Simon Müller, Fabienne Roche, Dietrich Manzey

Attitude indicator format

How difficult is the transition between different reference systems?

Müller, S., Roche, F., & Manzey, D. (2019). Attitude Indicator Format. Aviation Psychology and Applied Human Factors, 9(2), 95–105. https://doi.org/10.1027/2192-0923/a000168

Attitude Indicator Format: How Difficult Is the Transition Between Different Reference Systems?

Simon Müller, Fabienne Roche, and Dietrich Manzey

Work, Engineering, & Organizational Psychology, Department of Psychology and Ergonomics, Technische Universität Berlin, Berlin, Germany

Author Note

The authors thank all experimental participants and Stephan Pietschmann for his software support.

Corresponding author: Simon Müller, simon.mueller@tu-berlin.de, Department of Psychology and Ergonomics, Technische Universität Berlin, Marchstr. 12, 10587 Berlin, Germany

Abstract

A simulator study investigated the consequences of a transition between two alternative formats of the attitude indictor in aircraft cockpits, the moving-horizon and moving-aircraft format. Two groups of novices practiced performing two flight tasks (flight-path tracking and recovery from unusual attitudes) with one attitude-indicator format for six practice sessions, before transitioning to the other format. The results show that, after practice, participants were able to perform both tasks equally well with both attitude-indicator formats. However, the number of reversal errors in the recovery task increased considerably when transitioning from the moving-aircraft to moving-horizon format. No such effect emerged for the other direction. This suggests that the former transition is more difficult and represents a possible risk for flight safety.

Keywords: attitude indicator, transition, spatial disorientation, display design principles, display–control compatibility

Attitude Indicator Format: How Difficult Is the Transition Between Different Reference Systems?

On January 10, 2000, a Saab 340B crashed near Nassenwil, Zürich, Switzerland. Ten passengers (including three crewmembers) were killed. The accident occurred after the pilots failed to recover from a surprising bank movement of their aircraft to the right. The accident investigation report identified as one important contributing factor the pilot's inappropriate rudder input to the right, which further increased the unusually high bank attitude and thus aggravated the situation. This was attributed to the fact that the pilot flying, who was originally trained on Russian aircraft equipped with a so-called *moving-aircraft* (MA) attitude indicator (AI), lost his spatial orientation presumably owing to a misinterpretation of the Western *moving-horizon* (MH) AI (Eidgenössisches Departement für Umwelt, Verkehr, Energie und Kommunikation [UVEK], 2002).

This exemplary accident points to the importance of a classic human factors issue in aviation psychology, namely, the question of which reference frame should be used to display the aircraft's attitude information to pilots (Fitts, 1947; Previc & Ercoline, 1999). Over the years, several AI formats have been proposed, but only two were widely disseminated. The first one is the MH format, which has always been the common format in Western aircraft. The MH format depicts the attitude with a moving artificial horizon against a fixed aircraft symbol. A clockwise bank movement of the aircraft is depicted by a counterclockwise rotation of the artificial horizon in the AI and vice versa. An upward or downward pitching of the aircraft is shown by downward or upward movements of the horizon line. This is a direct analogy to what one would see from inside the cockpit – a horizon apparently moving relative to the aircraft (Previc & Ercoline, 1999). The second AI format, the MA format (Previc & Ercoline, 1999), has often been used by Russian aircraft manufacturer and is still a standard in current aircraft models, for example, in Tupolev Tu-134 and MiG fighters. In this display, bank movements of the aircraft are indicated by corresponding rotations of the aircraft symbol in the display, while the artificial horizon line is kept in a steady horizontal position. This directly corresponds to what one would see if one would observe the aircraft from behind. Only changes in the pitch angle are indicated by movements of the horizon line, corresponding to the depiction in the MH display.

Both formats are driven by different concepts of compatibility. While the MH format serves the *principle of pictorial realism*, that is, it corresponds to what is perceived to move in the outside world when the attitude of an aircraft changes, the MA format serves the *principle of moving part* (Roscoe, 1968), that is, the moving element in the display corresponds to what moves in the pilot's "mental representation" of flying an aircraft (Wickens, 2003, p. 152).

Since both competing AI formats are still in use today, two important questions arise. First, which AI format is superior in terms of higher performance and less prone to misinterpretation? This also involves the question of which of the two compatibility principles referred to earlier is the more dominant one in ensuring intuition of the displayed information. Second, and even more important, what are the consequences of a transition when a pilot has to switch between both AI formats for some reason? The first question has repeatedly been addressed in prior research. Most of this research was conducted during the 1950s-1970s and has been well summarized by Previc and Ercoline (1999). In particular, the results of research with flight novices has provided strong evidence for the MA format being more intuitive and exhibiting a general superiority in supporting quick and correct recoveries from unusual attitudes than the MH format.

More recent research has replicated these findings for the typical AIs integrated in the primary flight displays (PFD) of today's glass cockpits (Ding & Proctor, 2017; Müller, Sadovitch, & Manzey, 2018). Furthermore, there is evidence that the results gained with flight novices can also be generalized to experienced pilots (Müller et al., 2018; Ponomarenko, Lapa, & Lemeshchenko, 1990). For example, Müller et al. (2018) were surprised by the finding that pilots, certified for instrument flying based on training with MH AIs, were also less prone to reversal errors in recoveries from unusual attitudes when using the MA format compared with "their" MH format.

Several theories have been proposed to explain the supposed superiority of the MA format. The first set of theories considers it as a result of basic perceptual organization, namely, a reversal of *figure–ground perception* and the differentiation of various perceptual spaces when interacting with the three-dimensional environment (Johnson & Roscoe, 1972; Previc & Ercoline, 1999). The basic principle guiding the design of the MH format is that it should mimic the view out of the cockpit window showing the movements of the natural horizon in case of attitude changes. This presupposes that pilots perceive the displayed artificial horizon as the stable (back-)*ground* and any movements of this horizon as movements of the airplane (*figure*), as they intuitively do when interpreting movements of the natural horizon. However, the artificial horizon line in an MH format does not fulfill the usual characteristics of a ground. Instead it represents characteristics of a typical figure, that is, a comparatively small moving element included in a much larger and stable instrument panel. This can easily lead to a figure–ground reversal in a way that the pilot does not perceive the moving artificial horizon as the ground in front of which the airplane is moving, but as the figure that is directly controlled by the steering movements at the yoke or sidestick. As a consequence, they tend to execute a reversed control input compared with what was necessary to control the attitude of the aircraft when interacting with the MH AI.

This fits directly the distinction of perceptual spaces in the *neuropsychological reference model* proposed by Previc and Ercoline (1999). Specifically, they propose that different major brain systems are involved when interacting with the environment. The first and closest one is the *peripersonal system*. It processes information of the perceptual space within reaching distance. Movements of objects in this space are usually perceived as movements relative to oneself (as stable reference) and not as consequences of self-motions. By contrast, the *ambient extrapersonal system* processes information from far distances. Large-scale movements in this space are usually perceived as consequences of self-motions against a stable environment. Based on this theory, the main issue of the MH format is that it tries to induce a perceptual effect typically for movements processed by the ambient extrapersonal system by presenting the related information on a small instrument in the peripersonal space. Directly corresponding to the figure–ground reversal effect, this should lead pilots to intuitively perceive movements of the artificial horizon as movements caused by their steering inputs rather than by a consequence of their self-motion (Previc & Ercoline, 1999).

Another theoretical explanation of the assumed advantages of the MA versus MH format relates to *ideomotor theory* (Greenwald, 1970) and effects of *response–effect compatibility* (Janczyk, Pfister, Crognale, & Kunde, 2012; Janczyk, Yamaguchi, Proctor, & Pfister, 2015). According to this theory, "it is easier to produce actions that predictably produce consequences that are compatible rather than incompatible with the action itself" (Janczyk et al., 2012, p. 2). While the MA format is compatible in this respect, that is, right/left movements at the controls lead to directly compatible right/left movements of the moving part (aircraft symbol) in the display, the MH format is response–effect incompatible, that is, right/left movements at the controls lead to left/right movements of the artificial horizon. Numerous studies have shown that response–effect incompatible actions indeed need longer and are more prone to errors than compatible ones, both for discrete manual responses (e.g., Kunde, 2001) as well as for continuous rotary movements (Janczyk et al., 2015).

This also suggests that a transition from the MH to the MA format should be easier to achieve than the other way around. Transitioning from the MH to the MA format involves a transition from an inferior to a more intuitive and compatible format and, thus, should be achieved faster and with less risks of misinterpretation compared with the reverse transition. Evidence supporting this assumption is provided from studies with pilots who were extensively trained with an MH format and then had to perform simulated flight tracking tasks or recoveries from unusual attitudes with the MA format. In the majority of such studies, pilots were able to perform these transition tasks with only small or even no performance impairments (e.g., Browne, 1954; Cohen, Otakeno, Previc, & Ercoline, 2001; Gardner, Lacey, & Seeger, 1954; Müller et al., 2018; Roscoe & Williges, 1975). This is in line with findings of Kovalenko (1991) suggesting that almost one third of all experienced pilots, flying with the standard MH format, appear to still keep a mental representation of spatial bank movements that is more aligned with the depiction of attitude changes in an MA than an MH display.

In contrast, fewer studies have investigated the transfer from an MA to an MH display, which seem to be even more important given the fact that today pilots trained on Russian aircraft with MA AI often need to transfer to aircraft with Western technology. As far as we are aware of, the only exception is a study by Yamaguchi and Proctor

(2010). They trained two groups of 20 flight novices each to perform simulated flight path tracking tasks with the MH format or the MA format, respectively. Both groups showed a similar practice progress and finally performed the tasks equally well. Subsequently, both groups had to switch to the alternative AI format for a transition session. In contrast to our assumption, no asymmetric transition effects for MH to MA (easy) and MA to MH (difficult) transitions were found. Instead, both directions of transition led to tracking performance decrements to a similar degree, suggesting that risks of transitions might be comparable in both directions. However, the study involved somewhat artificial displays and the results were based on tracking tasks only. Tracking tasks solely require comparatively small continuous control movements in order to compensate slight attitude changes that can be monitored constantly, which might be performed well even with a different display format. Thus, it might be questioned to what extent the findings can be generalized to the typical AIs integrated in today's PFDs and the more critical recovery tasks. Recovery tasks are considered as particularly critical with respect to two aspects. First, they put higher demands on pilots than tracking tasks do, owing to their unexpected nature and the rather unusual flight attitudes, which have to be recovered. Second, recovery tasks simulate safety critical flight situations where incorrect responses can directly lead to fatal accidents.

Anecdotal reports from aviation indeed suggest that attitude recovery tasks might make a difference with respect to the relative risks of negative transfer effects involved in transitions of the AI format. In contrast to the relatively easy MH–MA transition, the switch from an MA to MH format seems to be coupled with negative performance consequences, especially in high-stress situations, and might even involve severe risks for flight safety. Dramatic examples of such anecdotal evidence are the crash of the Saab

340B Crossair CRX 498 in January 2000, described in our introductory example, and the crash of the Boeing 737-505 Aeroflot Flight 821 in September 2008 with 88 fatalities. In both accidents, the pilots, who were originally trained on MA-formatted instruments and had extensive experience on Russian aircraft, flew a Western aircraft equipped with an MH AI after a relatively short re-training phase, without a specific differential training on Russian and Western aviation. The reports of both accident investigation boards (Interstate Aviation Committee, 2008; UVEK, 2002) presumed that one important contributing factor to these fatal accidents was that the pilots responded falsely to the MH-formatted AI by falling back to their previously learned heuristics; that is, responding to unusual attitude change based on an MA AI.

In order to evaluate the possible risks of transitioning between the different AI formats, the current study directly compared the performance consequences of transitions from MH to MA and MA to MH formats, based on a display design closely resembling PFDs implemented in today's glass cockpits (i.e., A320). Because pilots trained exclusively on Russian aircraft are almost impossible to recruit for such research and given that our previous research comparing the AI formats showed similar effects for flight novices and pilots (Müller et al., 2018), we decided to base the entire research on flight novices. After providing intensive training with either MH- or MA-formatted AI, participants were requested to transfer to the AI format they were not trained with previously. Transfer-induced performance consequences were then evaluated for both simulated flight-path tracking and for recoveries from unusual attitudes. We expected to reproduce Yamaguchi and Proctor's (2010) results regarding the tracking task, but we assumed asymmetric transition effects to become apparent in the recovery task.

Method

Participants

A total of 31 participants were recruited to take part in the study. This sample size was based on an a priori power analysis using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007), which revealed that a sample size of 30 would enable us to detect at least a medium effect with a probability of 95% and a significance level of 5% for a between-group comparison.

The sample consisted of nine female and 22 male participants. Most of the participants (n = 23) had no prior knowledge of flying an aircraft whatsoever. The eight remaining participants had some limited experiences based on casually flying in flight simulators of different fidelity (including PC games). The participants were randomly divided into two groups. The mean age of the MH group was 31.4 years (*SD* = 7.2) and the MA group, 24.6 years (*SD* = 3.5). All participants of the simulator study were treated according to the Declaration of Helsinki. Participation was compensated by either course credits or a monetary expense allowance of € 10.

Apparatus

The experiment was conducted in a PC-based flight simulator consisting of a mock-up Cessna 172 Skyhawk SP G1000 cockpit panel with an integrated screen displaying a PFD. The input device was a Logitech Extreme 3D Pro joystick. The outside view, generated by an X-Plane flight simulation, was projected on the wall approximately 1.2 m in front of the cockpit. The PFD recreated the overall design of an Airbus A320 PFD (see Figure 1). However, some adjustments were made to meet the requirements of the experimental flight tasks and accommodate the fact that the participants were novices (i.e., the flight mode annunciator, indications of control limits, and bank indicator were hidden). The AI of the PFDs was 8.3 cm high and 7.1 cm wide and the participants were seated about 60 cm in front of the cockpit panel. The simulation was a simplified linear flight model with two degrees of freedom in pitch and bank so that novices could easily learn how to fly. The joystick inputs were linearly transferred into pitch and bank rates. There was no need for thrust control.

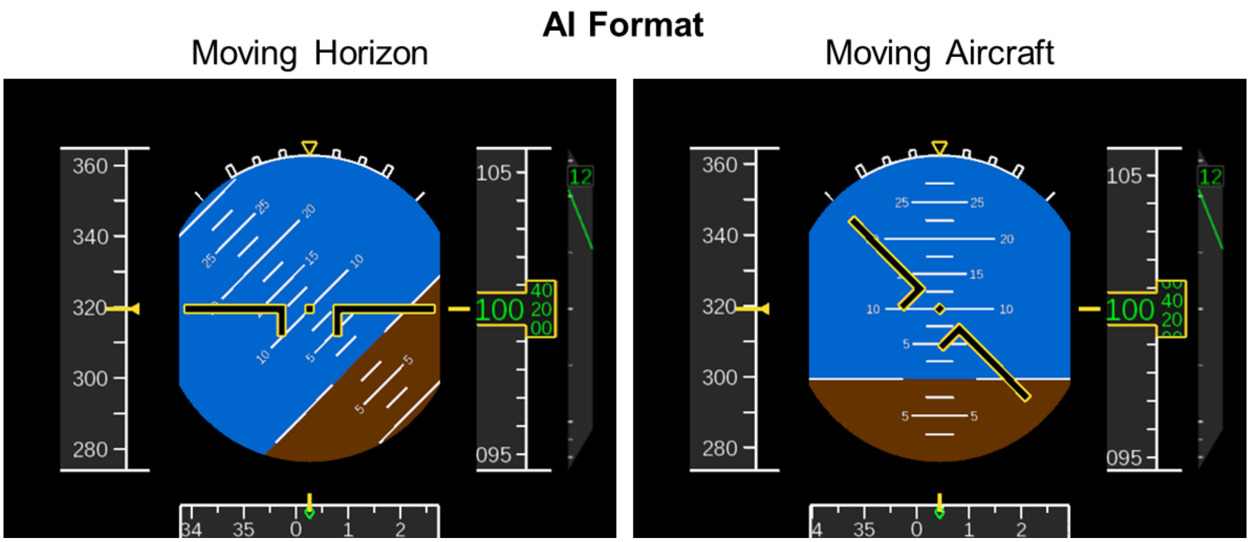

Figure 1. Both primary flight display (PFD) configurations used in the experiment. The left side shows the moving-horizon (MH) format and the right side the moving-aircraft (MA) format. Both PFDs show a bank angle of 45° to the right and pitch up of 10°. AI = attitude indicator.

Tasks

Tracking. Participants were required to hold a stable horizontal flight with pitch and bank angle of 0°. Continuous corrections on an *x*- and *y*-axis by means of the joystick were necessary to compensate two separate preprogrammed disturbance functions based on the study of Fracker and Wickens (1989). The amplitude in bank was three times higher than the amplitude of the pitch function.

Recovery. Sudden attitude changes of the aircraft by 45°, 90°, and 135° to the left or right had to be recovered to a stable horizontal flight. The pitch angle always was set to 0° at the moment of attitude change. As soon as the participants regained the horizontal attitude and kept it stable within ±2° for 2 s, the given trial was completed.

Design

The experiment involved a 2 (group) × 7 (session) × 3 (bank angle) design. The first factor was defined as a between-subjects factor and included two groups, that is, novices practicing the MH or the MA format, respectively. The second factor represented a within-subjects factor and included six practice sessions and one transition session. During the practice sessions, the participants practiced the two tasks with "their" AI format. In the seventh session, representing the transition session, participants had to perform the tasks with the AI format they had not practiced. The third factor was used only when analyzing the recovery-task performance and represented the three bank angle deflections to be recovered (i.e., 45°, 90°, and 135°).

Dependent Measures

Tracking task. Tracking performance was assessed in terms of tracking error, quantified by the root mean square error (RMSE) of bank and pitch angle in relation to 0°. This measure varied inversely to how precisely the participants were able to maintain a horizontal flight attitude with the given AI indicators. Because the AI formats did not differ in pitch representation, no effects of the AI format were expected for the pitch RMSE. The bank and pitch angle deflections used to calculate the RMSE were recorded in 60 Hz.

Recovery task. The first performance measure used was percentage of *reversal errors*. A reversal error was defined as an initial joystick input in a direction that amplifies instead of compensating a given bank angle change. As a second measure, the response time, defined as time from the occurrence of an attitude change to the first noticeable joystick input, was assessed. Only trials without reversal errors were included in this measure.

Workload. To assess the subjective workload, the NASA-TLX (Hart & Staveland, 1988) was used without weightings (Hart, 2006).

Procedure

The experimental procedure consisted of an instruction part, six practice sessions, and a transition session.

Instruction part. At the beginning of the instruction part, the participants read a standardized PC-based instruction, including general information on the test procedure and the task. Subsequently, a consent form was signed. This was followed by an accommodation phase, where the participants could familiarize themselves with the simulation and aircraft controls. During this accommodation phase, only the outside view was shown as a reference for movements of the aircraft, and the participants were asked to maneuver the aircraft (including different turns and level flights) and try to familiarize themselves with the joystick controls. This accommodation phase was followed by instructions on the use of each group's AI for a proper assessment of the aircraft's attitude. The subsequent familiarization phase then included flying with both an outside view as well as the PFD. After 2 minutes, the outside view was blanked and only the PFD was left for spatial orientation and attitude control of the simulated aircraft. Every participant completed several defined flight tasks and a free flight phase in order to become familiarized with the logic of their AI display.

Practice part. The following practice phase was introduced by a description of the tracking and recovery tasks followed by six practice sessions. Each practice session lasted about 7 min and included a 2-min trial of the tracking task and a set of 24 trials of the recovery task (3 bank angles × 2 directions × 4 replications). The latter trials were performed in random order with the time between trials varying from 5 to 12 s. The order of tracking and recovery tasks during the sessions was counterbalanced across participants. Small breaks of approximately 1 min were included between the sessions.

Transition part. The transition session directly followed the sixth practice session. This session started with an instruction explaining the basic principle of the new AI format. Then, the same set of tasks as in the practice sessions had to be performed.

Ratings of perceived workload were collected by means of the NASA-TLX after the first, the sixth, and the transition session. The experiment took about 90 min per participant.

Data Analysis

The statistical data analyses of the performance data were preceded by an outlier analysis. In the tracking task, participants were excluded if their bank RMSE in any of the sessions was more than three standard deviations above the mean of their experimental group. In the recovery task, only trials meeting two criteria were considered in the analysis: a response time of greater than 100 ms and a successful completion within 10 s. Participants who could not successfully finish more than 25% recovery trials in one of the sessions were removed entirely from the analysis.

Performance data derived from the tracking and the recovery task were then analyzed by analyses of variance (ANOVA) with repeated measures. An initial analysis investigated the possible effects of the AI format on the absolute level of performance and performance gain achieved through practice. This analysis included the two experimental groups and Practice Sessions 1–5. A second analysis included only Practice Session 6 and the transition session and addressed the performance consequences of the different transitions (MH–MA and MA–MH). Percentage data of reversal errors were arcsine transformed before the analyses in order to achieve better distribution characteristics (Sokal & Rohlf, 1981). We report the back-converted descriptive statistics for reversal errors in percent to facilitate interpretation. Since ratings of subjective workload, assessed by NASA-TLX, were only available from the first, sixth, and the transition session, these data were analyzed by 2 (group) × 3 (session) ANOVA. A significance level of $\alpha = 5\%$ was used to consider an effect as significant. If the sphericity assumption was violated (Mauchly, 1940), a correction of the degrees of freedom was performed according to the Huynh–Feldt procedure (Huynh & Feldt, 1976). The size of each effect is calculated in terms of partial eta-squared (η_p^2), while $\eta_p^2 = 0.01$ represents a small, $\eta_p^2 = 0.06$ a medium, and $\eta_p^2 = 0.14$ a large effect (Cohen, 1988).

Results

Tracking Task

Two participants of the MH group were considered as outliers and were not included in the statistical analysis. Figure 2 shows the mean RMSE scores for the bank and pitch angle of the remaining 29 participants, separated for the two groups and the different sessions of the experiment.

Effects of practice. Figure 2A shows that both groups performed almost equally well during the five practice phases, independent of the respective AI format. This holds true for both the absolute performance level and the performance gain over the course of practice. A significant main effect in the 2 (group) × 5 (session) ANOVA revealed that both groups improved in their bank tracking performance from the first (M = 4.39°, SE = 0.38°) to the fifth session (M = 3.10°, SE = 0.20°), $F(2.92, 78.82) = 12.22$, $p < .001$, $\eta_p^2 = .31$. Neither a significant main effect of group, $F(1, 27) = 0.02$, $p = .882$, $\eta_p^2 < .01$, nor a significant interaction effect of Group × Session, $F(2.92, 78.82) = 0.65$, $p = .578$, $\eta_p^2 = .02$, emerged.

Essentially, the same pattern was also found for pitch RMSE in Figure 2B. The pitch RMSE decreased significantly from Practice Session 1 (M = 2.40°, SE = 0.35°) to Session 5 (M = 1.41°, SE = 0.11°), $F(1.26, 34.03) = 8.87$, $p = .003$, $\eta_p^2 = .25$. Again, neither the main effect of group, $F(1, 27) = 0.53$, $p = .471$, $\eta_p^2 = .02$, nor the interaction effect of Group × Session, $F(1.26, 34.03) = 1.13$, $p = .309$, $\eta_p^2 = .04$, was significant.

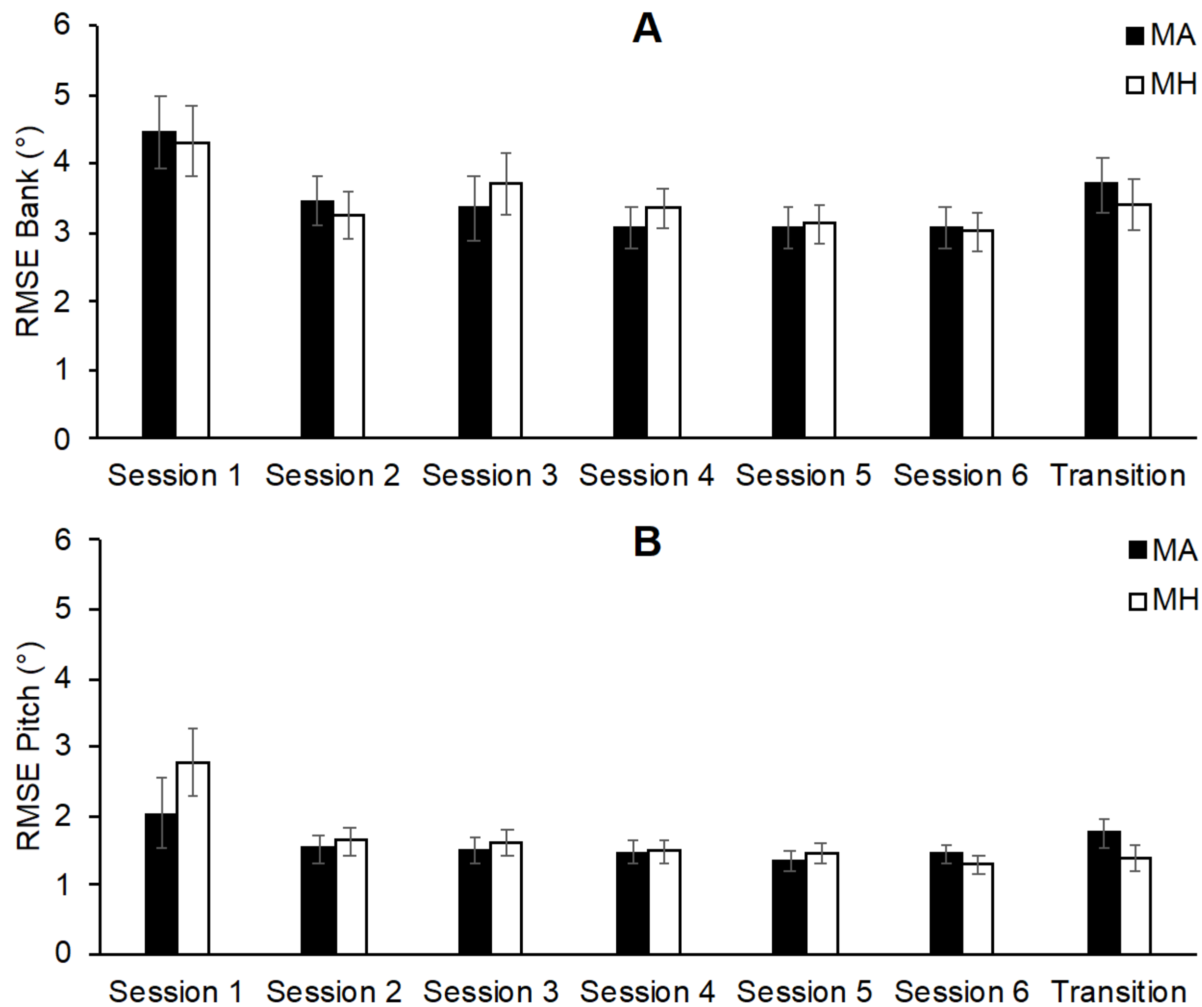

Figure 2. Means of (A) bank and (B) pitch root mean square error over six practice sessions and transition session for both groups, moving aircraft and moving horizon. Error bars represent standard errors. RMSE = root mean square error, MH = moving horizon, MA = moving aircraft.

Effects of transition. As becomes evident from Figure 2A, a significant increase of mean bank error was found in both groups when transferring from the sixth practice session (M = 3.05°, SE = 0.21°) to the alternative format in the transition session (M = 3.54°, SE = 0.27°), $F(1, 27) = 9.20$, $p = .005$, $\eta_p^2 = .25$. However, this effect was relatively small, and a post hoc analysis contrasting the performance in the first practice session and the transition session revealed that the performance in the transition session was still significantly better than the performance in the initial practice session, $F(1, 27) = 8.89$, $p = .006$, $\eta_p^2 = .25$. Neither a main effect of group, $F(1, 27) = 0.15$, $p = .701$, $\eta_p^2 = .01$, nor an interaction effect of Group × Session emerged, $F(1, 27) = 0.51$, $p = .480$, $\eta_p^2 = .02$.

Again, the effects for pitch RMSE were similar, albeit still weaker. The participants showed a slightly higher mean RMSE in the transition session ($M = 1.57°$, $SE = 0.14°$) compared with the final practice session ($M = 1.37°$, $SE = 0.09°$), but this effect was not significant, $F(1, 27) = 3.48$, $p = .073$, $\eta_p^2 = .11$. Again, neither a main effect of group, $F(1, 27) = 1.54$, $p = .225$, $\eta_p^2 = .05$, nor a Group × Session interaction was found, $F(1, 27) = 1.10$, $p = .303$, $\eta_p^2 = .04$.

Workload. Because of missing data, one additional participant had to be excluded from the analysis of the NASA-TLX data. The 2 × 3 (Group [MH, MA] × Session [1, 6, 7]) ANOVA did not reveal a significant difference in perceived workload between the two groups. Although the mean TLX ratings of the MH group were somewhat lower ($M = 33.2$, $SE = 4.4$) than the ratings of the MA group ($M = 44.9$, $SE = 4.8$), this effect was not significant, $F(1, 26) = 3.21$, $p = .085$, $\eta_p^2 = .110$. However, a main effect of session was found, $F(2, 52) = 14.35$, $p < .001$, $\eta_p^2 = .36$, indicating lower ratings after task practice in Session 6 ($M = 37.0$, $SE = 3.3$) and the transition session ($M = 33.6$, $SE = 3.6$), compared with the first practice session ($M = 49.0$, $SE = 3.8$). Most importantly, no significant interaction effect was observed, $F(2, 52) = 1.55$, $p = .222$, $\eta_p^2 < .06$. An inspection of the NASA-TLX subscales revealed that almost all subscales (i.e., *physical demand*, *temporal demand*, *performance*, *effort*, and *frustration*) contributed to the significant session effect in a similar way.

Recovery Task

Based on the outlier analyses, three participants were excluded from the analyses of recovery-task performance. Notably, all three outliers were found in the MH group. In addition, 3.0% of all individual trials were disregarded owing to unsuccessfully finished recoveries or response time constraints. Figure 3 shows the means of reversal error and the response time of the recovery task across all sessions of the experiment.

Effects of practice. As becomes evident from Figure 3A, a clear practice effect emerged in the two groups also for recovery performance. On average, the rate of recovery errors was reduced from 14.0% (SE = 2.6%) in the first practice session to 5.2% (SE = 1.4%) in the fifth practice session. This was reflected in a significant main effect of session in the 2 (group) × 5 (session) × 3 (bank angle) ANOVA, $F(4, 104) = 7.66$, $p < .001$, $\eta_p^2 = .23$. As with the tracking performance, neither the main effect of group, $F(1, 26) < 0.01$, $p = .961$, $\eta_p^2 < .01$, nor the Group × Session interaction, $F(4, 104) = 0.99$, $p = .418$, $\eta_p^2 = .04$, was significant. Yet, the bank angle deflection made a difference. As predicted, both groups made more reversal errors the higher the bank angle deflection was, particularly when comparing the 135° attitude change to both smaller bank angle changes, $F(1.49, 38.84) = 22.16$, $p < .001$, $\eta_p^2 = .46$. This main effect was not moderated by the AI format, as no interaction effect of Group × Bank Angle emerged, $F(1.49, 38.84) = 0.86$, $p = .400$, $\eta_p^2 = .03$. Furthermore, neither the interaction effect Session × Bank Angle, $F(8, 208) = 1.68$, $p = .105$, $\eta_p^2 = .06$, nor the interaction Group × Session × Bank Angle became significant, $F(8, 208) = 0.48$, $p = .866$, $\eta_p^2 = .02$.

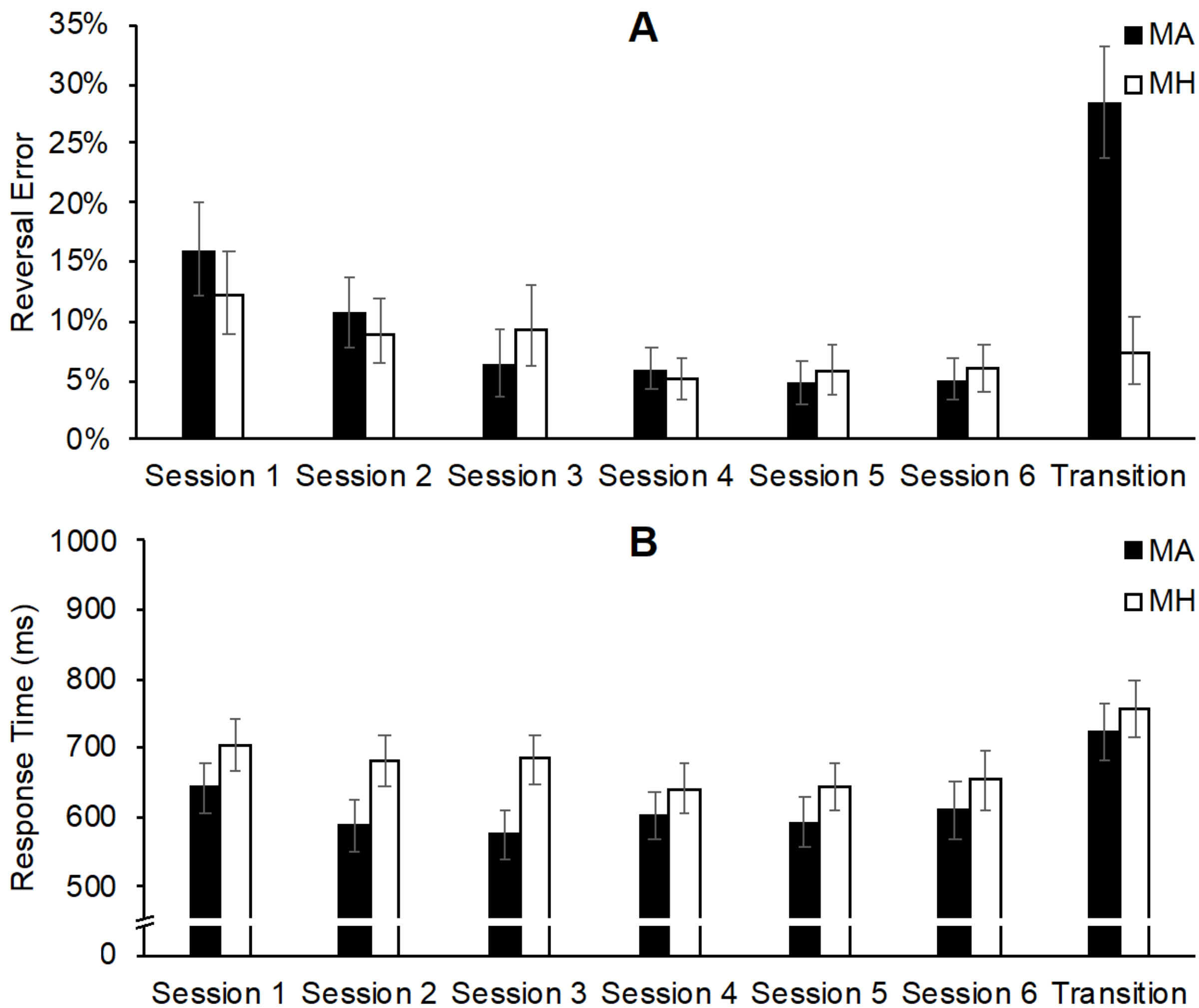

Figure 3. Means of (A) reversal error and (B) response time over six practice sessions and transition session for both groups, moving horizon and moving aircraft. Error bars represent standard errors. MH = moving horizon, MA = moving aircraft.

The participants' response times revealed a similar tendency as the recovery errors. This was reflected in a significant reduction of the mean response time in Session 5 (M = 618.1 ms, SE = 24.9 ms) compared with the first practice session (M = 673.5 ms, SE = 26.4 ms), $F(2.64, 68.63) = 2.94$, $p = .046$, $\eta_p^2 = .10$. Again, a significant effect of bank angle emerged, $F(1.31, 34.15) = 7.02$, $p = .007$, $\eta_p^2 = .21$, indicating a generally lower response time with the smallest bank angle (45°) compared with both higher angles. By contrast, neither the main effect of group, $F(1, 26) = 2.39$, $p = .134$, $\eta_p^2 = .08$, nor any interaction effect was significant, all $F < 2.0$, $p > .11$, $\eta_p^2 < .07$.

Effects of transition. As becomes evident from Figure 3A the mean reversal error rate increased in both groups when transitioning to the alternative AI format. As expected, this increase was much larger for the group transitioning from the MA to the MH format than for the reverse transition. While the mean reversal error rate of both groups was approximately identical in the last practice session (MA: $M = 5.0\%$, $SE = 1.8\%$; MH: $M = 5.9\%$, $SE = 2.0\%$), the reversal error rate of the MA group increased considerably more when switching to the MH format ($M = 28.4\%$, $SE = 4.8\%$) compared with the MH group switching to the MA format ($M = 7.3\%$, $SE = 2.8\%$). In the 2 (group) × 2 (session) × 3 (bank angle) ANOVA, this was reflected in significant main effects of group, $F(1, 26) = 6.31$, $p = .019$, $\eta_p^2 = .20$, and session, $F(1, 26) = 19.20$, $p < .001$, $\eta_p^2 = .42$, and a strong Group × Session interaction effect, $F(1, 26) = 13.82$, $p < .001$, $\eta_p^2 = .35$. In addition, a main effect of bank angle emerged, indicating that the participants committed the more reversal errors the higher the sudden change of bank angle was, $F(2, 52) = 5.95$, $p = .005$, $\eta_p^2 = .19$. This effect was not moderated by the direction of transition, as no interaction effects including the bank angle factor were significant, all $F < 1.9$, $p > .16$, $\eta_p^2 \leq .07$.

For recovery response times, a significant main effect of session emerged, indicating an increase of mean response times in the transition session ($M = 739.2$ ms, $SE = 29.4$ ms), compared with the last practice session ($M = 632.4$ ms, $SE = 30.2$ ms), $F(1, 26) = 24.91$, $p < .001$, $\eta_p^2 = .49$. Yet, neither a main effect of group, $F(1, 26) = 0.47$, $p = .498$, $\eta_p^2 = .02$, nor an interaction effect of Group × Session was observed,

$F(1, 26) = 0.06$, $p = .811$, $\eta_p^2 < .01$. Response times were generally lower for the 45° attitude change compared with the two larger changes, reflected in a main effect of bank angle, $F(1.90, 49.50) = 12.52$, $p < .001$, $\eta_p^2 = .33$. The interaction effect of Group × Bank Angle, $F(1.90, 49.50) = 3.18$, $p = .053$, $\eta_p^2 = .11$, as well as Session × Bank Angle, $F(2, 52) = 2.98$, $p = .060$, $\eta_p^2 = .10$, both just missed the usual level of significance. The threefold interaction of Group × Session × Bank Angle did not reach significance, $F(2, 52) = 1.70$, $p = .192$, $\eta_p^2 = .06$.

Workload. The mean NASA-TLX ratings for the recovery tasks performed in Sessions 1, 6, and 7 did not differ significantly between the two practice groups, $F(1, 26) = 1.25$, $p = .273$, $\eta_p^2 = .05$. In addition, neither the main effect session, $F(1.94, 50.34) = 0.28$, $p = .752$, $\eta_p^2 = .01$, nor the interaction effect Group × Session was significant, $F(1.94, 50.34) = 0.86$, $p = .425$, $\eta_p^2 = .03$.

Discussion

The results of the present experiment provide several interesting theoretical as well as practical insights. First, the effects of practice show that novices were equally able to acquire the skills to cope with tracking and recovery tasks effectively, independent of whether they practiced these tasks with an MA or MH display. Second, a transition from the practiced AI format to an alternative format led to performance decrements in both tasks. While these performance decrements were relatively small and independent of the direction of transition in the tracking task, highly asymmetric transition effects emerged for the recovery task. This suggests that different performance risks are associated with these two directions of transition.

Let us first consider the results of the tracking task. Already in the first practice session, no obvious differences between the groups using MH or MA format were found.

Thus, the short familiarization phase obviously was sufficient to balance any performance differences related to the two formats. This is in line with the findings of Yamaguchi and Proctor (2000) but contrasts the results from our previous work, where we found that novices were better able to perform the tracking task with the MA versus MH display (Müller et al., 2018). While in this previous research the AI format was varied within subjects, the current study as well as the study of Yamaguchi and Proctor (2010) varied the AI format between subjects. The latter might have reduced the sensitivity to find differences between the two formats on group level. Even more important, the task practice with the different AI formats had an equally positive effect on tracking performance during the practice phase and led only to small performance decrements when transitioning to the alternative format, independent of the direction of transition. Actually, tracking performance remained on a higher level than in the initial practice session even in the transition sessions. This suggests that the acquisition of tracking skills in interaction with a certain AI format can be transferred at least to some extent between the different formats. Practice effects were also reflected in the subjective workload ratings. For both AI formats, the ratings were lower after the sixth practice session compared with the first one. In addition, no effect on subjectively perceived workload emerged in the transition sessions, suggesting that the transition was not perceived as effortful by the participants.

The general finding of a symmetric transition effect in tracking performance confirms the results of Yamaguchi and Proctor (2010). Yet, in contrast to our work, they did not find any indication of a positive transfer between the two formats. The fact that their participants had an additional load in a secondary task during task practice with the different AI formats might have prevented the development of general and transferable tracking skills.

Considering these results for the tracking task in our experiment as well as the previous research by Yamaguchi and Proctor (2010), it might be concluded that a transition from MH to MA format is equally demanding as a transition from MA to MH format. Thus, both directions of transitions might easily be achievable for pilots with only limited risks of performance decrements.

However, a different picture emerged when considering the recovery task. The initial practice sessions also proved the principal suitability of both formats for learning an efficient recovery from unusual attitudes based on the AI, but a clear asymmetric performance effect appeared when transitioning from one format to the other. This effect emerged independent of the degree of bank angle deflections and directly supports our hypothesis. While reversal errors increased only slightly after switching from MH to MA format, reversal errors increased substantially when switching from MA to MH format. That the recovery performance did in fact show the expected transition effect while the tracking tasks did not can be explained by the different nature of responses required in both tasks. The tracking task requires continuous corrections of relatively small deviations of the aircraft attitude, which allow for constant feedback of the required responses, while spatial orientation is permanently maintained. The type of AI format does not make a big difference in supporting these movements. By contrast, the recovery task requires fast discrete responses to sudden and rather extreme attitude changes, which put high demands on quick orientation and response selection. It seems immediately plausible that this task has particularly provoked a quick and intuitive response behavior. Consequently, the accuracy of responses suffered more when transitioning from the intuitive MA to the unintuitive MH format than vice versa.

Remarkably, and in contrast to what was found for decrements in tracking performance, the recovery performance decrements induced by the MA to MH transition were so strong that the mean rate of reversal errors in this group was even worse than their average performance level in the first practice session. Obviously, the experience with the MA format reinforces the already more intuitive mental model that the movement in the instrument directly corresponds to the movement of the aircraft. Thus, it makes a shift to the MH format and its inverse logic of dynamics – that is, movements of the horizon to the left indicate movements of the aircraft to the right and vice versa, which is incompatible to the principle of moving part (Roscoe, 1968) – even more difficult than without much practice.

A transition effect was also evident in the response times. In both groups, the time to respond to a given attitude change increased considerably, independent of the degree of bank deflection, and fell back (or even became slightly worse) to the performance level during the first practice session. However, no asymmetric transition effect emerged in this measure. This excludes that the asymmetric effect observed in the reversal errors might be simply explained by a speed–accuracy effect. Interestingly, the (asymmetric) performance decrements in recovery performance were not reflected in the subjective workload ratings linked to this task. In fact, the TLX ratings of the participants did not show any effect of transition. This suggests that the participants subjectively were not aware of the higher demand after a transition to the alternative AI format, which makes the risks involved in such transitions, particularly the MA–MH transition, even higher.

Conclusion and Limitations

The results of the present study extend our previous findings suggesting that pilots trained with the MH format would not have many difficulties to change to the more intuitive MA format, neither with respect to a precise flight-path tracking nor regarding the more unusual recovery task (Müller et al., 2018). Consequently, a practical migration of the current MH AI displays to the superior MA format would likely be feasible with only limited effort of re-training and familiarization time.

At the same time, it also suggests that the reverse transition is much more problematic, supporting conclusions of the Crossair accident investigation (UVEK, 2002). The transition from the MA format to the less intuitive MH format led to a considerable increase in proneness to reversal errors. This effect was presumably caused by a reinforced bias to interpret the dynamic movements in the AI display as directly linked to the bank movements of the aircraft even when flying with the MH format combined with resulting issues of response–effect incompatibility. Theoretically, this bias is assumed to result from combined effects of perceptual organization as proposed by the figure–ground reversal hypothesis (Johnson & Roscoe, 1972) and the neuropsychological reference model (Previc & Ercoline, 1999). The main risk involved in this transition is that pilots will respond intuitively in stressful situations and fall back on earlier learned control behavior. Thus, special caution is advised when pilots change from MA- to MH-formatted flight displays. Intensive and long re-training phases with a particular focus on

AI differences will be needed to override this bias. At best, interpreting the new MH format should be over-trained to avoid the risk of fallback, before pilots transitioning from the MA format go into line operation.

One limitation of the present experiment may be the use of flight novices instead of trained pilots as participants. However, results of our earlier research suggest that the effects of different AI displays found with flight novices are not that different from findings with certified pilots as one might expect (Müller et al., 2018). An actual limitation can be seen in the use of sudden and discrete attitude changes, which represents a rather rare demand in real aviation. A further limitation is that we only tested AIs implemented in head-down PFDs. The situation might be different for synthetic-vision or head-up displays (Beringer & Ball, 2009; Ercoline, DeVilbiss, & Evans, 2004; Pongratz, Vaic, Reinecke, Ercoline, & Cohen, 1999), which remain a matter of future research.

References

Beringer, D. B., & Ball, J. D. (2009). Unknown-attitude recoveries using conventional and terrain-depicting attitude indicators: Difference testing, equivalence testing, and equivalent level of safety. *The International Journal of Aviation Psychology*, *19*(1), 76–97. https://doi.org/10.1080/10508410802597366

Browne, R. C. (1954). Figure and ground in a two dimensional display. *Journal of Applied Psychology*, *38*(6), 462–467. https://doi.org/10.1037/h0057045

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York, NY: Erlbaum.

Cohen, D., Otakeno, S., Previc, F. H., & Ercoline, W. R. (2001). Effect of “inside-out” and “outside-in” attitude displays on off-axis tracking in pilots and nonpilots. *Aviation, Space, and Environmental Medicine*, *72*(3), 170–176.

Ding, D., & Proctor, R. W. (2017). Interactions between the design factors of airplane artificial horizon displays. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *61*(1), 84–88. https://doi.org/10.1177/1541931213601487

Eidgenössisches Departement für Umwelt, Verkehr, Energie und Kommunikation. (2002). *Schlussbericht des Büros für Flugunfalluntersuchungen über den Unfall des Flugzeuges Saab 340B, HB-AKK, betrieben durch Crossair unter Flugnummer CRX 498* [Final report of the aircraft accident investigation bureau on the accident to the SAAB 340B aircraft, registration HB-AKK of Crossair flight CRX498]. Bern, Switzerland.

Ercoline, W. R., DeVilbiss, C. A., & Evans, R. H. (2004). Flight displays I: Head-down display topics for spatial orientation. In F. H. Previc & W. R. Ercoline (Eds.), *Spatial disorientation in aviation* (pp. 379–449) Reston, VA: American Institute of Aeronautics and Astronautics. https://doi.org/10.2514/4.866708

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191. https://doi.org/10.3758/BF03193146

Fitts, P. M. (1947). *Psychological research on equipment design*. Army Airforces Aviation Psychology Program Research, Report No. 19. Washington, DC: Army Airforces.

Fracker, M. L., & Wickens, C. D. (1989). Resources, confusions, and compatibility in dual-axis tracking: displays, controls, and dynamics. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(1), 80–96. https://doi.org/10.1037/0096-1523.15.1.80

Gardner, J. F., Lacey, R. J., & Seeger, C. M. (1954). *Speed and accuracy of response to five different attitude indicators*. Technical Report 54 (236). Springfield, OH: Wright Air Development Center.

Greenwald, A. G. (1970). Sensory feedback mechanisms in performance control: With special reference to the ideo-motor mechanism. *Psychological Review, 77*, 73–99. https://doi.org/ 10.1037/h0028689

Hart, S. G. (2006). NASA-Task Load Index (NASA-TLX); 20 years later. In HFES. (Eds.), *50th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 904–908). HFES. https://doi.org/10.1177/154193120605000909

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). Amsterdam, The Netherlands: North Holland Press. https://doi.org/10.1016/S0166-4115(08)62386-9

Huynh, H., & Feldt, L. S. (1976). Estimation of the box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of Educational Statistics, 1*(1), 69–82. https://doi.org/10.3102/10769986001001069

Interstate Aviation Committee. (2008). *Aeroflot Flight 821 accident report*. Moscow, Russia: Air Accident Investigation Commission.

Janczyk, M., Pfister, R., Crognale, M. A., & Kunde, W. (2012). Effective rotations: Action effects determine the interplay of mental and manual rotations. *Journal of Experimental Psychology: General*, *141*(3), 489–501. https://doi.org/10.1037/a0026997

Janczyk, M., Yamaguchi, M., Proctor, R. W., & Pfister, R. (2015). Response-effect compatibility with complex actions: The case of wheel rotations. *Attention, Perception, & Psychophysics, 77*(3), 930–940. https://doi.org/10.3758/s13414-014-0828-7

Johnson, S. L., & Roscoe, S. N. (1972). What moves, the airplane or the world? *Human Factors, 14*(2), 107–129. https://doi.org/10.1177/001872087201400201

Kovalenko, P. A. (1991). Psychological aspects of pilot spatial orientation. *ICAO Journal, 46*(3), 18–23.

Kunde, W. (2001). Response-effect compatibility in manual choice reaction tasks. *Journal of Experimental Psychology: Human Perception and Performance, 27*(2), 387–394. https://doi.org/ 10.1037/0096-1523.27.2.387

Mauchly, J. W. (1940). Significance test for sphericity of a normal n-variate distribution. *The Annals of Mathematical Statistics*, *11*(2), 204–209. https://doi.org/10.1214/aoms/1177731915

Müller, S., Sadovitch, V., & Manzey, D. (2018). Attitude indicator design in primary flight display: Revisiting an old issue with current technology. *The International Journal of Aerospace Psychology*, *28*(1–2), 46–61. https://doi.org/10.1080/24721840.2018.1486714

Pongratz, H., Vaic, H., Reinecke, M., Ercoline, W., & Cohen, D. (1999). Outside-in vs. inside-out: Flight problems caused by different flight attitude indicators. *SAFE Journal*, *29*(1), 7–11.

Ponomarenko, V. V., Lapa, V. A., & Lemeshchenko, N. A. (1990). Proyektirovaniya indikatsii prostranstvennogo polozheniya samolyota [Psychological foundation of aircraft's spatial attitude indication design]. *Psykhologicheskiy Zhurnal*, *11*(1), 37–46.

Previc, F. H., & Ercoline, W. R. (1999). The "outside-in" attitude display concept revisited. *The International Journal of Aviation Psychology*, *9*(4), 377–401. https://doi.org/10.1207/s15327108ijap0904

Roscoe, S. N. (1968). Airborne displays for flight and navigation. *Human Factors*, *10*(4), 321–332. https://doi.org/10.1177/001872086801000402

Roscoe, S. N., & Williges, R. C. (1975). Motion relationships in aircraft attitude and guidance displays: A flight experiment. *The International Journal of Aviation Psychology*, *17*(4), 374–387. https://doi.org/10.1177/001872087501700409

Sokal, R. R., & Rohlf, F. J. (1981). *Biometry: The principles and practice of statistics* (2nd ed.). New York, NY: W. H. Freeman.

Wickens, C. D. (2003). Aviation displays. In P. S. Tsang & M. A. Vidulich (Eds.), *Principles and practice of aviation psychology* (pp. 147–200). Mahwah, NJ: Erlbaum.

Yamaguchi, M., & Proctor, R. W. (2010). Compatibility of motion information in two aircraft attitude displays for a tracking task. *The American Journal of Psychology*, *123*(1), 81–92. https://doi.org/10.5406/amerjpsyc.123.1.0081