditions, and Studies 3 and 4 indicated that people have difficulty perceiving the angle of slanted lines. Eye tracking data (though unreliable) further indicated that pilots focused most of their attention on the lower half of the runway, towards the aim point. Error was worse at the lower half of the dotted lines presented in Studies 3 and 4, implying that pilots focus on an area where they are least able to judge line angles. Together, these three studies indicate that pilots may misperceive the relative angles of the runway edge lines at night. As demonstrated in Study 2, the effect of this misperception appears to be that the pilots judge the implicit horizon to be too low in the visual field. However, Study 2 also demonstrated that responses were more accurate at greater splay angles, implying that estimates of the implicit horizon would become less biased as the pilot approaches the runway. Thus, absolute H angle would appear to increase over time, leading the pilot to fly lower in an attempt to keep this cue stable. Pilots' ability to use other information sources may be compromised as well if the information gained from expansion of the near threshold towards the glare shield becomes more difficult to use as a secondary cue in the absence of redundant information from expansion of the far threshold towards the horizon, or if the position of the glare shield in the sight picture fluctuates due to turbulence or other factors (affecting relative H angle).

We further speculate on a mechanism by which longer runways (Study 1) may compound BHI effects. If viewed from a set distance, a runway of a given length will appear longer in the visual field when viewed from a higher elevation than a lower elevation. This perception is largely irrelevant in the context of a visible, stable horizon which can guide visual interpretation of the runway as its apparent size and shape change over the course of the approach. However, if the pilot is *estimating* the location of the horizon and places it too low in the visual field, the apparent closeness of the far threshold to the horizon may be interpreted as a cue that the pilot is too high and prompt a lower approach. As the far runway threshold expands toward the horizon in the visual field during descent, the pilot may continue flying lower to compensate.

Resolving Inconsistencies in our Findings

Though Studies 2-4 present a consistent series of findings regarding underestimates of the intersection point and what may lead to such estimates, one apparent discrepancy across the studies is worth addressing. Study 2 indicated that participant error was greater at smaller line angles off the vertical, whereas Studies 3 and 4 indicated that participant error was greater at larger line angles. We believe that response error in our studies can be considered to be a product of uncertainty, which can be thought of as a combination of the distance over which an estimate must be made and the tilt of the stimulus lines. Error in Study 2 was likely driven by the distance over which the estimate had to be made at the smaller angles compared to the larger angles. In contrast, Studies 3 and 4 did not require estimating over such large distances. Error in Studies 3 and 4 was therefore driven primarily by uncertainty related to the tilt of the stimulus lines. In addition, the way in which the stimuli were presented across the studies may also have contributed to the seemingly contradictory pattern of effects.

Stimuli for Studies 3 and 4 were presented as a single line with no other frame of reference aside from the response line (which did not vary from trial to trial). Differences in error across the different line angles can therefore be attributed solely to the changes in line angle. In contrast, Study 2 presented two line segments to the participants, such that each may have provided a reference point for the other that changed with every trial. Recall from our earlier discussion of Study 2 that we maintained a constant distance between the end points of the line segments in order to reduce possible confounds. Because of this, the intersection point was much lower on the screen for larger angles than smaller angles. This had two effects: first, the distance over which estimates had to be made was smaller at greater angles (potentially reducing error), and second, there may have been a floor effect as the closer intersection point provided less margin for participants to be too low in their estimates. In addition, the two lines appeared to “point” toward one another at greater angles, potentially making estimates simpler. Despite the apparent discrepancy between the direction of our findings across these studies, we note that we still observed error at the smaller angles in Studies 3 and 4 that was consistent with a flattening of the lines. The proposed explanation for the tendency to place the intersection point too low in the visual field therefore remains plausible.

What is the Role of Other Strategies and Cues?

Our findings are best interpreted in the context of a visual approach strategy of maintaining a stable absolute H angle to control glide path, but other visual cues and strategies are possible as well. For example, all else being equal, the splay angle of parallel lines appears smaller when viewed from higher altitudes and larger when viewed from lower altitudes. Flach et al. (1997) found evidence that the splay angle of parallel lines can be used as a cue to help regulate altitude in a simulated flight task. However, the stimuli used in Flach et al.’s study contained far more visual information that would aid in judging splay angle (e.g., parallel lines all ran completely to the horizon) than would be found under BHI conditions, and we do not believe splay is a primary cue pilots use during their approach to land for a variety of reasons.

Changes in splay angle are an excellent trend monitor for judging whether one is ascending or descending, but splay by itself is of limited value for absolute altitude judgments. Splay angle of the convergence point at the horizon varies with the spacing between the parallel lines. Wider runways would have a different apparent splay than narrow runways when viewed from the same altitude. Further, our informal observations during the flight simulations indicate that changes in splay are extremely gradual and difficult to see with the naked eye until relatively late in the approach to the runway (1nm or less), making it seem unlikely that this strategy could be used at greater distances even in daytime approaches. In addition, Study 1 demonstrated that runway length affected glide path, but runway length does not affect splay angle. We also found that participants perceived the line segments as too horizontal. Such a misperception would imply a greater splay angle and thus a perception of being too low rather than too high. We therefore do not believe that our results support line splay as the primary visual strategy.

However, splay angle may still provide an important cue allowing pilots to recognize the occurrence of BHI and correct prior to impact with the ground (leading to the curved approach paths characteristic of BHI in ours and prior simulator studies). As mentioned above, splay angle does not change rapidly until relatively close to the runway. When this angle does begin to change rapidly, however, it provides an unmistakable cue regarding rate of descent. Galanis and colleagues (1998) asserted that although the horizon-based H angle provides the least uncertainty at greater distances from the runway, runway shape relations become a more reliable cue at closer distances. The sudden increase in splay of the edge lines in a BHI approach compared to a normal approach likely serves as a cue that helps pilots recognize the presence of BHI and alter their glide path prior to impacting the ground. Future research should examine the role of secondary cues such as line splay, and how they may interact with the cues/strategies described above.

Limitations and Future Research

The results of Studies 2-4 form a coherent narrative consistent with the hypothesis that pilots may misperceive the horizon under certain conditions, believing it to be lower in the visual field than it actually is. However, the immediate implications for the flight environment are limited. The stimuli for these studies were limited to very simple computer tasks that did not fully replicate the appearance of runway edge lights at night. Our stimuli were larger than runway edge lines would appear in the visual field, particularly early in the approach phase.

Further, the spacing of the dots in the stimulus lines was much greater than would be observed during a nighttime approach towards lighted runway edge lines, and the static display lacked the dynamic visual changes that would be associated with descending towards the runway. As such, we cannot declare with certainty that the same effects observed here would be observed under visually representative flight conditions. Future work will directly test how well people are able to judge line angle under stimulus conditions more representative of nighttime approaches.

In addition to the limitations that prevent us from asserting that the line bias effect occurs in the flight environment, we cannot speak to whether the presence of the line bias effect would actually lead to BHI in-flight. The current studies did not allow us to directly test the hypothesis that the line bias illusion leads to BHI. We did not assess whether pilots try to estimate the location of the horizon (consciously or unconsciously) as part of a nighttime landing strategy, nor did we assess the impact of underestimates of the horizon's location on the pilots' adopted glide path. Future studies will investigate whether pilots do in fact appear to estimate the location of the horizon, and whether low estimates of the horizon lead to altered approach paths. Of particular interest would be the effect of changes in the estimate over the course of the approach.

Our study utilized static stimuli in order to better isolate the effects of stimulus line angle and dot spacing on participants' estimates. However, static stimuli do not capture the dynamic changes of the cues in the visual field (or the interaction between them) as the pilot approaches the runway. The influence of the line bias illusion in a dynamic setting is an area ripe for future research. Change over time during an approach is likely a key additional piece of information that allows the whole to become more than the sum of its visual parts. For instance, the apparent length of the runway edge lines and the splay angle created by those lines each increase as the pilot approaches the runway and descends. This interaction may provide redundant cues that make altitude interpretation more reliable, and the rate of change in these cues provides additional information regarding the speed and angle of descent.

We have interpreted our findings in the context of a strategy focused on maintaining a stable absolute H angle, rather than one based on runway aspect ratio or splay angles. Future work should examine these various strategies to determine how each contributes to pilot judgements of altitude and glide path, but more importantly how they interact with one another. Multiple cues and strategies likely contribute to pilots' ability to judge altitude during an approach (Galanis, Jennings, & Beckett, 1998; Kim, Palmisano, Ash, & Allison, 2010). Identifying these strategies and describing how they interact with one another will help identify ways in which they may fail during BHI conditions. This in turn will allow us to develop and test mitigation strategies in the form of new training, new displays, or new runway indicators.

Conclusion

The studies described in this report collectively serve to help narrow the possible causes of BHI, while introducing the line bias effect as a novel hypothesis not previously described. Whereas prior studies have largely focused on failures within a particular sensory system (e.g., the visual system), we have attempted to place observed effects in the context of a broader cognitive spatial strategy. Rather than identify what pilots may do differently during night approaches, we have assumed that pilots attempt to use the same spatial strategies during day and night approaches and then attempted to explain how this strategy may fail in the absence of various cues at night. Adopting this more holistic stance may help identify likely failure points that lead to BHI, along with potential interventions. We believe that our findings are consistent with the hypothesis that maintaining a stable absolute H angle is the primary spatial strategy, but more work remains to be done to identify the corpus of strategies that pilots use during approach and how these strategies interact with one another. Such an understanding will facilitate better predictions about the conditions that will cause BHI, whether a pilot will recognize BHI, and how BHI can be prevented in the first place.

References

Andrade, E. (2011). Excessive confidence in visually-based estimates. *Organizational Behavior and Human Decision Processes, 116,* 252-261.

Boeing (2006). Statistical summary of commercial jet airplane accidents.

Erkelens, C. (2015a). Perception of perspective angles. *I-Perception, 6,* 1-11.

Erkelens, C. (2015b). The extent of visual space inferred from perspective angles. *i-Perception, 6,* 5-14.

Flach, J., Warren, R., Garness, S., Kelly, L., & Stanard, T. (1997). Perception and control of altitude: Splay and depression angles. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 1764-1782.

Galanis, G., Jennings, A., & Beckett, P. (1998). A mathematical model of glide-slope perception in the visual approach to landing. *The International Journal of Aviation Psychology, 8,* 83-101.

Gibb, R. (2007). Visual spatial disorientation: Revisiting the black hole illusion. *Aviation, Space, & Environmental Medicine, 78,* 801-808.

Gibb, R., Schvaneveldt, R., & Gray, R. (2008). Visual misperception in aviation: Glide path performance in a black hole environment. *Human Factors, 50,* 699-711.

Gibson, J. J. (1966). The senses considered as perceptual systems. Boston, MA: Houghton Mifflin.

Gillingham (1992). The spatial disorientation problem in the United States Air Force. *Journal of Vestibular Research, 2,* 297-306.

Jacobs, D., Morice, A., Camachon, C., & Montagne, G. (2018). Eye position affects flight altitude in visual approach to landing independent of level of expertise of pilot. *PLoS ONE, 13, e0197585.*

Khatwa R, Helmreich R. L. (1998). Analysis of critical factors during approach and landing in accidents and normal flight: Data acquisition and analysis working group final report. *Flight Safety Digest,* Killers in Aviation: FSF task force presents facts about approach-and-landing and controlled-flight-into-terrain accidents 1998; Nov-Dec, Jan-Feb: 1-77.

Kim, J., Palmisano, S., Ash, A., & Allison, R. (2010). Pilot gaze and glideslope control. *ACM Transactions on Applied Perception, 7,* article 18 (18 pp.).

Lintern, G., & Koonce, J. (1991). Display magnification for simulated landing approaches. *The International Journal of Aviation Psychology, 1,* 59-72.

Lintern, G., & Liu, Y. (1991). Explicit and implicit horizons for simulated landing approaches. *Human Factors, 33,* 401-417.

Lintern, G., & Walker, M. (1991). Scene content and runway breadth effects on simulated landing approaches. *The International Journal of Aviation Psychology, 1,* 117-132.

Matthews RS, Previc F, & Bunting A. (2002). "USAF spatial disorientation survey". Paper presented at the Research and Technology Organization/Human Factors and Medicine Symposium on Spatial Disorientation in Military Vehicles: Causes, Consequences, and cures. April; La Coruna, Spain. HFM-085, RTO-mp-086.

Murray, R., Allison, R., & Palmisano, S. (2009). Glideslope perception during aircraft landing. In E. Leigh (Ed.), SimTect 2009 Conference Proceedings (pp. 87-91). Sydney, Australia: Simulation Industry of Australia. Nicholson & Stewart, 2013.

Nicholson, C., & Stewart, P. (2013). Effects of lighting and distraction on the black hole illusion in visual approaches. *The International Journal of Aviation Psychology, 23,* 319-334.

Patterson, F., Williams, H., Folga, R., & Arnold, R. (2015). Effects of spatial disorientation countermeasures training on black hole illusion (BHI) onset distance and recovery altitude. *Paper presented at AsMA 2015, May 10-14, Lake Buena Vista, Fl.*

Perrone, J. (1982). Visual slant underestimation: A general model. *Perception, 11,* 641-654.

Perrone, J. (1983). Visual slant misperception and the "black-hole" landing situation. *NASA technical memorandum 85866.*

Poulton, E. (1985). Geometric illusions in reading graphs. *Perception & Psychophysics, 37,* 543-548.

Sipes, W., & Lessard, C. (2000). A spatial disorientation survey of experienced instructor pilots. *IEEE Eng Med and Biol Mag, Mar-Apr,* 35-42.

Thompson, R. (2010). The "Black Hole" nigh visual approach: Calculated approach paths resulting from flying a constant visual vertical angle to level and upslope runways. *The International Journal of Aviation Psychology, 20,* 59-73.

Appendix 1: Code for implementing dot spacing in studies 2-4.

```
for i in range(0, 10):
		if global.even:
			var thisdot = dotsprite.instance()
			var dotpos = Vector2()
			var thisradius = global.lineradius/9 * global.evenspacecenter[i]
			dotpos.x = -thisradius * sin(deg2rad(trialangle))
			dotpos.y = thisradius * cos(deg2rad(trialangle))
			thisdot.set_pos(dotpos)
			dottedline.add_child(thisdot)
		else:
			var thisdot = dotsprite.instance()
			var dotpos = Vector2()
			var thisradius = -global.lineradius * ((global.unevenspace[i] + 50.5) /49.5)
			dotpos.x = -thisradius * sin(deg2rad(trialangle))
			dotpos.y = thisradius * cos(deg2rad(trialangle))
			thisdot.set_pos(dotpos)
			dottedline.add_child(thisdot)


var unevenspace			= [-100, -81, -64, -49, -36, -25, -16, -9, -4, -1]


lineradius = 200 also
well global.lineradius in the code
```