

Article

Comparing Response Behaviors to System-Limit and System-Malfunction Failures with Four Levels of Operational Proficiency

Junmin Du ^{*}, Padun Yunusi, Shuyang He and Peng Ke 

School of Transportation Science and Engineering, Beihang University, Beijing 100191, China

^{*} Correspondence: dujm@buaa.edu.cn

Abstract: Commercial aircraft are becoming highly automated, but pilots must take control if automation systems fail. Failures can be due to known limitations (system-limit failures) or unforeseen malfunctions (system-malfunction failures). This study quantifies the impact of these failures on response performance and monitoring behavior, considering four levels of operational proficiency. In a flight simulator with pitch, roll, and yaw, 24 participants experienced both types of failures at different proficiency levels. The results showed that system-malfunction failure response times were 3.644, 2.471, 2.604, and 4.545 times longer than system-limit failure response times at proficiency levels 1 to 4. Monitoring behaviors (fixation duration, saccade duration, fixation rate) differed between failure types and proficiency levels. Considering these differences in response performance and monitoring behavior between failure types, it is important to differentiate between system-limit and system-malfunction failures in the literature and not overlook the influence of proficiency. Furthermore, due to the unpredictability of system-malfunctions, it is crucial to develop pilots' psychological models and training theories regarding the operation of automated systems, fostering their core competency to excel in handling unknown situations.

Keywords: system-limit failure; system-malfunction failure; operational proficiency levels; response time; monitoring behavior; commercial aircraft pilot



Citation: Du, J.; Yunusi, P.; He, S.; Ke, P. Comparing Response Behaviors to System-Limit and System-Malfunction Failures with Four Levels of Operational Proficiency. *Appl. Sci.* **2023**, *13*, 8304. <https://doi.org/10.3390/app13148304>

Academic Editor: João M. F. Rodrigues

Received: 20 May 2023
Revised: 8 July 2023
Accepted: 11 July 2023
Published: 18 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the widespread application of automation technology in commercial aircraft, the complexity of aircraft automation systems has been increasing, highlighting the importance of studying the interaction between pilots and aircraft automation systems. Currently, aircraft automation has reached an advanced level. Modern commercial aircraft are equipped with sophisticated automatic flight systems, including autopilot, auto-throttle management, automatic navigation, and automatic landing capabilities. These automation systems provide a high level of autonomy and precision during various flight phases such as takeoff, cruise, descent, and landing. This high level of automation allows pilots to focus more on monitoring and decision-making tasks.

However, despite the convenience brought by automation, there have been numerous instances where aircraft automation failures occur and pilots fail to correctly identify the failure and promptly take over to prevent accidents [1–3]. A notable recent example is the Boeing 737-MAX8 accident, where the Maneuvering Characteristics Augmentation System (MCAS) misjudged the aircraft's state during flight, continuously providing erroneous nose-down commands without adequately considering the pilots' reactions and understanding of the system. The captain of the involved aircraft failed to fully recognize the system error and engaged in a continuous struggle by pulling up against the downward commands from the flight control system. After a series of struggles, the aircraft ultimately lost control and crashed [1].

Airline pilots are given ultimate responsibility and final authority over their aircraft to ensure the safety and well-being of all its occupants [4,5]. In today's airline operations, pilots continue to serve as supervisors and operators, requiring an understanding and adaptation to the workings and limitations of automation systems, as well as the capability to handle system failures and unpredictable circumstances. Human–automation interaction issues persist as a prevalent problem nowadays, thus calling for extensive research into various aspects of pilot behavior in unexpected situations to enhance the synergy between pilots and automation systems and seek effective approaches to problem solving.

1.1. Reaction to Unexpected Events

Research has examined pilot reactions to unexpected events, specifically focusing on startle and surprise on the flight deck [6,7]. In recent years, the prevalence of flight deck automation surprises caused by unexpected events and mode confusions has been investigated. When faced with surprising and startling events such as technical malfunctions or automation surprises, pilots' performance can be significantly impaired, and the likelihood of negative outcomes following unexpected events increases [8–11]. Adverse manifestations of pilot responses to unexpected events may be longer reaction times, greater increases in heart rate and pupil dilation, reduced eye movement scanning, higher workload, or other examples of poorer performance [11–13].

The performance issues in unexpected situations can often be traced back to insufficient adaptation of one's frame to the situation [11]. Studies on automation system failures in vehicle autonomous driving have found that a better understanding of the limitations of driving automation is associated with faster responses. When drivers can anticipate failures, they are more prompt in taking control [14].

Automation may fail as a result of known system limits (system-limit failure) or of malfunctions that are unforeseen by system designers (system-malfunction failure) [15]. The failures caused by known limits in automation systems can be mitigated through training, enabling operators to anticipate them. However, there are certain failures that even the system designers themselves cannot foresee, making it impractical to address them through training.

From the perspective of pilots, for system-limit failures, they are aware of the known system limitations and potential failures. They can anticipate and respond to these failures based on their understanding of the automation system, environmental cues, and self-knowledge. However, for system-malfunction failures that even the designers cannot foresee, pilots cannot rely on pre-established accurate cues or self-knowledge to anticipate and prepare for them. It is worth mentioning that some failures, although anticipated by automation system designers, may not be adequately addressed in pilot training due to various reasons. If pilots lack experience and knowledge in dealing with such failures, they may find themselves in a situation similar to facing system-malfunction failures, wherein they are unable to anticipate and prepare in advance. Therefore, from the perspective of pilots, their ability to anticipate and prepare for failures can serve as a key indicator for distinguishing between system-limit failures and system-malfunction failures, which may be more easily understood from the pilot's viewpoint.

Despite numerous studies conducted on pilot reactions to aircraft automation system failures, over the years, there has been a lack of distinction between system-limit failures and system-malfunction failures, resulting in a partial understanding of performance in responding to these failures. Therefore, there is a need to differentiate between automation system failure types from the perspective of pilots and conduct research on their reactions.

1.2. Operational Proficiency and Failures Type Factor

In fields such as road transportation, studies have examined the impact of different types of system failures on monitoring behavior and takeover performance. It has been found that drivers exhibit different behavioral patterns depending on the type of system failure encountered [14–16]. For example, drivers tend to take control of the vehicle

at a faster speed in the case of system-limit failures compared to system-malfunction failures [15]. The impact of system failure types on drivers has gained attention from researchers, as evidenced by existing studies on trust and other factors in different types of automation failures [17–19].

In the aviation domain, researchers typically study pilot behavior when encountering system-limit failures or system-malfunction failures [5,20–24]. Previous studies often focused on investigating a single type of system failure, and predominantly system-limit failures. There has been limited direct comparison research that specifically investigates the differences in monitoring behavior between different types of failures.

Extensive research has demonstrated the significant impact of pilot experience and operational proficiency on behavioral performance. Pilots are commonly categorized as novices or experts, and studies have investigated the differences between these two groups in terms of their control behavior, eye-tracking characteristics, and more. The results consistently show that experts outperform novices in various aspects. For example, experts demonstrate more effective eye movement patterns compared to novices, higher perceptual efficiency (more fixations with shorter durations), balanced visual attention allocation, more complex and detailed visual scanning patterns, and faster and more accurate decision-making [23–25]. The majority of existing studies have only considered a two-level categorization of experience or operational proficiency (novice and expert), while in reality, pilot experience or operational proficiency encompasses levels beyond two categories.

Previous research has obtained expert detection abilities by comparing the behaviors of novices and experts, which is often used as guidance for training novices. While this approach may be applicable in most cases, it overlooks an important issue: even experienced expert pilots may not have an advantage when facing events such as system-malfunction failures that even designers cannot anticipate. Multiple accident cases have highlighted the limitations of relying solely on experience in such situations, where experienced captains have made erroneous actions [1,26,27]. Therefore, it is necessary to further explore the effects of failure types and operational experience on monitoring behaviors to deepen our understanding of behavioral performance in unexpected events and find solutions.

1.3. Present Study

The current study aims to address a gap in the existing literature by exploring the influence of failure type and operational proficiency on performance and monitoring behavior during unexpected events in flight tasks.

In particular, the following hypotheses are studied.

- (1) Higher levels of operational proficiency will result in faster resolution of failures. However, due to the unpredictable nature of system-malfunction failures, the facilitating effect of proficiency on resolution speed may not be as pronounced as in the case of system-limit failures.
- (2) Visual monitoring-related metrics, such as fixation duration, saccade duration, fixation rate, etc., will vary across different types of failures.

Based on the aforementioned content, this study presents a flight simulation experiment aimed at quantitatively comparing the differences in performance and monitoring behavior between automation system-limit and system-malfunction failures, considering different levels of operational proficiency. Due to the previous lack of distinction between system-limit and system-malfunction failures, this study specifically considers four levels of operational proficiency to examine whether the impact of proficiency level on performance and behavior is consistent across the two types of failures.

2. Materials and Methods

2.1. Participants

This study recruited 24 participants through an online platform by posting recruitment information. The number of participants was determined using Latin square design calcu-

lation. Considering the need for balance in the experimental order, the participants were divided into 2 groups, requiring a multiple of 12 (or 2) for the sample size. Additionally, reference was made to similar studies on simulator experiments, where a sample size of around 20 participants is commonly seen. Taking into account the impact of COVID-19 prevention and control measures, it was also challenging to gather a larger number of participants. Therefore, a final decision was made to select a sample size of 24 participants.

The participants' ages were 23.917 years on average ($SD = 1.754$). To meet the requirements of the experiment, these participants had a background in flight dynamics, flight control, or aircraft manipulation, and had prior exposure to flight simulators but no actual flying experience. Before participating in the experiment, they self-reported that they had normal hearing, vision (without glasses or wearing contact lenses), and color perception. They were in good physical condition and had not taken any medications for allergies, colds, or stomach discomfort in the past 48 h, nor had they used any psychotropic medications with anticholinergic properties. Their mental state was stable, and they had not engaged in behaviors such as staying up late or consuming alcohol that could affect their performance in the past 48 h. Prior to the experiment, the participants underwent vision and color perception tests to confirm their normal visual abilities and their ability to see the experimental interface and operate the instruments.

During the experiment, participants were instructed not to wear glasses and could only wear contact lenses. This was because the Tobii Glasses-2 eye tracker used in this experiment is a wearable eye-tracking device, and wearing framed glasses may cause errors in pupil calibration.

Each participant who completed the experiment received a compensation of 120 RMB.

2.2. Apparatus

This experiment utilized a range of equipment, including a flight simulator, a desktop computer with MATLAB, Simulink models, and Xplane software, as well as a Tobii Glasses-2 eye tracker. The flight simulator provided a simulated flight environment, while the desktop computer controlled the simulation using MATLAB, Simulink models, and Xplane software. The Tobii Glasses-2 eye tracker, along with a laptop computer running Tobii Pro Glasses Controller and Tobii Pro Lab software, was used to track and analyze participants' eye movements. The experiment also included a computer host and audio equipment, among other auxiliary devices. The exterior of the flight simulator was designed in a semi-circular shape, resembling the cockpit of an aircraft. These devices were crucial for creating an immersive flight simulation and collecting eye-tracking data for the purpose of examining participants' performance and behavior during the experiment.

The flight simulator used in this experiment is shown in Figure 1a. It had three degrees of freedom: pitch, roll, and yaw. It consisted of a pilot seat supported by hydraulic rods and a hydraulic drive unit. Additionally, there was a set of control devices including a control stick, throttle, and pedals, which were used for flight operation inputs. The control system of the flight simulator was designed to mimic the 737–800 aircraft, with the use of fly-by-wire flight control systems for the control stick, throttle control, and pedals to provide flight operation commands. During the experiment, the participant's control commands were transmitted to the computer-controlled simulation model through the control stick, throttle, and pedals to calculate the current flight attitude. Different hydraulic-driven rods were used to control the angle of the central pilot's seat, thereby simulating the three degrees of freedom of pitch, roll, and yaw. The control stick was equipped with buttons and toggle switches for additional functions. By using the set of control devices, participants were able to simulate various stages of flight operations, including takeoff, cruise, descent, and landing. They could also use the troubleshooting button on the control stick to resolve failures.

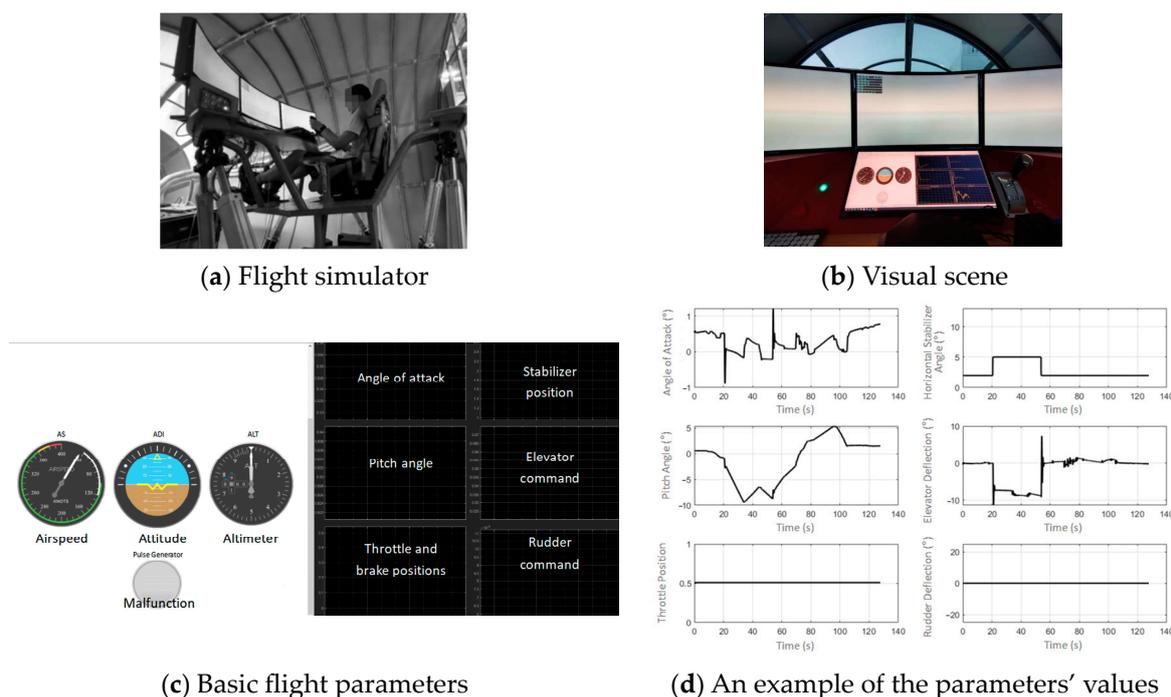


Figure 1. Flight simulator and visual scene.

The simulator was equipped with four environment display monitors with a resolution of 1920×1080 pixels, providing visual information for the simulated flight environment. Figure 1b illustrates the simulated pilot’s perspective and the corresponding visual scene. The display system consisted of three circular monitors positioned at eye level to simulate the external view, and one flat monitor tilted at a 45-degree angle below eye level to simulate instrument parameters. These monitors provided visual information to the pilot. The basic flight parameters were shown on the flat monitor (see Figure 1c). The left side of the flat monitor displayed flight instruments and malfunction indicators, including an airspeed indicator, attitude indicator, and altimeter. The bottom of the flat monitor displayed malfunction indicators. On the right side of the flat monitor, there was a parameter curve area that shows key parameters during the flight and parameters required for troubleshooting. Examples of the former include angle of attack, pitch angle, throttle, and brake positions, while examples of the latter include stabilizer position, elevator command, and rudder command. Figure 1d illustrates an example of the parameters’ values displayed on the parameter curve area during the flight in the experiment. The participants were able to observe the real-time display of these flight parameters during the flight.

The eye-tracking device used in this experiment was the Tobii Glasses-2 (refer to Figure 2), developed by Tobii, a Swedish company. It was a head-mounted, non-invasive eye-tracking device that utilized four cameras (two per eye) with a sampling rate of 100 Hz, as well as an ultra-wide-angle scene camera for eye-tracking. The device was designed with a head-mounted tracking module to ensure freedom of movement and accuracy of eye-tracking data. The scene camera had a field of view of 90 degrees and used a 16:9 aspect ratio. It was recorded at a horizontal angle of 82 degrees and a vertical angle of 52 degrees, with a resolution of 1920×1080 pixels and a frame rate of 25 frames per second.

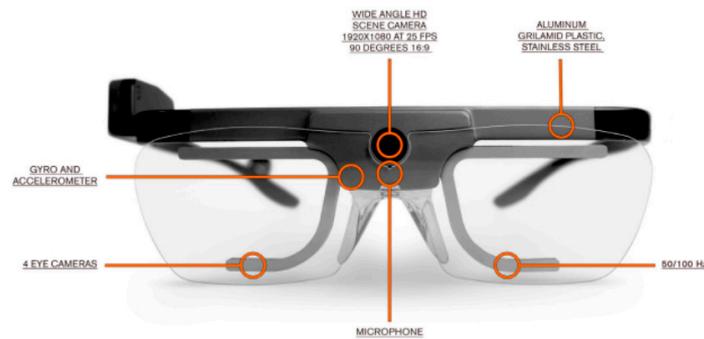


Figure 2. Tobii Glasses-2 eye-tracking instrument.

During the experiment, the simulated pilots (participants) controlled the flight simulator by providing command inputs to the control simulation model. The simulation model simulated the current aircraft attitude and position, calculated relevant parameters, and then outputted them to both the Xplane visual software and the simulator itself. This allowed the participants to experience visual scenes and perceive the simulated motion and displacement. Meanwhile, the Tobii Glasses-2 eye tracker captured the participants' eye movement data throughout the experimental session.

2.3. *Flighting Task*

Participants were required to simultaneously perform regular flight tasks and failure recovery tasks (detailed as follows). At the beginning of the experiment, the aircraft was in a cruising state at an altitude of 6000 m. The participants needed to control the aircraft to descend smoothly by 1000 m. If any failures were detected during the descent, the participants were required to resolve the failures and try to maintain the aircraft's heading, lateral motion (yaw and roll), and airspeed stability during the descent. Finally, the participants needed to level the aircraft: starting from the altitude of 5100 m during the descent and continuing until reaching the altitude of 5000 m.

For the failure recovery task, participants were required to identify the type of failure when they perceive a potential failure occurrence in the aircraft, and to press the corresponding button on the control stick to resolve the failure.

The following failure types were set:

1. System-limit failure: this included an error in the elevator sensor signal (pitch direction) and an error in the upper rudder sensor signal (yaw and roll directions).

When there was an error in the elevator sensor signal (pitch direction), the aircraft would immediately experience a pitch-down tendency. The elevator command signal would abruptly change to a non-zero constant value, causing the aircraft to descend in altitude and the pitch angle to decrease abruptly before gradually descending. The corresponding button to resolve this failure was button 1.

When there was an error in the upper rudder sensor signal (yaw and roll directions), the aircraft would immediately exhibit a side-slip tendency. The rudder command signal would abruptly change to a non-zero constant value. Due to the coupled effects of the aircraft's lateral and longitudinal dynamics, the aircraft would exhibit a yawing and rolling motion. The corresponding button to resolve this failure was button 2.

2. System-malfunction failure: This involved a failure in the horizontal stabilizer sensor signal (pitch direction). When this failure occurred, the horizontal stabilizer sensor signal transitioned from a normal signal to a non-zero value, causing the aircraft to pitch down. The corresponding button to resolve this failure was button 3.

The key difference between the system-limit and system-malfunction failures lies in whether the pilot can anticipate the failure in advance. Therefore, during pre-training, the participants were informed about the type of failure that would occur in the case of system-limit failure, including the possible timing, underlying principles, and resolve

methods of the failure. On the other hand, for system-malfunction failures, the participants were only informed that an unknown type of failure would occur, without knowledge of the possible scenarios, timing, and underlying principles of the failure, but they were aware of the resolve method. It should be noted that during the experiment, when facing a system-malfunction failure, although the participants lack self-knowledge and accurate cues about the failure, they could still perceive the failure through factors such as the visual environment, important parameter indicators, and the pilot's proprioception.

2.4. Procedure

The experimental procedure was as follows:

1. Recording Participant Information: Upon arrived at the laboratory, participants were requested to complete the necessary documentation, which included signing the consent form and providing personal information (name, age, gender, asymptomatic characteristics, etc.).
2. Preparing the Participants for the Experiment: The participants received training and familiarization with the flight simulator, including operational tests to ensure their proficiency in operating the failure-free simulator, indicating their readiness for the task. The experimental requirements, including participant tasks, types of failures, methods of resolving them, and safety precautions, were explained to the participants. These pre-training and experimental instructions were provided immediately before the commencement of the experimental trial.
3. Calibration of Eye-tracking Equipment and Flight Simulator: The eye-tracking equipment and the flight simulator were calibrated to ensure proper functioning.
4. Commencing the Experiment: Each participant completed 8 trials, with each trial lasting approximately 120 s. The experiment was conducted as a crossover design, considering both the types of system failures and the participants' proficiency levels. To control for the effects of the sequence of failure occurrence and the order of handling, the 24 participants were randomly assigned to two groups of equal size. In the first group, during each round of trials, they were first exposed to system-limit failures followed by system-malfunction failures; conversely, the second group experienced the reverse order, starting with system-malfunction failures followed by system-limit failures. During the experiment, the equipment collected response time, eye-tracking data, and flight data.
5. After each trial, the participants completed a training proficiency questionnaire to consolidate their skills. The training proficiency questionnaire primarily consisted of questions regarding participants' retrospective and self-assessment of their control over aircraft stability, handling of failures, and task competence. These questions aimed to reinforce the participants' acquired experience and skills in task performance and failure handling.
6. Upon completion of all the experiments, the participants were provided with compensation for their participation.

2.5. Data Analysis

The performance measure for response was represented by the response time, which refers to the time taken from the occurrence of the failure to the successful resolution of the failure by the participant.

The behavioral metrics selected for analysis included the representative indicators of pupil diameter, fixation duration, saccade duration, fixation rate, and saccade rate. Due to the short duration of each trial, the participants' attention was highly focused during the trial period, with their gaze primarily directed towards the screen and controls. Eye-tracking data were not divided into areas of interest (AOIs), but time of interest (TOI) was defined. The region of time interest was defined as the time period from when the failure occurred until it was resolved by the participant. For pupil diameter, the average value within the region of interest was calculated. Fixation duration and saccade duration

were determined by first identifying the categories of eye movement behavior and then averaging the duration of individual eye movement behaviors within each category. As for the fixation rate and saccade rate, eye movement behavior categories were identified, and then the ratio was calculated by dividing the total number of eye movement indicators in each category within the region of interest by the total duration of the region of interest.

The operational proficiency level in this study was represented by the number of experimental trials. A total of 24 participants completed 8 trials each, consisting of 4 trials of system-limit failures and 4 trials of system-malfunction failures. For the four encounters with system-limit failures, proficiency levels were assigned as Level 1 (inexperienced), Level 2 (low proficiency), Level 3 (moderate proficiency), and Level 4 (high proficiency). Similarly, for the four encounters with system-malfunction failures, proficiency levels were also assigned from Level 1 to Level 4. The use of experimental trials to represent proficiency levels took into account the participants' experience and skill accumulation during the task, as well as the improvement in proficiency with increasing trial numbers.

The data for the aforementioned measures were subjected to a repeated measures multifactor analysis of variance (ANOVA) to perform statistical analysis considering the variables of proficiency levels and types of failures.

3. Results

3.1. The Response Time at the Four Proficiency Levels

The response times for system-malfunction and system-limit failures at the four proficiency levels are shown in Table 1. For facilitating visual observation, Figure 3 illustrates the data of response time and SD from Table 1. In the figures and tables, "Malfun" stands for system-malfunction failure, while "Limit" stands for system-limit failure. "SD" stands for standard deviation. These notations will be used consistently throughout the subsequent figures and tables.

Table 1. The response time for system-malfunction and system-limit failures at four proficiency levels.

Proficiency Level	Failure Type	Response Time	SD	RML-RT ¹	MProf-RT ²	LProf-RT ³
1. Inexperienced	Malfun	39.194	22.106	3.644	1.671	1.133
	Limit	10.756	15.871	—	—	—
2. Low Proficiency	Malfun	23.450	24.754	2.471	1.437	1.514
	Limit	9.490	15.737	—	—	—
3. Moderate Proficiency	Malfun	16.317	18.241	2.604	1.089	1.900
	Limit	6.267	9.732	—	—	—
4. High Proficiency	Malfun	14.988	18.786	4.545	—	—
	Limit	3.298	1.404	—	—	—

¹ RML-RT stands for the ratio of system-malfunction failure response time to system-limit failure response time.

² MProf-RT stands for the ratio of response time between low proficiency level and the adjacent higher proficiency level in the case of system-malfunction failure. ³ LProf-RT stands for the ratio of response time between low proficiency level and the adjacent higher proficiency level in the case of system-limit failure.

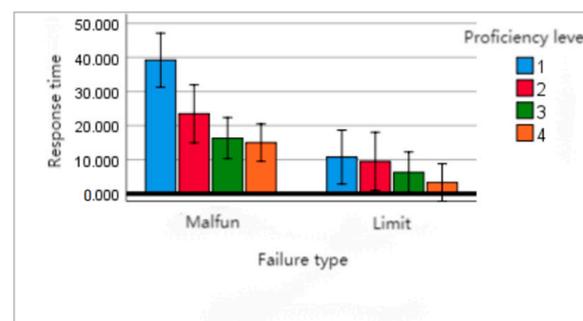


Figure 3. The response time for system-malfunction and system-limit failures at four proficiency levels.

The statistical analysis results showed that at various levels of proficiency, the response time for resolving failures under system-malfunction conditions was significantly longer than the response time under system-limit conditions ($p = 0.000 < 0.001$, $F = 18.138$). The ratio of response time between system-malfunction and system-limit failure, at proficiency levels 1 (inexperienced) to 4 (highly proficient), was 3.644, 2.471, 2.604, and 4.545, respectively. These ratios suggested that at lower proficiency levels (levels 1–3), the difference between response times for system-malfunction and system-limit failures remained relatively constant. However, at higher proficiency level (level 4), the gap between these two types of failure response times widens, indicating that highly proficient operators became more responsive as their understanding and proficiency in handling system-limit failures deepened, while showing less improvement in responding to system-malfunction failures.

Proficiency level had a significant impact on the response time for resolving failures under system-malfunction conditions ($p < 0.05$). As proficiency level increased, the response time for resolving failures significantly decreased. The ratio of response time between adjacent proficiency levels, from lower to higher, was 1.671, 1.437, and 1.089. This indicated that at lower proficiency levels, increasing proficiency greatly reduced the response time. However, when the proficiency level was already relatively high, the promoting effect on response time became less pronounced.

Proficiency level had a significant impact on the response time for resolving failures under system-limit conditions ($p < 0.05$). As proficiency level increased, the response time for resolving failures significantly decreased. The ratio of response time between adjacent proficiency levels, from lower to higher, showed a monotonic increase: 1.133, 1.514, and 1.900. This indicated that as proficiency level became higher, the beneficial promotion in reducing response time became more pronounced.

3.2. The Pupil Diameter at Four Proficiency Levels

Pupil diameter for system-malfunction and system-limit failures at four proficiency levels is shown in Table 2.

Table 2. The pupil diameter for system-malfunction and system-limit failures at four proficiency levels.

Proficiency Level	Failure Type	Pupil Diameter	SD
1. Inexperienced	Malfun	4.661	0.774
	Limit	4.776	0.637
2. Low Proficiency	Malfun	4.736	0.624
	Limit	4.823	0.682
3. Moderate Proficiency	Malfun	4.731	0.663
	Limit	4.737	0.595
4. High Proficiency	Malfun	4.758	0.752
	Limit	4.801	0.648

The statistical analysis results indicated that at various proficiency levels, the effect of failure type on pupil diameter was not significant ($p = 0.740 > 0.05$, $F = 0.111$). The proficiency level also did not have a significant impact on pupil diameter under system-malfunction or system-limit failure conditions ($p = 0.317 > 0.05$, $F = 18.138$).

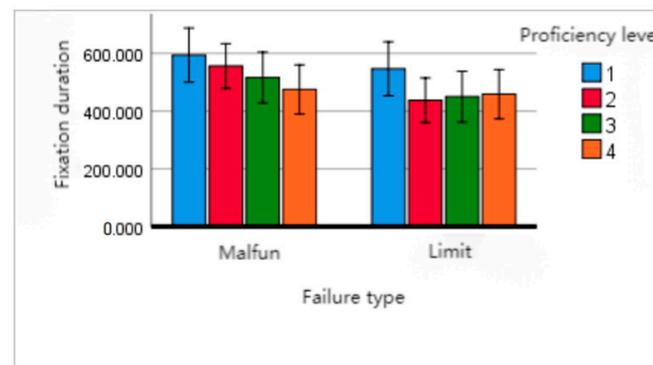
3.3. The Fixation Duration at the Four Proficiency Levels

The fixation duration for system-malfunction and system-limit failures at four proficiency levels is shown in Table 3. For facilitating visual observation, Figure 4 illustrates the data of fixation duration and SD from Table 3.

Table 3. The fixation duration for system-malfunction and system-limit failures at four proficiency levels.

Proficiency Level	Failure Type	Fixation Duration	SD	RML-FD ¹	MProf-FD ²	LProf-FD ³
1. Inexperienced	Malfun	594.078	206.020	1.087	1.068	1.249
	Limit	546.767	247.075	—	—	—
2. Low Proficiency	Malfun	556.101	183.085	1.270	1.077	0.972
	Limit	437.742	192.525	—	—	—
3. Moderate Proficiency	Malfun	516.341	171.139	1.147	1.086	0.981
	Limit	450.121	250.306	—	—	—
4. High Proficiency	Malfun	475.261	145.657	1.036	—	—
	Limit	458.761	255.310	—	—	—

¹ RML-FD stands for the ratio of system-malfunction failure fixation duration to system-limit failure fixation duration. ² MProf-FD stands for the ratio of fixation duration between low proficiency level and the adjacent higher proficiency level in the case of system-malfunction failure. ³ LProf-FD stands for the ratio of fixation duration between low proficiency level and the adjacent higher proficiency level in the case of system-limit failure.

**Figure 4.** The fixation duration for system-malfunction and system-limit failures at four proficiency levels.

The statistical analysis results indicated that at various proficiency levels, the impact of system-malfunction and system-limit failures on fixation duration was not significant ($p = 0.163 > 0.05$, $F = 2.006$). In terms of numerical values, the ratio of fixation duration between system-malfunction and system-limit failure, at proficiency levels 1 (inexperienced) to 4 (highly proficient), was 1.087, 1.270, 1.147, and 1.036, respectively. Across all proficiency levels, the fixation duration for system-malfunction failure was consistently higher than that for system-limit failure, although the difference was not statistically significant.

In the case of system-malfunction failures, proficiency level had a significant impact on the fixation duration required to resolve the failures ($p < 0.05$). As proficiency level increased, the fixation duration for resolving the failures significantly decreased. Specifically, at proficiency level 1, the fixation duration was significantly different from level 2 ($p < 0.05$), level 3 ($p < 0.001$), and level 4 ($p < 0.001$). The ratio of fixation duration between lower proficiency level and adjacent higher proficiency level showed a monotonic increase: 1.068, 1.077, 1.086. Fixation duration could reflect the difficulty of extracting information to some extent. The decrease in average fixation duration with increasing proficiency level suggested that operators faced reduced difficulty in extracting information.

In the case of system-limit failures, there was no consistent decreasing trend in fixation duration for resolving the failures as proficiency level increased. Instead, there was a significant decrease followed by a slight increase. The ratio of fixation duration between lower proficiency level and adjacent higher proficiency level was 1.249, 0.972, 0.981, respectively.

3.4. The Saccade Duration for the Four Proficiency Levels

The saccade duration for system-malfunction and system-limit failures at four proficiency levels is presented in Table 4. For facilitating visual observation, Figure 5 illustrates the data of saccade duration and SD from Table 4.

Table 4. The saccade duration for system-malfunction and system-limit failures at four proficiency levels.

Proficiency Level	Failure Type	Saccade Duration	SD	RML-SD ¹	MProf-SD ²	LProf-SD ³
1. Inexperienced	Malfun	39.070	6.470	1.038	1.041	1.018
	Limit	37.623	9.814	—	—	—
2. Low Proficiency	Malfun	37.548	4.870	1.016	1.030	1.067
	Limit	36.963	5.921	—	—	—
3. Moderate Proficiency	Malfun	36.451	5.561	1.052	1.015	1.039
	Limit	34.658	8.649	—	—	—
4. High Proficiency	Malfun	35.921	4.461	1.076	—	—
	Limit	33.370	7.383	—	—	—

¹ RML-SD stands for the ratio of system-malfunction failure saccade duration to system-limit failure saccade duration. ² MProf-SD stands for the ratio of saccade duration between low proficiency level and the adjacent higher proficiency level in the case of system-malfunction failure. ³ LProf-SD stands for the ratio of saccade duration between low proficiency level and the adjacent higher proficiency level in the case of system-limit failure.

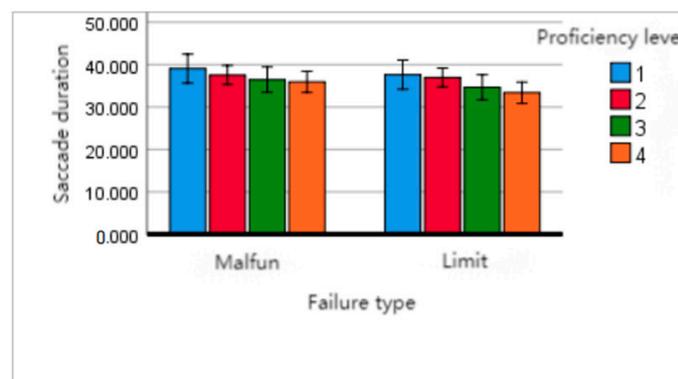


Figure 5. The saccade duration for system-malfunction and system-limit failures at four proficiency levels.

The statistical analysis results indicated that at various proficiency levels, the impact of system-malfunction and system-limit failures on saccade duration was not significant ($p = 0.244 > 0.05$, $F = 1.349$). In terms of numerical values, the ratio of saccade duration between system-malfunction and system-limit failures, at proficiency levels 1 (inexperienced) to 4 (highly proficient), was 1.038, 1.016, 1.052, and 1.076, respectively. Across all proficiency levels, the saccade duration for system-malfunction was consistently higher than that for system-limit, although the difference was not statistically significant.

In the case of system-malfunction failures, proficiency level had a significant impact on the saccade duration required to resolve the failures ($p < 0.05$). As proficiency level increased, the saccade duration for resolving the failures significantly decreased. The ratio of saccade duration between lower proficiency level and adjacent higher proficiency level showed a monotonic decrease: 1.041, 1.030, 1.015. Saccade duration could reflect the difficulty of extracting information to some extent. The decrease in saccade duration with increasing proficiency level suggested that operators faced reduced difficulty in visual search, enabling them to locate the visual area where the failure was present more quickly and to obtain the necessary information.

In the case of system-limit failures, as proficiency level increased, the saccade duration for resolving the failures significantly decreased ($p < 0.05$). The ratio of saccade duration

between lower proficiency level and adjacent higher proficiency level was 1.018, 1.067, 1.039, respectively.

3.5. The Fixation Rate for the Four Proficiency Levels

The fixation rate for system-malfunction and system-limit failures at four proficiency levels is presented in Table 5. For facilitating visual observation, Figure 6 illustrates the data of fixation rate and SD from Table 5.

Table 5. The fixation rate for system-malfunction and system-limit failures at four proficiency levels.

Proficiency Level	Failure Type	Fixation Rate	SD	RML-FR ¹	MProf-FR ²	LProf-FR ³
1. Inexperienced	Malfun	1.622	0.568	1.056	0.941	0.836
	Limit	1.536	0.623	—	—	—
2. Low Proficiency	Malfun	1.724	0.609	0.938	1.062	0.933
	Limit	1.838	0.659	—	—	—
3. Moderate Proficiency	Malfun	1.624	0.526	0.824	0.906	0.975
	Limit	1.971	0.726	—	—	—
4. High Proficiency	Malfun	1.793	0.657	0.887	—	—
	Limit	2.021	0.736	—	—	—

¹ RML-FR stands for the ratio of system-malfunction failure fixation rate to system-limit failure fixation rate. ² MProf-FR stands for the ratio of fixation rate between low proficiency level and the adjacent higher proficiency level in the case of system-malfunction failure. ³ LProf-FR stands for the ratio of fixation rate between low proficiency level and the adjacent higher proficiency level in the case of system-limit failure.

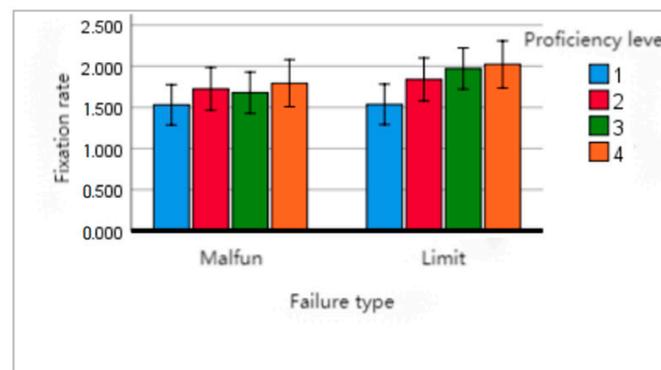


Figure 6. The fixation rate for system-malfunction and system-limit failures at four proficiency levels.

The statistical analysis results indicated that at various proficiency levels, the impact of system-malfunction and system-limit failures on fixation rate was not significant ($p = 0.290 > 0.05$, $F = 1.144$). In terms of numerical values, for all proficiency levels except level 1 (inexperienced), the fixation rate for system-malfunction failure was lower than that for system-limit failure, and the ratio between the two varied as follows: 1.056, 0.938, 0.824, and 0.887 at proficiency levels 1 to 4, respectively.

In the case of system-malfunction failures, proficiency level had a significant impact on the fixation rate required to resolve the failures ($p < 0.05$). As proficiency level increased, the fixation rate for resolving the failures exhibited a pattern of increase–decrease–increase. The ratio of fixation rate between lower proficiency level and adjacent higher proficiency level was 0.941, 1.062, 0.906, respectively.

In the case of system-limit failures, as proficiency level increased, the fixation rate for resolving the failures significantly increased ($p < 0.05$). The ratio of fixation rate between lower proficiency level and adjacent higher proficiency level showed a monotonic increase: 0.836, 0.933, 0.975, respectively.

The fixation rate increased with higher proficiency level, indicating that as proficiency increased, operators were able to attend to more points of interest within a given time. They could obtain effective information in shorter fixation durations and then shift their attention to the next area, exhibiting a more balanced distribution of attention.

3.6. The Saccade Rate for the Four Proficiency Levels

The saccade rate for system-malfunction and system-limit failures at four proficiency levels is presented in Table 6.

Table 6. The saccade rate for system-malfunction and system-limit failures at four proficiency levels.

Proficiency Level	Failure Type	Saccade Rate	SD
1. Inexperienced	Malfun	2.403	1.176
	Limit	2.267	1.401
2. Low Proficiency	Malfun	2.198	0.842
	Limit	2.990	1.901
3. Moderate Proficiency	Malfun	2.313	1.072
	Limit	2.869	1.436
4. High Proficiency	Malfun	2.331	1.239
	Limit	2.372	1.465

The statistical analysis results indicated that at various proficiency levels, the impact of failure types on saccade rate was not significant ($p = 0.226 > 0.05$, $F = 1.508$). In the case of system-malfunction or system-limit failures, proficiency level did not have a significant impact on the saccade rate ($p = 0.427 > 0.05$, $F = 0.931$).

4. Discussions and Conclusions

The purpose of this study is not to differentiate the differences in response performance and monitoring behavior between different failure types; rather it is to quantify these differences and consider the level of operational proficiency in order to assess the impact of failure types on response performance and monitoring behavior in different proficiency scenarios. It is hoped that researchers will pay attention to these differences, and we encourage them to consciously consider the proficiency levels of their participants and the types of system failures implemented in their studies.

Previous studies have found that individuals have faster response times to expected events compared to unexpected events [28]. This is because expected events can be anticipated and prepared for, whereas unexpected events cannot. However, some of the previous research comparing the takeover performance between system-malfunction failures and system-limit failures did not find significant differences [16]. Another study indicated that there were differences between the two, but the differences in takeover time were minimal (e.g., less than 1 s) [15]. In contrast, the quantitative analysis of response time differences in this study for the two types of failures revealed substantial disparities. Across proficiency levels 1 to 4, the response time for system-malfunction failures was 3.644, 2.471, 2.604, and 4.545 times longer than that for system-limit failures. Looking at the absolute values, at proficiency level 1, the response times for resolving system-malfunction failures and system-limit failures were 39.194 s and 10.756 s, respectively; at proficiency level 4, they were 14.988 s and 3.298 s, respectively. Therefore, this study confirms significant and substantial differences in response times between the two types of failures, highlighting the importance of not disregarding the influence of failure types on response performance. Furthermore, compared to driving a vehicle, the experimental task in this study simulated a complex flight mission. The complexity of the flight task itself and the manifestation of failures may contribute to longer response times and larger time differences. Therefore, it is possible that as tasks and failures become more complex, response times become longer, and the differences between system-malfunction failures and system-limit failures increase. This aspect needs to be further investigated in future research.

When unexpected events occur during flight, experienced pilots require shorter response times to make appropriate operations compared to novices. This has been confirmed by numerous studies [29–31]. However, these previous research articles usually focused on predictable situations (similar as system-limit failures), neglecting the consideration of unpredictable circumstances and the performance differences caused by different types of

failures at various proficiency levels. In addition to confirming the results consistent with previous studies, which indicate that participants' response times for resolving failures significantly decrease as proficiency increases, this study further reveals that the effect of proficiency levels on reducing response time varies across different types of failures. In the case of system-malfunction failure, increasing proficiency has a significant effect on reducing response time at lower proficiency levels, but its effect becomes less apparent at higher proficiency levels. Conversely, in the case of system-limit failures, higher proficiency levels have a greater facilitative effect on reducing response time. This suggests that proficiency levels had varying effects on the response time for resolving failures in different failure types, with higher proficiency being more effective in dealing with system-limit failures but less effective in addressing system-malfunction failures. Therefore, further research may be needed in order to explore the underlying mechanisms and factors influencing the effectiveness of higher proficiency in resolving system-malfunction failures.

Previous studies have extensively demonstrated that expert pilots exhibit more systematic and efficient eye movement patterns compared to intermediate and novice pilots [29,32,33]. This study further revealed that eye movement patterns vary in high proficiency (expert) participants depending on the type of failure. For instance, in the case of system-malfunction failures, the proficiency ratio of fixation duration monotonically increased, while in the case of system-limit failures, this ratio initially decreased and then increased (see Table 3). The trends in saccade duration and fixation rate varied in system-failure and system-limit failure as well (see Tables 4 and 5). This inconsistent pattern of proficiency ratio changes suggested differences in eye movement patterns, attention allocation and monitoring behavior of participants at different proficiency levels for system-malfunction and system-limit failures. It also implied that different cognitive and coping strategies may have been employed for different types of failures. The existence of such inconsistencies provided clues for further investigating the cognitive and coping strategies for different types of failures, aiming to enhance pilots' understanding and response capabilities to automation system failures.

Given the substantial variations in response performance and monitoring behavior across different proficiency levels, depending on the type of failure, it is important for the literature to distinguish between system-malfunction and system-limit failures and not overlook the influence of proficiency.

The findings of this study not only emphasize the importance of distinguishing between system-malfunction and system-limit failures but also support the significance of pilots' mental models of automated systems and the improvement of pilot training. System-limit failures are predictable because pilots receive training and have knowledge of specific system failure scenarios. However, if they have not received sufficient training or their proficiency level is not adequate, system-limit failures may become difficult and unpredictable for them. Additionally, there are system-malfunction failure conditions that even the designers themselves may not have anticipated, making the situations even more unpredictable. Practitioners responsible for system operations may encounter an increasing number of abnormal events, sometimes even unprecedented ones. Research findings on how to predict, respond to, and train for rare, uncertain, and unexpected events will help them succeed in these challenging and difficult processes [34]. The results of this study also support this perspective. Therefore, in addition to recognizing the distinction between system-malfunction and system-limit failure, it is crucial to focus on the development of pilots' mental models of automated systems and training theories, fostering their core competencies and enhancing their overall abilities to deal with complexity, uncertainty, and unknown situations. This ensures that they can effectively cope with system failures when they occur, and it is of paramount importance to enhancing the quality of pilot training and improving the safety design of aircraft automation systems.

One limitation of this study is its relatively small sample size, consisting of only 24 participants. It is important to acknowledge that this sample may not fully represent the diverse population of commercial aircraft pilots. However, despite being a small sample, the results of this study still contribute valuable information and insights. The

findings may serve as a foundation for further studies in this field, allowing for more in-depth investigations with larger and more diverse samples. Additionally, the information provided by this sample can be utilized for exploratory research and provide preliminary evidence. While caution should be exercised in generalizing these findings, this study still offers valuable insights into the topic and serves as a stepping stone for future research in this area.

Author Contributions: J.D.: conceptualization, methodology, resources, writing—original draft preparation, supervision, funding acquisition. P.Y.: conceptualization, investigation, writing—review and editing, visualization. S.H.: conceptualization, methodology, software, validation, formal analysis, data curation. P.K.: conceptualization, methodology, resources. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China (71601007) and the Aviation Science Foundation of China (*51003).

Informed Consent Statement: Informed consent was obtained from all participants involved in the study.

Data Availability Statement: The authors are unable or have chosen not to specify which data has been used.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. AAIBE. *Aircraft Accident Investigation Preliminary Report: B737-8 (MAX) Registered ET-AVJ*; Report No. AI-01/19; AAIBE: Addis Ababa, Ethiopia, 2019.
2. National Transportation Safety Board. *Accident Report NTSB/AAR-14/01 PB2014-105984, Descent below Visual Glidepath and Impact with Seawall Asiana Airlines Flight 214 Boeing 777-200ER*; National Transportation Safety Board: Washington, DC, USA, 2014.
3. Australian Transport Safety Bureau. *ATSB Transport Safety Report, Aviation Occurrence Investigation AO-2008-070, In-Flight Upset 154 km West of Learmonth, WA 7 October 2008 VH-QPA Airbus A330-303*; Australian Transport Safety Bureau: Canberra, Australia, 2011.
4. FAA. *Code of Federal Regulations Title 14, Section 91.3—Responsibility and Authority of the Pilot in Command*; FAA: Washington, DC, USA, 2023.
5. Holford, W.D. An Ethical Inquiry of the Effect of Cockpit Automation on the Responsibilities of Airline Pilots: Dissonance or Meaningful Control? *J. Bus. Ethics* **2022**, *176*, 141–157. [[CrossRef](#)]
6. Martin, W.L.; Murray, P.S.; Bates, P.R. What would you do if...? Improving pilot performance during unexpected events through in-flight scenario discussions. *Aeronautica* **2011**, *1*, 8–22.
7. Wickens, C.D.; Hoey, B.L.; Gore, B.F.; Sebok, A.; Koenecke, C.; Salud, E. Predicting pilot performance in off-nominal conditions: A meta-analysis and model validation. In Proceedings of the 53rd Annual Meeting of the Human Factors and Ergonomics Society, San Antonio, TX, USA, 19–23 October 2009; SAGE Publications: Thousand Oaks, CA, USA, 2009; pp. 86–90.
8. Sarter, N.B.; Woods, D.D.; Billings, C.E. Automation surprises. In *Handbook of Human Factors and Ergonomics*, 2nd ed.; Salvendy, G., Ed.; Wiley: New York, NY, USA, 1997; pp. 1926–1943.
9. Kochan, J.A.; Breiter, E.G.; Jentsch, F. Surprise and unexpectedness in flying: Database reviews and analyses. In Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society, New Orleans, LA, USA, 20–24 September 2004; HFES: Santa Monica, CA, USA; SAGE Publications: Thousand Oaks, CA, USA, 2004; pp. 335–339.
10. Sarter, N.B. Investigating mode errors on automated flight decks: Illustrating the problem-driven, cumulative, and interdisciplinary nature of human factors research. *Hum. Factors J. Hum. Factors Ergon. Soc.* **2008**, *50*, 506–510. [[CrossRef](#)]
11. Landman, A.; Groen, E.L.; Van Paassen, M.M.R.; Bronkhorst, A.W.; Mulder, M. Dealing with Unexpected Events on the Flight Deck: A Conceptual Model of Startle and Surprise. *Hum. Factors* **2017**, *59*, 1161–1172. [[CrossRef](#)]
12. Kinney, L.; O'Hare, D. Responding to an Unexpected In-Flight Event: Physiological Arousal, Information Processing, and Performance. *Hum. Factors* **2020**, *62*, 737–750. [[CrossRef](#)]
13. Xing, G.; Sun, Y.; He, F.; Wei, P.; Wu, S.; Ren, H.; Chen, Z. Analysis of Human Factors in Typical Accident Tests of Certain Type Flight Simulator. *Sustainability* **2023**, *15*, 2791. [[CrossRef](#)]
14. Seppelt, B.D.; Lee, J.D. Keeping the Driver in the Loop: Dynamic Feedback to Support Appropriate Use of Imperfect Vehicle Control Automation. *Int. J. Hum.-Comput. Stud.* **2019**, *125*, 66–80. [[CrossRef](#)]
15. DeGuzman, C.A.; Hopkins, S.A.; Donmez, B. Driver Takeover Performance and Monitoring Behavior with Driving Automation at System-Limit versus System-Malfunction Failures. *Transp. Res. Rec.* **2020**, *2674*, 140–151. [[CrossRef](#)]
16. Dogan, E.; Rahal, M.C.; Deborne, R.; Delhomme, P.; Kemeny, A.; Perrin, J. Transition of Control in a Partially Automated Vehicle: Effects of Anticipation and Non-Driving-Related Task Involvement. *Transp. Res. Part F Traffic Psychol. Behav.* **2017**, *46*, 205–215. [[CrossRef](#)]

17. Lee, J.; Abe, G.; Sato, K.; Itoh, M. Developing human-machine trust: Impacts of prior instruction and automation failure on driver trust in partially automated vehicles. *Transp. Res. Part F Traffic Psychol. Behav.* **2021**, *81*, 384–395. [[CrossRef](#)]
18. Lee, J.; Abe, G.; Sato, K.; Itoh, M. Impacts of System Transparency and System Failure on Driver Trust During Partially Automated Driving. In Proceedings of the 2020 IEEE International Conference on Human-Machine Systems (ICHMS), Rome, Italy, 7–9 September 2020; pp. 1–3.
19. Mishler, S.; Chen, J. Effect of automation failure type on trust development in driving automation systems. *Appl. Ergon.* **2023**, *106*, 103913. [[CrossRef](#)]
20. Ephrath, A.R.; Curry, R.E. Detection by pilots of system failures during instrument landings. *IEEE Trans. Syst. Man Cybern.* **1977**, *7*, 841–848. [[CrossRef](#)]
21. Salmon, P.M.; Walker, G.H.; Stanton, N.A. Pilot error versus sociotechnical systems failure: A distributed situation awareness analysis of Air France 447. *Theor. Issues Ergon. Sci.* **2016**, *17*, 64–79. [[CrossRef](#)]
22. Kramer, L.J.; Etherington, T.J.; Bailey, R.E.; Kennedy, K.D. Quantifying pilot contribution to flight safety during hydraulic systems failure. In *Advances in Human Aspects of Transportation, Proceedings of the AHFE 2017 International Conference on Human Factors in Transportation, The Westin Bonaventure Hotel, Los Angeles, CA, USA, 17–21 July 2017*; Springer International Publishing: Cham, Switzerland, 2018; pp. 15–26.
23. Schriver, A.T.; Morrow, D.G.; Wickens, C.D.; Talleur, D.A. Expertise differences in attentional strategies related to pilot decision making. *Hum. Factors* **2008**, *50*, 864–878. [[CrossRef](#)]
24. Russi-Vigoya, M.N.; Patterson, P. Analysis of Pilot Eye Behavior during Glass Cockpit Simulations. *Procedia Manuf.* **2015**, *3*, 5028–5035. [[CrossRef](#)]
25. Ziv, G. Gaze Behavior and Visual Attention: A Review of Eye Tracking Studies in Aviation. *Int. J. Aviat. Psychol.* **2016**, *26*, 75–104. [[CrossRef](#)]
26. Junta Investigadora de Accidentes Aereos (JIAA). *Aviation Accident Report: Birgenair Flight ALW-301*; Junta Investigadora de Accidentes Aereos (JIAA): Caracas, Venezuela, 1996.
27. Air Line Pilots Association; Engineering and Air Safety. *Aircraft Accident Report: Accident Reconstruction by Evaluation of Injury Patterns*; Division of Aerospace Pathology: Washington, DC, USA, 1977.
28. Green, M. How Long Does It Take to Stop? Methodological Analysis of Driver Perception-Brake Times. *Transp. Hum. Factors* **2000**, *2*, 195–216. [[CrossRef](#)]
29. Bruder, C.; Hasse, C. Differences between experts and novices in the monitoring of automated systems. *Int. J. Ind. Ergon.* **2019**, *72*, 1–11. [[CrossRef](#)]
30. Santos, S.; Parraca, J.A.; Fernandes, O.; Villafaina, S.; Clemente-Suarez, V.J.; Melo, F. The Effect of Expertise during Simulated Flight Emergencies on the Autonomic Response and Operative Performance in Military Pilots. *Int. J. Environ. Res. Public Health* **2022**, *19*, 9141. [[CrossRef](#)]
31. Zheng, Y.; Lu, Y.; Jie, Y.; Zhao, Z.; Fu, S. Test Pilot and Airline Pilot Differences in Facing Unexpected Events. *Aerosp. Med. Hum. Perform.* **2023**, *94*, 18–24. [[CrossRef](#)]
32. Hua, Y.; Fu, S.; Lu, Y. The Effect of Pilots' Expertise on Eye Movement and Scan Patterns During Simulated Flight Tasks. In *International Conference on Human-Computer Interaction*; Springer International Publishing: Cham, Switzerland, 2022; Volume 13307, pp. 290–299.
33. Harris, D.J.; Arthur, T.; de Burgh, T.; Duxbury, M.; Lockett-Kirk, R.; McBarnett, W.; Vine, S.J. Assessing expertise using eye tracking in a Virtual Reality flight simulation. *Int. J. Aerosp. Psychol.* **2023**, 1–21. [[CrossRef](#)]
34. Hancock, P.A. Reacting and responding to rare, uncertain and unprecedented events. *Ergonomics* **2023**, *66*, 454–478. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.