

Relationships Between Pilots' Startle and Surprise Responses and Information-Processing Performance During Simulated In-Flight Events

Human Factors
2026, Vol. 0(0) 1–23
© 2026 Human Factors
and Ergonomics Society



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/00187208261434426
journals.sagepub.com/home/hfs



Jiayu Chen^{1,2} , Annemarie Landman^{1,3} , Alexis Derumigny¹ ,
Olaf Stroosma¹ , M. M. (René) van Paassen¹ , and Max Mulder¹ 

Abstract

Objective: We aim to investigate how pilots' startle and surprise responses affect information-processing performance during simulated in-flight events.

Background: Startle and surprise are distinct constructs, each with their own potential effects on pilot's performance during unexpected in-flight events. Prior research suggests that startle may impair performance through stress-induced cognitive interference, whereas surprise may do so via cognitive demands associated with sensemaking. Thus, we hypothesized that both startle and surprise would negatively affect information-processing performance on a secondary auditory cognitive task.

Method: Using a motion-based hexapod simulator and a twin-propeller aircraft model, 26 pilots each performed eight single-pilot flight scenarios, which were designed to elicit varying levels of startle and surprise responses. Linear mixed-effects models were employed to analyse the relationships between self-report startle and surprise with secondary task performance, while controlling for individual differences and differences between the scenarios.

Results: The results revealed that higher startle was significantly associated with reduced information-processing speed. For surprise, no significant association was found.

Conclusion: The findings suggest that, within the context of the tested scenarios, startle appeared to impose a more pronounced disruptive effect on pilots' information-processing performance than surprise.

Application: The study underscores the need for tailored interventions to enhance pilots' resilience to startle and calls for further research on ecologically valid methods to induce surprise for research and training purposes.

¹Delft University of Technology, The Netherlands

²China Automotive Engineering Research Institute Co., Ltd, China

³TNO, The Netherlands

Corresponding Author:

Jiayu Chen, Faculty of Aerospace Engineering, TU Delft, Kluyverweg 1, Delft 2629 HS, The Netherlands; e-mail: Jiayu.ChenJade@outlook.com

Keywords

aviation, cognition, stress, working memory, incapacitation

Introduction

During non-normal in-flight events, high workload and pressure may pose severe challenges to pilot cognitive functioning. Cognitive performance impairments such as attentional tunnelling, break-down of teamwork, hasty decision making, or automation mode confusion have been described (De Boer & Dekker, 2017; Parasuraman et al., 2000; Sarter et al., 2017; Wickens, 2008; Young & Stanton, 2002). These challenges may be exacerbated when pilots experience startle and surprise (Casner et al., 2013). Recognizing this impact, the European Union Aviation Safety Agency and Federal Aviation Administration have integrated startle and surprise into the regulatory framework for Upset Prevention and Recovery Training (UPRT) to mitigate risks (European Aviation Safety Agency, 2015; Federal Aviation Administration, 2015).

Startle refers to a coinciding emotional and physiological response to abrupt, intense stimuli perceived as potentially threatening. It is characterized by involuntary physiological startle reflexes and generalized stress responses (Dreissen et al., 2012; Holand et al., 1999; Koch, 1999; Martin et al., 2015). The startle reflex, typically within 100 milliseconds after a stimulus, includes eye-blink, head movement, facial grimacing, shoulder elevation, arm abduction, elbow bending, forearm pronation, finger flexion, and abdominal contraction (Ladd et al., 2000; Leuchs et al., 2019). If the threat persists, it triggers a generalized stress response involving activation of the autonomic nervous system and the release of cortisol (Martin et al., 2015). This leads to rapid breathing, increased heart rate, systolic blood pressure, pupil dilation, and amplified sensory arousal (Dreissen et al., 2012; Holand et al., 1999; Jansen et al., 1995; Papadimitriou & Priftis, 2009).

Research has shown that startle and resulting stress can temporarily disrupt cognitive

processes, impair sustained attention, and reduce cognitive efficiency (Sehlström et al., 2022; Valls-Sole et al., 2008). Neurobiological models indicate that startle activates survival circuits, prioritizing threat detection at the expense of cognitive processing, thereby interfering with executive function (LeDoux, 2012). This disruption could limit working memory and attentional control by diverting cognitive resources toward stimuli-driven processing, thereby compromising goal-directed behaviour (Eysenck et al., 2007). This overload may narrow attention, reduce situation awareness, impair decision making and information-processing capacity (Casner et al., 2013; Dehais et al., 2013). The severity and duration of these effects vary depending on individual differences (Bradford et al., 2014), contexts, and stimuli (Lang et al., 1990). The disruptions are particularly concerning in high-risk operational domains such as aviation, where precise and timely decision making is critical (Landman et al., 2017; Rivera et al., 2014). Empirical evidence demonstrates that unexpected in-flight events can elicit elevated physiological arousal (e.g. increased heart rate and pupil dilation), impair processing of critical information, and degrade overall performance (Kinney & O'Hare, 2020; Martin et al., 2016). While several studies have documented transient cognitive impairments following startle (Deniel et al., 2024), others have reported minimal effects or even performance enhancements under high cognitive load conditions (Schwartz et al., 2025). These mixed findings underscore the importance to investigate the effect of startle within highly ecologically valid contexts.

Surprise is a cognitive and affective response triggered by unexpected, schema-discrepant events that are (momentarily) difficult to explain (Horstmann, 2006; Meyer et al., 1997). Surprise interrupts ongoing automatic cognitive processes, directing attention

towards analysing unexpected events, and updating one's understanding of situations (Reisenzein et al., 2019). This interruptive effect has been quantified with the latency of verbal and motor response tasks in labs (Wessel & Aron, 2017). The sensemaking and "reframing" processes require effortful, goal-directed cognitive processing (Klein et al., 2007; Landman et al., 2017). Goal-directed processes that are required for sensemaking are susceptible to disruption under high stress, which could hinder the acquisition of relevant information and the execution of appropriate actions (Eysenck et al., 2007; Landman et al., 2017).

Thus, although startle and surprise are theoretically and psychometrically distinct constructs (Chen et al., 2025a, 2025b), both startle and surprise may impair pilots' information-processing performance by disrupting ongoing cognitive processes. Startle could induce additional mental workload by diverting attention to the startling stimuli, or due to rapid engagement of survival circuits that prioritize a quick response to threat over executive functioning (LeDoux, 2012). Surprise could impair information-processing by imposing additional demands on working memory to make sense of the surprising events.

Therefore, we aim to investigate the relationships between subjective startle and surprise severity on pilots' information-processing performance. Building upon research performed in lab settings (Duchevet et al., 2025), the approach will be to use a more ecologically valid context consisting of single-pilot basic flying tasks performed in a motion-based simulator. Using a dual-task paradigm (Wickens, 2002), information-processing performance is

quantified using a secondary auditory cognitive task, that was performed concurrently with in-flight events. An auditory task was chosen to not interfere with visual information processing.

We hypothesize that both startle and surprise would impair information-processing performance, reducing secondary task performance. To test the hypotheses, pilots perform eight flight scenarios designed to elicit a wide range of startle and surprise responses. The objective is to test whether higher levels of startle or surprise are associated with poorer secondary task performance, while controlling for scenario differences and general individual differences in performance. To measure startle and surprise, the developed and validated multi-item Startle Inventory and Surprise Inventory are employed, respectively (Chen et al., 2025a, 2025b). Insights into the effects of startle and surprise on information processing in a context that is relatively high in ecological validity may inform the development of training programs aimed at enhancing cognitive resilience during in-flight emergencies.

Method

Participants

A total of 26 professional pilots (25 males, 1 female) participated in the study. Among them, 18 held a full Airline Transport Pilot License (ATPL), and 8 held a frozen ATPL, indicating they had completed all theoretical exams under the EASA licensing framework but had not yet accrued the required 1,500 flight hours for issuance of the full license. In terms of current professional roles, the participants

Table 1. Characteristics of the participants

	Mean	SD	Min	Max
Age (yrs)	43.8	13.0	23.0	67.0
Employed time (yrs)	17.8	13.2	0.5	44.0
Flight hours	8633.9	7082.1	280.0	25500.0
Large aircraft	6566.1	6607.8	0.0	22000.0
Business jet	1257.7	2803.7	0.0	10000.0
Small aircraft	810.1	1258.4	0.0	5000.0

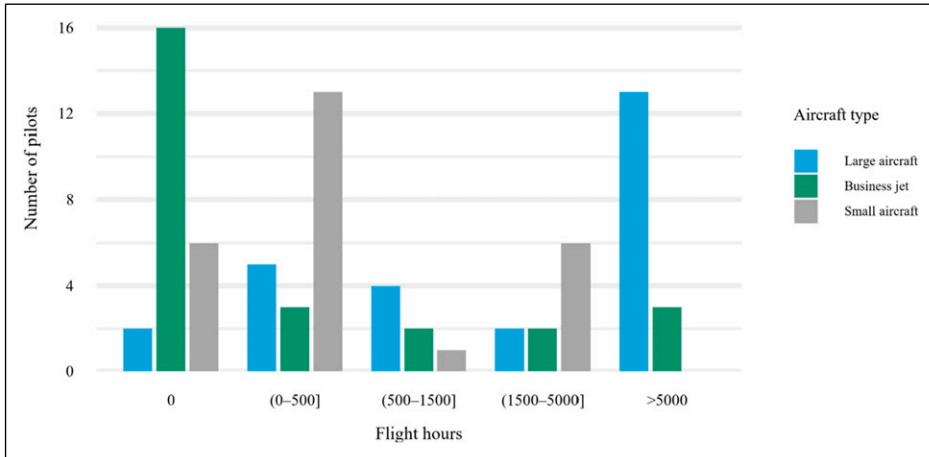


Figure 1. Distribution of pilots' flight hours across aircraft types (N = 26). Each pilot may have experience in multiple aircraft types. No pilot reported > 5000 flight hours on small aircraft

included 14 captains, 8 first officers, 3 second officers, and 1 pilot employed in a non-airline aviation position.

The characteristics of participants are summarized in Table 1. Figure 1 illustrates the distribution of flight hours across different aircraft types (N = 26). All participants provided informed consents. This research complied with the American Psychological Association Code

of Ethics and the Research Ethics Committee of the Delft University of Technology approved the research design (No. 4056).

Apparatus

The experiment was performed using the SIMONA Research Simulator at the Delft University of Technology (Figure 2). This is a full-



Figure 2. The SIMONA research simulator



Figure 3. The experimental setup (simulated flight deck)

motion simulator equipped with a hydraulic hexapod platform providing six degrees of freedom. The simulator has a collimated 180 degrees horizontal by 40 degrees vertical field of view for outside vision rendered with FlightGear. A 5.1 surround sound system was installed for realistic 3D sound effects of potential startling or surprising events, alarms, flaps, retractable gear, aerodynamic noise, ground rumble, and engines. During the experiment, participants wore single-ear intercom headsets.

The experiment employed an aerodynamic model of the Piper PA-34 Seneca III, a light twin-propeller aircraft. The flight deck (see Figure 3 for daytime and night settings) featured flight controls, including a control column with pitch trim, rudder pedals with force feedback, throttle, gear, and flaps with three settings: 0° (UP), 25°, and 40° (LAND). The

avionics consisted of a primary flight display (PFD) similar to a G1000 PFD, a backup primary flight display, and a multifunction display for engine, configuration, and navigation data. Information on airspeed, altitude, attitude, engine parameters, flaps position, and gear status was available via the avionics displays.

General Procedure

An overview of the experimental procedure is illustrated in Figure 4. Pilots performed tasks on a single day and as single-pilot crew. The total duration of the experiment per participant was approximately 2 hours, including the briefing, familiarization, test session (comprising eight test scenarios), and debriefing.

All pilots were briefed about the aircraft model, simulator features, experimental tasks,

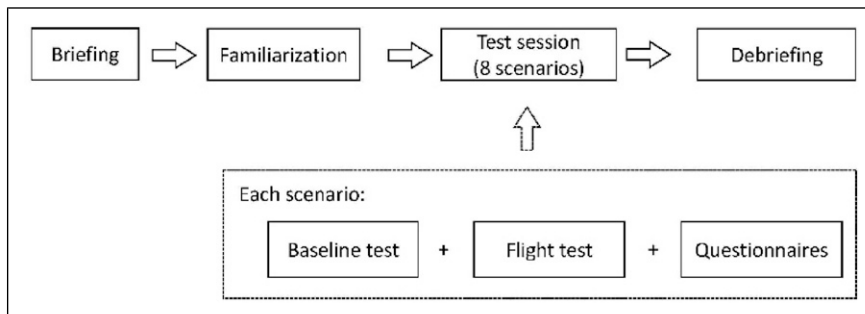


Figure 4. The experimental procedure

and definitions of startle and surprise. Each familiarization and test scenario required the pilot to fly (part of) a left-handed traffic circuit (Figure 5) for runway 18C, Schiphol Airport (EHAM). The circuit would need to be flown at 1,000 ft with a speed of 115 kt. A flaps setting of 0 (UP) was required for take-off, 25 during base leg, and 40 (LAND) in final. The circuit required rotate speed of 80 kt, minimum control speed of 80 kt, best rate of climb speed of 92 kt, and landing approach speed of 90 kt. These configurations were also available on keypad in the simulator.

During familiarization, pilots also practiced the secondary auditory task (see the Secondary Auditory Task section) once on the runway, and once while performing the circuit. At the end of the familiarization, all pilots confirmed that they could handle the aircraft model, none required help in determining the turn points of the circuit, and none had difficulty with performing the secondary auditory task.

Pilots then proceeded with the test session, which consisted of eight test scenarios (see the Startle and Surprise Events section), presented in a semi-counterbalanced order defined by a Latin square (Hinkelmann & Kempthorne, 2007). Test scenarios began either from the take-off position on runway 18C, or in-flight position at 800 ft ahead of runway 18C, with an airspeed of 99 kt. In all

cases, participants were required to complete the circuit and land safely on the same runway.

Before each scenario, participants received a briefing on wind strength, direction, and weather code through a Meteorological Aerodrome Report (METAR). Before and during each scenario, pilots performed blocks of the auditory task (see the Secondary Auditory Task section). Immediately following each scenario, pilots completed a questionnaire that included the Startle and Surprise Inventories (see the Dependent Measures section). After completing all scenarios, pilots were informed about all simulated events in a debriefing.

Startle and Surprise Events

Test scenarios were designed to induce relatively high surprise or high startle, both, or neither, based on the characteristics of the preset events. This was done to maximize variation between scenarios in the pilots' responses, which is necessary for investigating relationships between the responses and information-processing performance. Events that were rare, unfamiliar, or difficult to immediately explain were used to induce relatively high surprise (Izard et al., 1993; Klein et al., 2007; Landman et al., 2017; Meyer et al., 1997). Events that were sudden, loud, or

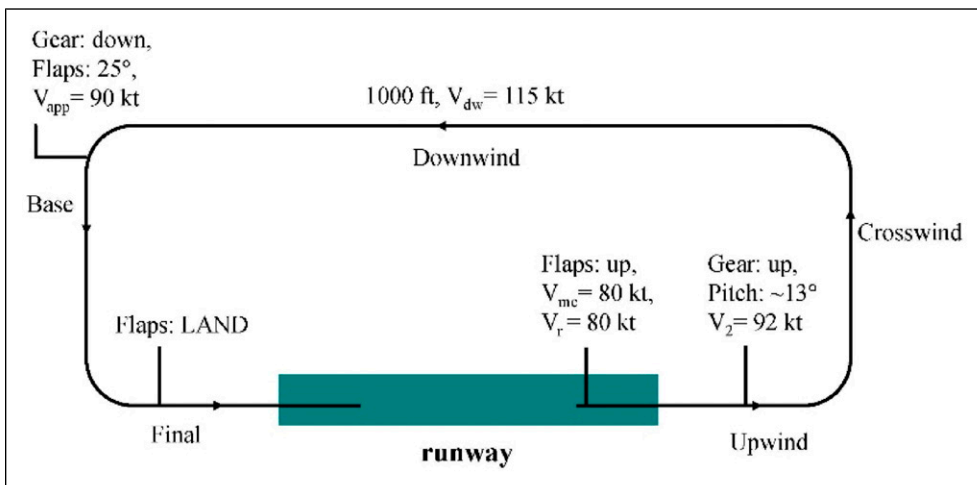


Figure 5. The standard traffic circuit with target settings

immediately threatening were used to induce relatively high startle (Blumenthal, 1988; Bradley et al., 2005; Koch, 1999; Martin et al., 2015). More “high surprise” than “high startle” scenarios were included because we expected difficulty with surprising a sufficient proportion of pilots to induce sufficient variation for analysis. The characteristics of the events and secondary auditory tasks are listed in Table 2.

The following three scenarios required pilots to manually respond to controllability issues. In the engine failure scenario (ENF), the right engine failed during climb after reaching 900 ft, causing a roll and yaw moment that could be counteracted using the column and pedals. The pilots’ unfamiliarity with the specific aircraft response to the engine failure was expected to induce surprise. Since there was no immediate threat nor intense stimuli, a limited startle response was expected. In the flaps asymmetry scenario (FLAP), the left flap remained in the UP position when pilots selected Flaps 25 during the turn to base leg. This caused

an unexpected roll and yaw moment, which need to be counteracted using the control column. The pilots’ unfamiliarity with the specific aircraft response to the flap failure was expected to induce surprise. Since there was no immediate threat nor intense stimuli, a limited startle response was expected. In the cargo shift scenario (CARGO), a simulated piece of heavy cargo broke loose and shifted towards the tail after take-off, with a loud scraping and collision noise coming from the back of the aircraft. This event temporarily moved the aircraft’s centre of gravity backward, resulting in a violent pitch-up motion that pilots had to correct using the control column. The novelty of the event and difficulty in explaining it were expected to induce high levels of surprise, while the sudden (upset) motion of the aircraft and the accompanying loud scraping noise were expected to induce startle response.

The following scenarios required no intervention from pilots and featured no imposed changes to simulator or aircraft motions. The

Table 2. The characteristics of event and secondary auditory task in each test scenario

ID	Event description	Event inserted ^a	Lead time ^b	Event timing	Non-target (distraction) block timing
ENF	The right engine failed shortly after take-off.	Right engine failure during take-off	10 s	5 s after reaching 900 ft	Climbing through 100 ft
FLAP	When selecting Flaps 25 in the base leg, the left flap remained UP.	The aircraft response when you selected Flap 25	5 s	At Flaps 25 selection	Climbing through 800 ft
CARGO	Cargo moved backward during take-off with a loud sound and pitch-up motion.	The pitch-up motion (and noise) after rotation	15 s	10 s after reaching 200 ft	-
LTS	While flying at night, lightning struck the aircraft with a bright flash and loud thunder sound.	The lightning strike	5 s	5 s after descending to 500 ft	Descending through 800 ft
PFD	The PFD turned black.	The malfunction of the PFD	5 s	Descending to 600 ft	-
STALL	A bird strike triggered a false stall alarm with stick shaker.	The stall alarm	5 s	20 s after reaching 800 ft	Climbing through 150 ft
NTO	Normal take-off without preset malfunction.	The level-off manoeuvre	-	-	-
NLO	Normal landing without preset malfunction.	The effect of crosswind	-	-	-

Note. PFD = primary flight display.

^aThe specific content that replaced the placeholder ‘[the stimulus]’ in the Startle and Surprise Inventories.

^bLead time as in Figure 6b.

lightning strike scenario (LTS) started at the in-flight position in the night condition, with weather report indicating the presence of thunderstorms. The scenario was designed to be highly startling due to the sudden bright flash and loud thunder sound, but not (limited) surprising due to the stated weather conditions. Between 400 and 500 ft, a lightning strike was simulated (Chen et al., 2025b), accompanied by a thunder sound played over a surround sound system, and a strobe light flash presented via the simulator's projection system.

The second no-intervention scenario was the primary flight display failure scenario (PFDf). Due to its unfamiliar nature, this scenario was intended to elicit a high level of surprise. However, in the absence of intense stimuli or immediate threat, a limited startle response was expected. It started at in-flight position. The PFD turned to black at 600 ft. Pilots could use the outside view and the backup display to continue landing. The third no-intervention scenario was the false stall warning scenario (STALL). After take-off, a bird struck the angle of attack at 800 ft, triggering a continuous false stall alarm. Due to the lack of context for a stall event, it was expected to be surprising, and due to the sudden loud auditory stall alarm and stick shaker it was expected to be startling.

Finally, two more scenarios, normal take-off (NTO) and normal landing (NLO), were

included to present events that were expected to induce low levels of startle and surprise. NTO started at the take-off position and NLO started at the in-flight position, with pilots performing a landing under a 5 kt crosswind from the east.

Secondary Auditory Task

The pilots were informed that the auditory task was designed to assess their capacity to process auditory information. In line with standard procedures, they were instructed to always prioritize aircraft control over the auditory task.

A "block" consisted of ten randomly generated numbers, ranging from 0 to 9, pronounced in the ICAO Phonetic Alphabet, where presented over the pilots' headset with 2.5 second intervals resulting in a total block duration of 28 seconds. Each block was preceded by an auditory warning: "The auditory task is coming". The target block, where performance was collected, started at 5, 10, or 15 seconds before the preset event with 2, 4, or 6 additional numbers, respectively, and always continued for 28 seconds (10 numbers) after event onset (see Table 2 and Figure 6).

Non-target (distraction) blocks were also included in the ENF, FLAP, STALL, and LTS. These were presented at different moments than the startle/surprise events (i.e. non-target (distraction) block timing in Table 2) and always lasted 28 seconds. Their only purpose

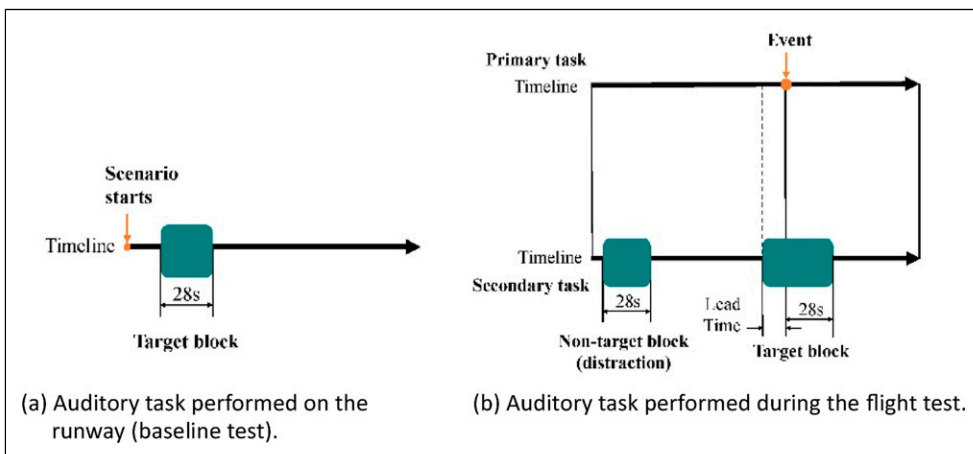


Figure 6. Auditory tasks in the familiarization and test session

was to reduce participants' expectation of startle/surprise events, which were consistently paired with target blocks.

Participants were instructed to single-click the autopilot disconnect button with their thumb if the number was odd, and to double-click it if the number was even. For double-clicks, the interval between clicks had to be less than 500 milliseconds. Otherwise, the response would be recorded as one single-click and one invalid click.

Both aircraft control and auditory task were implemented in the Delft University Environment for Communication and Activation (DUECA). This ensured synchronisation between flight phase, auditory stimulus presentation, and responses acquisition.

Dependent Measures

Auditory Task Reaction Time. The main measure of the information-processing performance was the auditory task reaction time. The autopilot disconnect button was sampled at a frequency of 100 Hz, after which reaction times were calculated with a resolution of 0.01 seconds. The mean reaction time of the correct responses was obtained for the target block (after the event) and for the baseline test for each scenario. To clean the data, responses that were either extremely fast (i.e. quicker than the average response time in baseline) or missing, were excluded as invalid. The mean reaction time during the flight test was then corrected by subtracting the mean baseline reaction time (obtained on the runway), resulting in the Delta Reaction Time (ΔRT).

Auditory Task Accuracy. As additional measure of information-processing performance, accuracy within the target block (after the event) was assessed. It was defined as the ratio of correct responses to the total number of presented numbers in the target block (i.e. 10). To account for individual baseline performance, flight test accuracy was baseline-corrected by subtracting the baseline accuracy (obtained on the runway), yielding the Delta Accuracy (ΔAC).

Measures of Startle and Surprise. Following each test scenario, participants completed the

Startle Inventory (Startle-I) and Surprise Inventory (Surprise-I) regarding each in-flight event (Table 2). The Startle-I and Surprise-I are validated self-report measures designed to assess startle and surprise responses to specific stimuli (Chen et al., 2025a, 2025b). The Startle-I consists of six statements: "It startled me.", "It made me physically flinch.", "It caused my heart to suddenly beat harder or faster.", "It immediately made me feel scared or angry.", "It shocked me.", and "It immediately caused stress or frustration to me.", to which the participant rates agreement on 1–5 Likert scales (1 = Strongly disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly agree). These statements aim to capture operationally meaningful levels of startle, rather than transient startle reflexes. The Surprise-I consists of five statements: "It surprised me.", "It was consistent with my expectations." (reverse-coded), "I predicted it beforehand." (reverse-coded), "I did not see it coming.", and "It was unexpected.", which are scored in the same way as the Startle-I. The total score of each inventory is the average of the items' scores, ranging from 1 to 5.

McDonald's ω (McDonald, 2013) for the current sample ranged from $\omega = 0.88$ to $\omega = 0.96$ for the Startle-I, and $\omega = 0.77$ to $\omega = 0.96$ for the Surprise-I across scenarios, indicating acceptable to excellent internal consistency.

Statistical Analyses

To account for the repeated-measures design and non-independence of observations across both participants and scenarios, a modified 'between and within formulation' (Hox et al., 2010) was applied. For each dependent variable, a separate linear mixed-effects model was fitted, to control for the participant-level and scenario-level and variability.

For testing our hypotheses on the relationships between perceived startle and surprise with ΔRT or ΔAC across different scenarios, heteroscedastic linear mixed-effects models were applied. The fixed effects were the ratings of startle (*Startle-I*), the ratings of surprise (*Surprise-I*), and scenario (*Scenario*) modelled

as a categorical variable with eight levels. This means that potential scenario effects on ΔRT or ΔAC were controlled for. Additionally, if the sequence of scenarios (*Sequence*) was found to have a significant effect on ΔRT or ΔAC , the sequence of the scenarios was also included in the linear mixed-effects model, to control for this as well. To account for general individual differences in ΔRT or ΔAC , participant number (*ID*) was included as a random effect.

Linear mixed-effects models were fitted using the `lme` function from the `nlme` package in R. To account for heteroscedasticity in the residuals, the `varIdent` function was used to accommodate variance differences across scenarios. Effect size was measured as the value of an (estimated) coefficient divided by the total variance. This total variance was estimated as the sum of the between-participants variance and the (average) within participants variance (where the average was taken with respect to all possible scenarios).

Furthermore, the Intraclass correlation coefficient (ICC; Raudenbush & Bryk, 2002) was calculated to assess the proportion of the total variance attributable to differences between participants. The ICCs were derived from random effect results, with a higher ICC indicating the notable between-participant variability, supporting the application of linear mixed-effects models.

As additional exploratory analyses, three correlation matrices were computed: (1) between-participant correlations, using participant-level random effects; (2) between scenario correlations, based on estimated marginal means within scenarios; and (3) residual correlations, derived from model residuals. We added age and flight hours to the matrices, as it would be interesting to check if, for instance, more experienced pilots reported lower surprise overall in the experiment. The between-scenario correlations were obtained to check if, for instance, more startling scenarios also affected secondary task performance more. This is highly exploratory as the number of pairs is very low ($n = 8$) and significant correlations are likely to be confounded by other scenario differences.

Finally, to visualize the temporal pattern of pilot performance during the different

scenarios, the mean and standard deviation of ΔRT were calculated across the sequence of ten numbers, as well as the proportion of valid responses for each number in the sequence.

Results

Relationships Between Startle and Surprise with ΔRT

No significant effect of the scenario sequence (*Sequence*) on ΔRT was found; hence, it was not included in the model. The linear mixed-effects model of ΔRT was fitted as follows:

$$\Delta RT = 1 + Startle - I + Surprise - I + Scenario + (1|ID).$$

The results of the linear mixed-effects model of ΔRT are presented in Table 3. Higher Startle-I scores were significantly associated with an increase in ΔRT , $\beta = 0.049$, $SE = 0.017$, $t(171) = 2.817$, $p = 0.005$. This indicated that for every point scored higher on the Startle-I, ΔRT increased by 49 ms, after controlling for the effect of *Surprise-I*, *Scenario*, and *ID*. In contrast, no significant effect of the *Surprise-I* was found on ΔRT .

The estimated distributions of residuals and random effects were close to a normal distribution, with a Kolmogorov–Smirnov distance (i.e. distance with respect to the supremum-norm) of 0.075 and 0.076, respectively. The residuals were slightly positively skewed (i.e. right-tailed, with a skewness of 0.83) and leptokurtic (with a kurtosis of 5.91), and the random effects were unskewed (skewness of -0.04) and slightly platykurtic (slightly light tails, with a kurtosis of 2.65). Heterogeneity in ΔRT variance was observed across scenarios (Figure 9). This variability supports the use of a heteroscedastic model structure to appropriately account for non-constant variance across scenarios. The estimated standard deviation of the random effect of participant was 0.073, corresponding to an estimated ICC of 0.294. This indicates that approximately 29.4% of the total variance in ΔRT is attributable to between-participant differences (after removing the effect of *Surprise-I*, *Startle-I*, and *Scenario*), supporting the application of

Table 3. Summary of the linear mixed-effects model of the Delta Reaction Time

Effect	Estimate	SE	t	p	Variance factor ^a	Effect size
Fixed effects						
(Intercept)	0.065	0.036	1.8243	0.070	-	0.486
Startle-I (1–5)	0.049	0.017	2.817	0.005**	-	0.363
Surprise-I (1–5)	0.002	0.012	0.139	0.890	-	0.012
Scenario _{ENF}	0.257	0.053	4.849	<0.001**	2.128	1.906
Scenario _{FLAP}	0.131	0.060	2.181	0.031*	2.756	0.974
Scenario _{CARGO}	0.137	0.058	2.379	0.019*	2.547	1.017
Scenario _{LTS}	-0.107	0.046	-2.315	0.022*	1.162	-0.796
Scenario _{PFDf}	-0.008	0.039	-0.202	0.841	1.341	-0.058
Scenario _{STALL}	0.076	0.055	1.388	0.167	2.155	0.567
Scenario _{NLO}	-0.045	0.023	-1.981	0.049*	0.895	-0.333
Random effects (SD)						
Between-participant	0.073	-	-	-	-	-
Within-participant (NTO)	0.086	-	-	-	-	-

Note. Number of observations = 206; Coefficients for each Scenario are in reference to the NTO scenario.

^aThe residual (i.e. within-participant) variance is reported as an estimated multiplicative factor relative to the scenario NTO. For example, the variance of the residuals for scenario ENF is $2.128 \times 0.086 = 0.183$.

* $p < 0.05$ (two-tailed). ** $p < 0.01$ (two-tailed).

mixed-effects model to account for the non-independence of observations.

When controlling for the effect of Startle-I, Surprise-I, and individual differences, ENF, FLAP, and CARGO resulted in significantly higher ΔRT than the reference scenario NTO. In ENF, ΔRT was 257 ms higher, $\beta = 0.257$, $SE = 0.053$, $t(171) = 4.849$, $p < 0.001$. In FLAP, ΔRT was 131 ms higher, $\beta = 0.131$, $SE = 0.060$, $t(171) = 2.181$, $p = 0.031$, and in CARGO, ΔRT was 137 ms higher, $\beta = 0.137$, $SE = 0.058$, $t(171) = 2.379$, $p = 0.019$. Participants showed significantly lower ΔRT in LTS and NLO compared to the NTO. ΔRT was 107 ms lower in LTS, $\beta = -0.107$, $SE = 0.046$, $t(171) = -2.315$, $p = 0.022$, and ΔRT was 45 ms lower in NLO, $\beta = -0.045$, $SE = 0.023$, $t(171) = -1.981$, $p = 0.049$.

Relationships Between Startle and Surprise with ΔAC

The sequence of scenarios had a significant effect on ΔAC . *Sequence* was therefore included in the linear mixed-effects model of ΔAC :

$$\Delta AC = 1 + \text{Startle} - I + \text{Surprise} - I \\ + \text{Scenario} + \text{Sequence} + (1|ID).$$

Results from the linear mixed-effects model of ΔAC are presented in Table 4. The model revealed no significant effect of *Surprise-I* nor *Startle-I* on ΔAC .

The estimated distributions of residuals and random effects were close to a normal distribution, with a Kolmogorov–Smirnov distance (i.e. distance with respect to the supremum-norm) of 0.162 and 0.196, respectively. Both residuals and random effects were negatively skewed (i.e. left-tailed, with skewness -1.56 and -1.75 , respectively) and leptokurtic (slightly fat tails, with kurtosis of 8.63 and 6.76). Heterogeneity in ΔAC variance across scenarios (Figure 10) further supports modelling heteroscedasticity to account for variance differences between scenarios. The estimated standard deviation of the random intercept was 4.498, corresponding to an approximately ICC of 0.225. This finding shows that approximately 22.5% of the total variance in ΔAC stems from between-participant differences, further justifying the use of a mixed-effects model.

Table 4. Summary of the linear mixed-effects model of the Delta Accuracy

Effect	Estimate	SE	t	p	Variance factor ^a	Effect size
Fixed effects						
(Intercept)	-0.594	2.845	-0.209	0.835	-	-0.063
Startle-I (1-5)	-0.840	1.267	-0.663	0.508	-	-0.089
Surprise-I (1-5)	-0.640	0.861	-0.744	0.458	-	-0.067
Sequence	0.197	0.304	0.646	0.519	-	0.021
Scenario _{ENF}	-25.412	6.038	-4.208	<0.001**	5.447	-2.677
Scenario _{FLAP}	-8.115	4.123	-1.968	0.051	3.149	-0.855
Scenario _{CARGO}	-8.629	4.734	-1.823	0.070	3.994	-0.909
Scenario _{LTS}	-0.745	3.235	-0.230	0.818	1.324	-0.079
Scenario _{PFDF}	1.707	2.627	0.650	0.517	1.494	0.180
Scenario _{STALL}	-4.710	4.114	-1.145	0.254	2.937	-0.496
Scenario _{NLO}	0.053	2.623	0.020	0.984	2.437	0.006
Random effect (SD)						
Between-participant	4.498	-	-	-	-	-
Within-participant (NTO)	5.506	-	-	-	-	-

Note. Number of observations = 208; coefficients for each Scenario are in reference to the NTO scenario.

^aEstimated multiplicative factor of the residual (i.e. within-participant) variance with respect to the reference scenario NTO. For example, the variance of the residuals for scenario ENF is $5506 \times 5.447 = 29.991$.

* $p < 0.05$ (two-tailed); ** $p < 0.01$ (two-tailed).

The fixed-effects intercept was not statistically significant, indicating that ΔAC in NTO was not significantly different from zero. Using NTO as reference, ENF exhibited significantly lower ΔAC . ΔAC in ENF was $-0.594 - 25.412 = -26.006$, with $\beta = -25.412$, $SE = 6.038$, $t(172) = -4.208$, $p < 0.001$.

Descriptive Statistics

One case of ΔRT was missing in ENF and one case in CARGO, due to insufficient numbers of correct responses.

Table 5 lists an overview of dependent measures across the eight scenarios. Figures 7 and 8 present box plots illustrating Startle-I and Surprise-I responses, respectively, across these scenarios. The figures demonstrate that the scenarios elicited high individual variability in startle and surprise responses, although the means of each scenario did not always match our expectations. LTS was not rated very low on surprise, ENF, FLAP, and CARGO resulted in somewhat similar mean startle ratings. In general, Surprise-I scores were

Table 5. Means and standard deviations of the dependent measures across scenarios

Scenario	ΔRT (s)	ΔAC (%)	Startle-I (1-5)	Surprise-I (1-5)
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
ENF	0.46 (0.21)	-30.00 (29.26)	2.76 (0.66)	4.01 (0.71)
FLAP	0.33 (0.25)	-12.69 (18.45)	2.63 (0.76)	4.22 (0.65)
CARGO	0.33 (0.25)	-13.08 (22.59)	2.58 (0.75)	4.09 (0.65)
LTS	0.12 (0.13)	-5.38 (8.11)	3.28 (0.93)	3.44 (1.10)
PFDF	0.17 (0.14)	-2.31 (9.08)	2.07 (0.61)	3.94 (0.62)
STALL	0.29 (0.23)	-9.62 (16.37)	2.85 (0.81)	4.32 (0.47)
NTO	0.13 (0.11)	-1.92 (5.67)	1.17 (0.31)	1.91 (0.86)
NLO	0.09 (0.09)	-1.92 (10.96)	1.29 (0.37)	1.82 (0.74)

Note. ΔRT = Delta Reaction Time; ΔAC = Delta Accuracy; Startle-I = Startle Inventory; Surprise-I = Surprise Inventory.

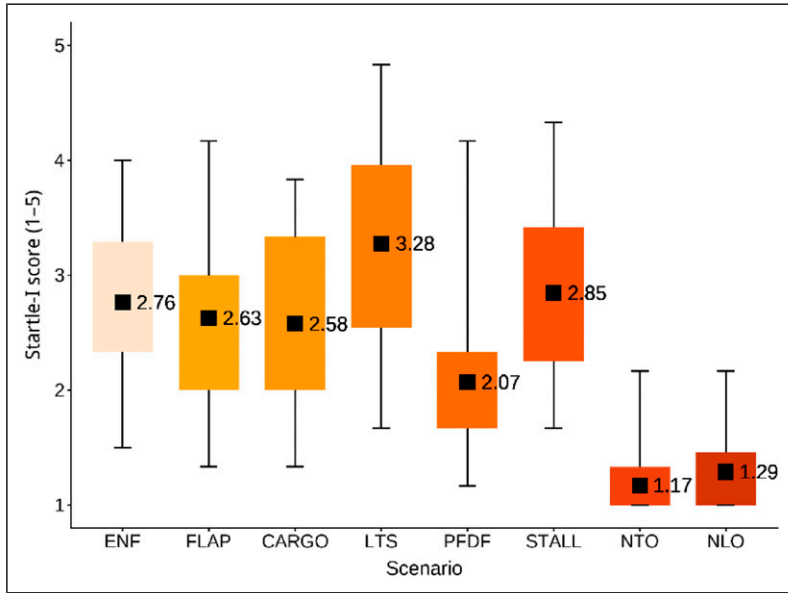


Figure 7. The Startle-I scores in the eight scenarios (square markers indicate means, whiskers indicate interquartile ranges)

higher than Startle-I scores. As expected, LTS produced relatively high startle and low surprise, while STALL produced relatively high startle as well as surprise.

PFDf scored as expected relatively low on startle and high on surprise.

Figures 9 and 10 present box plots of ΔRT and ΔAC for each scenario. ENF, FLAP, and

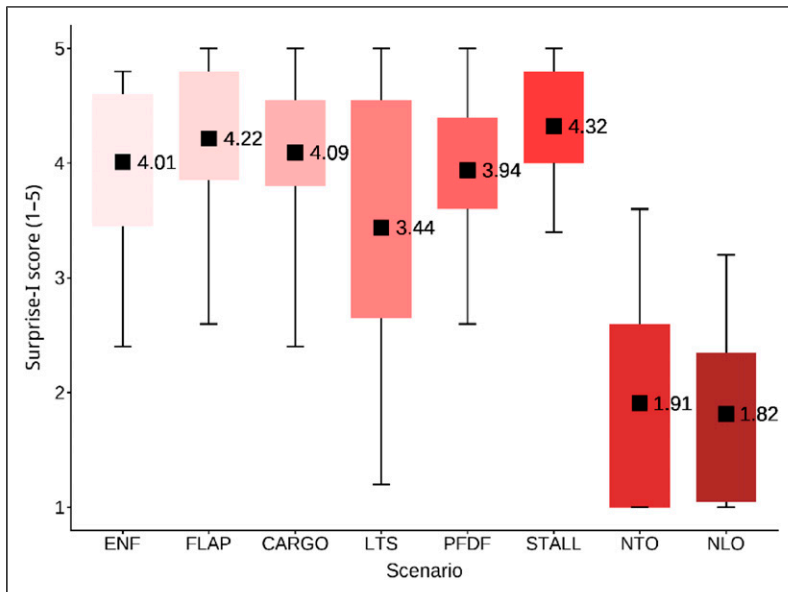


Figure 8. The Surprise-I scores in the eight scenarios (square markers indicate means, whiskers indicate interquartile ranges)

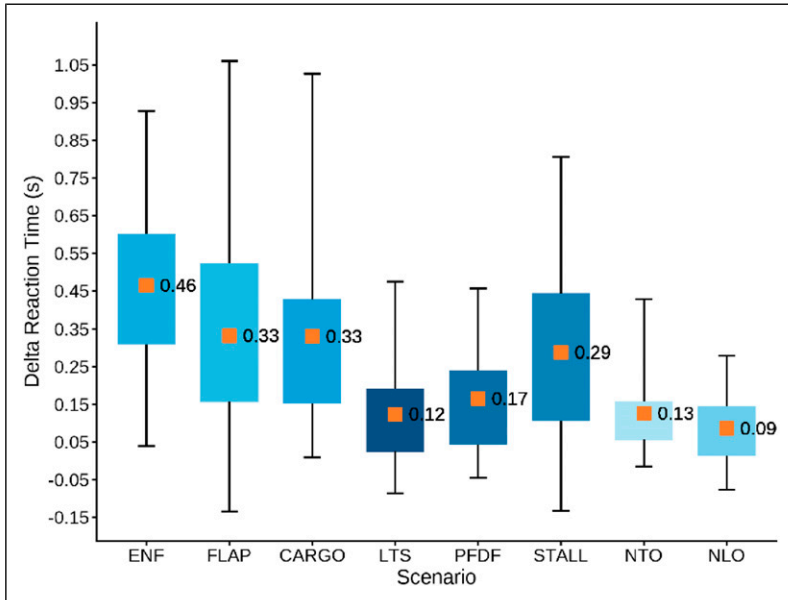


Figure 9. The Delta Reaction Time of the auditory task in the eight scenarios (square markers indicate means, whiskers indicate interquartile ranges)

CARGO, each requiring manual intervention, showed the highest mean ΔRT s, lowest mean ΔAC s, and the highest variance in both measures. Although STALL did not involve

manual intervention, it appeared similarly impactful. In contrast, LTS (highly startling, less surprising) and PFDF (highly surprising, less startling) showed relatively low impact.

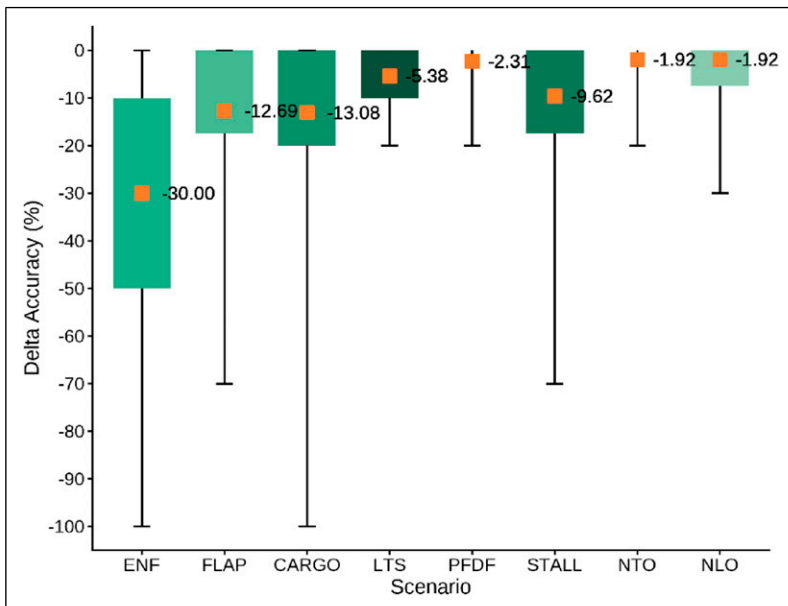


Figure 10. The Delta Accuracy of the auditory task in the eight scenarios (square markers indicate means, whiskers indicate interquartile ranges)

Exploratory Correlation Analyses

Table 6 lists the Pearson correlations between Age, Flight hours, ΔRT , ΔAC , Startle-I, and Surprise-I scores at between-participant, between-scenario, and residual levels.

At the between-participant level, both Age and Flight hours were significantly negatively associated with ΔRT , $r = -0.596$ and $r = -0.599$, respectively. These findings suggest that older and more experienced pilots demonstrate lower ΔRT on the secondary task.

At the between-scenario level, Surprise-I was significantly positively correlated with ΔRT , $r = 0.718$, and also with Startle-I, $r = 0.817$. These findings suggest that, within the context of the tested scenarios, the relatively more surprising scenarios also tended to be relatively more startling, and were associated with higher ΔRT on the secondary task. There

was no significant correlation between Startle-I and ΔRT on the between-scenario level.

Additionally, ΔAC showed a significant negative correlation with ΔRT at between-scenario, $r = -0.932$, and between-participant level, $r = -0.401$. These findings indicate that both participants and scenarios with longer ΔRT tended to exhibit more errors on the secondary task.

At the residual level, Startle-I was significantly positively correlated with ΔRT , $r = 0.162$ and with Surprise-I, $r = 0.161$. This suggests that unusually high Startle-I ratings were associated with higher-than-expected Surprise-I ratings and longer-than expected ΔRT . In addition, ΔAC showed a significant negative correlation with ΔRT , $r = -0.161$. This means that if, in a given scenario, ΔRT was higher than expected for that participant and scenario, more errors were likely to occur on the secondary task.

Table 6. Correlation matrices of the study variables

	Age	FH	ΔRT	ΔAC	Startle-I	Surprise-I
Between-participant correlation matrix						
Age	1.000					
FH	0.917**	1.000				
ΔRT	-0.596**	-0.599**	1.000			
ΔAC	-0.107	-0.118	-0.401*	1.000		
Startle-I	-0.242	-0.301	0.237	-0.041	1.000	
Surprise-I	0.097	0.112	-0.038	-0.152	0.134	1.000
Between-scenario correlation matrix						
Age	-					
FH	-	-				
ΔRT	-	-	1.000			
ΔAC	-	-	-0.932**	1.000		
Startle-I	-	-	0.528	-0.508	1.000	
Surprise-I	-	-	0.718*	-0.543	0.817*	1.000
Residual correlation matrix						
Age	-					
FH	-	-				
ΔRT	-	-	1.000			
ΔAC	-	-	-0.161*	1.000		
Startle-I	-	-	0.162*	0.008	1.000	
Surprise-I	-	-	0.026	-0.044	0.161*	1.000

Note. FH = Flight hours; ΔRT = Delta Reaction Time; ΔAC = Delta Accuracy; Startle-I = Startle Inventory; Surprise-I = Surprise Inventory.

** $p < 0.01$ (two-tailed); * $p < 0.05$ (two-tailed).

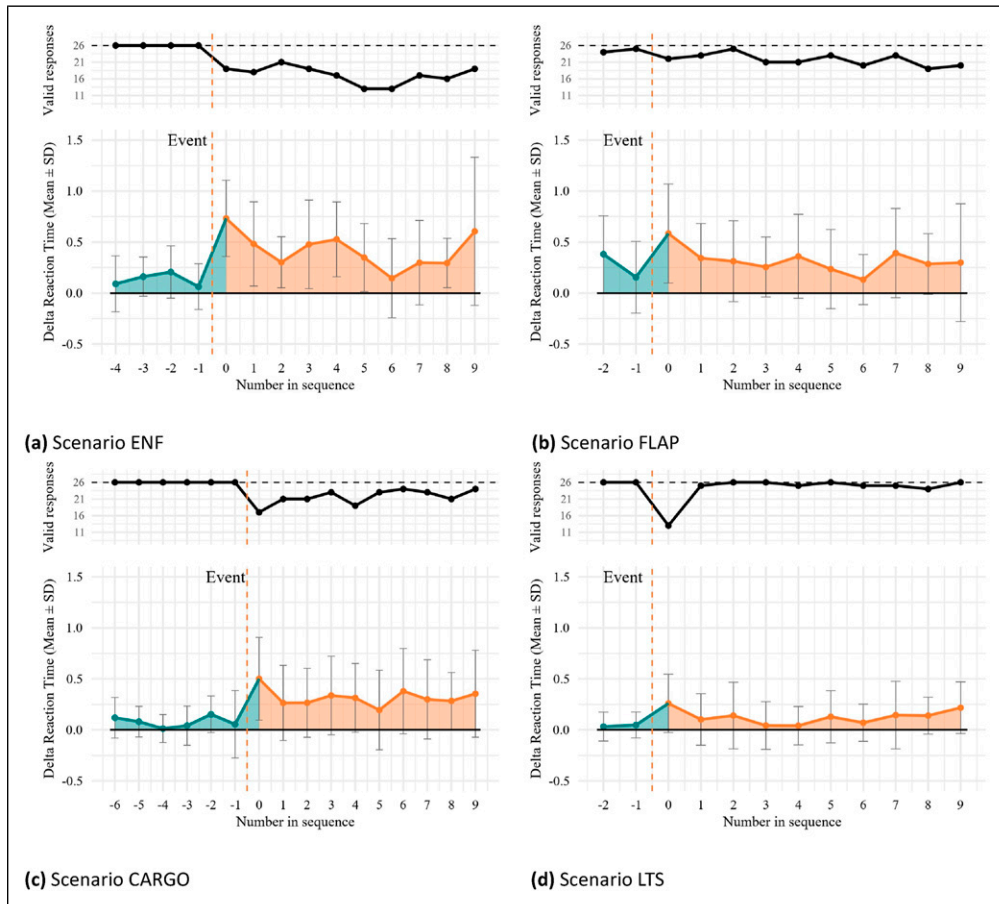


Figure 11. Mean Delta Reaction Time (± 1 SD) with valid responses in target block per scenario

Temporal Patterns of ΔRT and Valid Responses

Figures 11 and 12 provide an overview of the ΔRT and valid responses across the secondary task duration in the target block. In the ΔRT plots, the black line represents a zero change in Reaction Time, while the shaded areas highlight deviations from this reference. In valid responses plots, the gray dashed line indicates the total number of participants ($N = 26$).

In all scenarios with preset events, ΔRT showed a sharp increase immediately after the event (dashed vertical line), peaking at the first number in the sequence and coinciding with a drop in valid responses. For ENF, FLAP, and CARGO, where pilots had to intervene,

changes in ΔRT and valid responses remained evident throughout the entire measurement period. Additionally, these scenarios exhibited greater postevent variability in ΔRT , as reflected by larger standard deviations. A similar pattern was observed in STALL, possibly due to the persistent warning sound. The impact of LTS and PFDF subsided quickly, suggesting a brief impairment in information processing.

Discussion

We tested whether there were significant relationships between pilots' startle and surprise responses and information-processing performance during simulated in-flight events, while

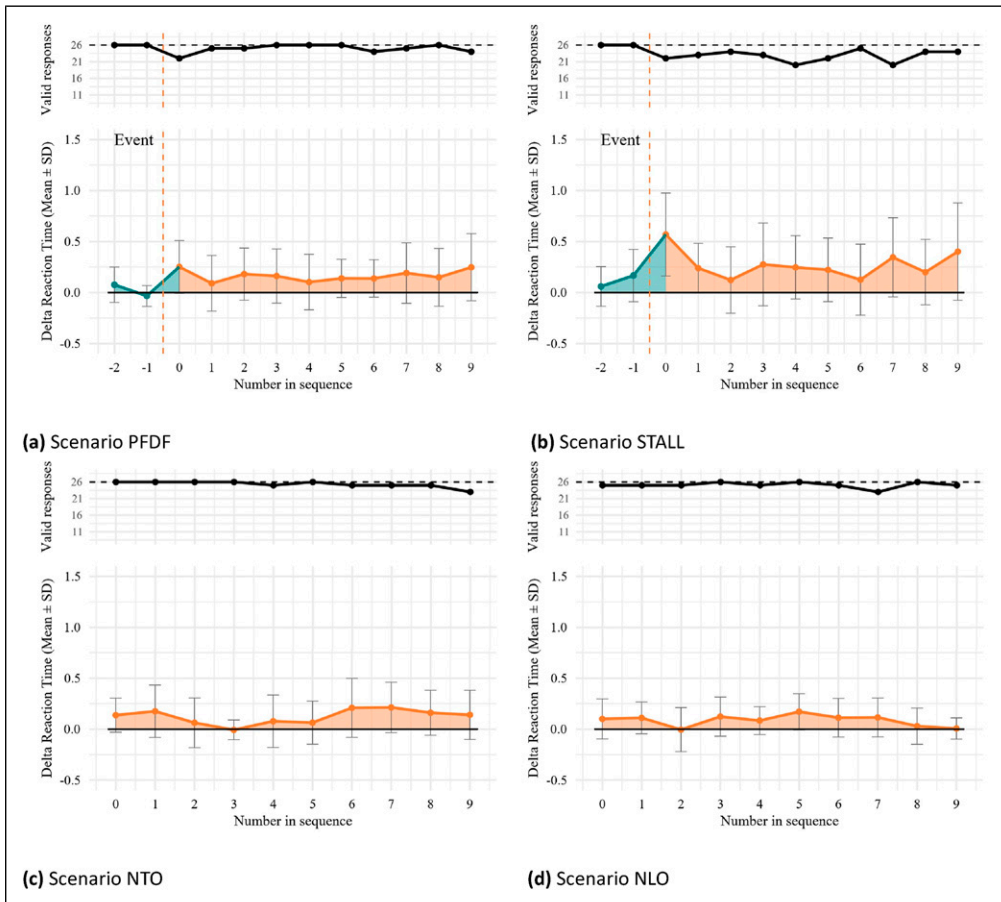


Figure 12. Mean Delta Reaction Time (± 1 SD) with valid responses in target block per scenario (continued)

controlling for general differences between these events, and for general differences between individuals. Our results indicate that more severe startle responses were significantly related to larger impairments of information-processing speed, as evidenced by increased ΔRT on the secondary auditory task (see Table 3). This was not additionally reflected in a significant positive correlation between startle and secondary task performance on the scenario level (see Table 6), meaning that other differences between scenarios, such as workload or manual control, overruled this effect when comparing scenarios. The findings align with prior research in lab settings (Duchevet et al., 2025) and in-flight simulators (Martin et al., 2015) suggesting that startle leads to temporary

cognitive disruption, likely due to the rapid engagement of survival circuits that prioritize threat responses over executive functioning (Eysenck et al., 2007; LeDoux, 2012). The involuntary nature of startle appears to divert attentional resources from ongoing tasks, requiring additional cognitive demands to reorient focus, which is supported by evidence of increased cerebral blood flow in prefrontal cortex regions (Parasuraman & Caggiano, 2005).

Interestingly, these findings stand in contrast to recent research (Schwartz et al., 2025), which focused on a controlled cognitive task in a single-task paradigm. In this context, no performance impairment following startle was observed, and slight performance improvements were even reported under high cognitive

load conditions. One possible explanation for this discrepancy lies in the differences between controlled laboratory environment and the more complex task in our study. In laboratory contexts, the absence of perceived threat may enable participants to rapidly compensate for the effect of startle, limiting the disruptive effect on ongoing tasks. However, real-world startle responses are more likely to occur in highly demanding and threatening situations, where task complexity and pressure may overwhelm compensatory mechanisms and exacerbate performance disruptions.

Contrary to our hypothesis, but in line with a previous study (Duchevet et al., 2025), surprise did not have a distinct significant impact on secondary task performance. While surprise has been theorized to impose cognitive demands by prompting a need for sensemaking and situational “reframing” (Klein et al., 2007; Landman et al., 2017), our results show that its immediate effect on information processing were not significant in the in-flight events featured in our experiment. It could be that the surprising events in our study may not have been complex enough to elicit sufficient variance in pilots’ “reframing” efforts that would induce sufficient additional workload to impact cognitive task performance. This discrepancy highlights the complexity of modelling surprise effects and suggests that its impact may be more contextual, indirect, or delayed, compared to the more reflexive and acute nature of startle.

Although the mixed-effects model did not reveal a significant relationship between surprise and information-processing performance, this does not imply that surprise is less relevant for pilot training. In fact, a significant correlation between surprise ratings and ΔRT was observed at the between-scenario level, suggesting that more surprising scenarios are generally associated with greater disruption to information-processing performance. Additionally, the temporal pattern analysis highlights the influence of surprise on response dynamics. One possible explanation for this discrepancy is that the between-scenario correlations reflect general trends across scenarios, whereas the mixed-effects model accounts for both between-participant and

between-scenario variability while controlling for other contributing factor. After adjusting for startle ratings, scenario level variability, and individual differences, the effect of surprise on ΔRT and ΔAC was non-significant. Nevertheless, previous studies have shown that unexpectedness in training is important to build cognitive flexibility and higher-level competences to deal with a wide variety of possible events (Helmreich et al., 2017; Landman et al., 2018; Salas et al., 2006). Given its importance for training, more research is needed into means to systematically introduce surprise in aviation scenarios in a highly ecologically valid manner.

There was a negative correlation between flight experience and ΔRT , suggesting that more experienced pilots were better at managing cognitive disruptions caused by in-flight events. This finding supports the notion that expertise and training can mitigate the impact of unexpected events (Causse et al., 2019), potentially by enabling pilots to rely on well-practiced procedural knowledge that requires fewer cognitive resources under stress. This underscores the importance of tailored training programs designed to enhance resilience to startle, particularly for less experienced pilots.

The findings further highlight the necessity of incorporating startle management strategies into pilot training programs. Although current UPRT protocols acknowledge the role of startle and surprise in aviation safety, our research suggests that targeted interventions specifically designed to mitigate startle-induced cognitive disruptions may be beneficial (Vlaskamp et al., 2024). Techniques such as stress management, exposure-based training, and cognitive resilience exercises may help pilots reduce the disruptive effects of startle and maintain optimal performance under pressure (Causse et al., 2011; Driskell & Johnston, 1998; Landman et al., 2020).

Several limitations of the current study should be acknowledged, along with corresponding recommendations for future research. First, although the data revealed both within-scenario and between-scenario variation in Startle-I ratings, the intensity of the startle

responses observed was likely lower than those typically experienced during real-world aviation emergencies. This limitation may be attributed to the controlled nature of the simulation context, in which participants were aware that no actual threat was present. Future research could employ more immersive simulation techniques, introduce social stressors (e.g. performance evaluation and peer observation) or incorporate sudden-onset stimuli to elicit stronger startle responses, and further validate their effects under more ecologically intense conditions.

Second, the quantification of startle and surprise relied on self-report measures, which, although informative and psychometrically validated, are subject to potential biases such as individual differences in interpreting questionnaire items. Future research could enhance the assessment of startle and surprise by integrating self-report data with physiological indicators (e.g. reflex electromyogram (EMG) (Blumenthal et al., 2005), pupillometry (Leuchs et al., 2019), and heart rate (Kinney and O'Hare, 2020)).

Third, the experimental setup involved single-pilot crew operating a twin-propeller model, which may not fully replicate the dynamics of real-world multi-crew operations. Investigating how startle and surprise interact in team-based settings could provide valuable insights into crew resource management strategies and training.

Fourth, certain scenario characteristics, most notably the requirement for manual control, were inherently confounded with scenario membership. As a consequence, manual control could not be modelled as an independent predictor alongside scenario effects without introducing perfect collinearity. Nevertheless, startle was assessed at the individual level and exhibited substantial variability within and across scenarios, allowing its association with Delta Reaction Time to be estimated beyond scenario-level differences. Overall, these results indicate that secondary task performance decrements likely reflect the combined influence of startle responses and task-related demands (e.g. manual control). Future studies could address this limitation by directly measuring control activity or by

experimentally manipulating manual control independently of scenario context.

Fifth, the lack of significant findings regarding ΔAC might be due to low resolution of the secondary task measurement or ceiling effects. Future studies might benefit from integrating additional metrics, such as physiological measures or increasing the difficulty of the cognitive task, to better capture the fluctuation of information-processing performance.

Conclusion

In summary, the study demonstrates that higher levels of startle were significantly associated with impaired information-processing performance in pilots, whereas no such association was found for surprise. These findings reinforce the importance of training interventions aimed at managing startle effects in high-risk environments. Additionally, the role of experience in mitigating these effects highlights the value of continued practice and exposure-based training. Future research should explore more nuanced measures of cognitive capacity and extend these findings to multi-crew settings to further enhance aviation safety protocols.

Key Points

- Higher levels of startle, as measured by self-report, was significantly associated with impaired information-processing speed, whereas no such association was found for surprise. The information-processing performance were assessed using a secondary auditory cognitive task.
- The study employed eight different in-flight scenarios in a highly ecologically valid setting.
- The flight scenarios induced an adequate range of startle and surprise responses in the pilots.
- The findings highlight the need for tailored interventions and training protocols to mitigate the potentially disruptive effects of startle responses, enhancing pilot performance during in-flight emergencies.


ORCID iDs

Jiayu Chen  <https://orcid.org/0009-0006-0084-1661>

Annemarie Landman  <https://orcid.org/0000-0003-3678-9210>

Alexis Derumigny  <https://orcid.org/0000-0002-6163-8097>

Olaf Stroosma  <https://orcid.org/0000-0002-4578-3171>

M. M. (René) van Paassen  <https://orcid.org/0000-0003-4700-1222>

Max Mulder  <https://orcid.org/0000-0002-0932-3979>

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

- Blumenthal, T. D. (1988). The startle response to acoustic stimuli near startle threshold: Effects of stimulus rise and fall time, duration, and intensity. *Psychophysiology*, *25*(5), 607–611. <https://doi.org/10.1111/j.1469-8986.1988.tb01897.x>
- Blumenthal, T. D., Cuthbert, B. N., Filion, D. L., Hackley, S., Lipp, O. V., & van Boxtel, A. (2005). Committee report: Guidelines for human startle eyeblink electromyographic studies. *Psychophysiology*, *42*(1), 1–15. <https://doi.org/10.1111/j.1469-8986.2005.00271.x>
- Bradford, D. E., Kaye, J. T., & Curtin, J. J. (2014). Not just noise: Individual differences in general startle reactivity predict startle response to uncertain and certain threat. *Psychophysiology*, *51*(5), 407–411. <https://doi.org/10.1111/psyp.12193>
- Bradley, M. M., Moulder, B., & Lang, P. J. (2005). When good things go bad: The reflex physiology of defense. *Psychological Science*, *16*(6), 468–473. <https://doi.org/10.1111/j.0956-7976.2005.01558.x>
- Casner, S. M., Geven, R. W., & Williams, K. T. (2013). The effectiveness of airline pilot training for abnormal events. *Human Factors*, *55*(3), 477–485. <https://doi.org/10.1177/0018720812466893>
- Causse, M., Chua, Z. K., & Rémy, F. (2019). Influences of age, mental workload, and flight experience on cognitive performance and prefrontal activity in private pilots: A fMRI study. *Scientific Reports*, *9*(7688), 7688. <https://doi.org/10.1038/s41598-019-44082-w>
- Causse, M., Dehais, F., & Pastor, J. (2011). Executive functions and pilot characteristics predict flight simulator performance in general aviation pilots. *The International Journal of Aviation Psychology*, *21*(3), 217–234. <https://doi.org/10.1080/10508414.2011.582441>
- Chen, J., Landman, A., Derumigny, A., Stroosma, O., van Paassen, M. M., & Mulder, M. (2025a). Development and validation of the startle and surprise inventories and visual analogue scales. *Ergonomics*, 1–14. <https://doi.org/10.1080/00140139.2025.2529317>
- Chen, J., Landman, A., Derumigny, A., Stroosma, O., van Paassen, M. M., & Mulder, M. (2025b). Preliminary multilevel confirmatory factor analysis of the startle & surprise inventories in an operational aviation context. *Journal of Cognitive Engineering and Decision Making*. <https://doi.org/10.1177/15553434251401784>
- De Boer, R. J., & Dekker, S. W. A. (2017). Models of automation surprise: Results of a field survey in aviation. *Safety*, *3*(20), 20. <https://doi.org/10.3390/safety3030020>
- Dehais, F., Causse, M., Vachon, F., Regis, N., Menant, E., & Tremblay, S. (2013). Failure to detect critical auditory alerts in the cockpit: Evidence for inattentive deafness. *Human Factors*, *56*(4), 631–644. <https://doi.org/10.1177/0018720813510735>
- Deniel, J., Neubert, J., Schwartz, F., & Causse, M. (2024). Impact of startle reflex on cognitive performance, face temperature and brain activity. In D. Harris & W.-C. Li (Eds.), *Engineering psychology and cognitive ergonomics* (pp. 36–50). Springer Nature. https://doi.org/10.1007/978-3-031-60728-8_4
- Dreissen, Y. E. M., Bakker, M. J., Koelman, J. H. T. M., & Tijssen, M. A. J. (2012).

- Exaggerated startle reactions. *Clinical Neurophysiology*, 123(1), 34–44. <https://doi.org/10.1016/j.clinph.2011.09.022>
- Driskell, J. E., & Johnston, J. H. (1998). Stress exposure training. In J. A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 191–217). American Psychological Association. <https://doi.org/10.1037/10278-007>
- Duchevet, A., Imbert, J.-P., Garcia, J., Lamirault, B., & Causse, M. (2025). Investigating the independent and combined effects of startle and Surprise in a simulated flight task. *Human Factors*, 67(11), 1170–1187. <https://doi.org/10.1177/00187208251342100>
- European Aviation Safety Agency. (2015). *Loss-of-Control prevention and recovery training (no. NPA 2015-13)*. Retrieved from <https://www.easa.europa.eu/sites/default/files/dfu/NPA2015-13.pdf>
- Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, 7(2), 336–353. <https://doi.org/10.1037/1528-3542.7.2.336>
- Federal Aviation Administration. (2015). *Upset prevention and recovery training (No. Advisory Circular 120-111)*. Retrieved from https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_120-111_CHG_1_Ed_Upd_3.pdf
- Helmreich, R. L., Merritt, A. C., & Wilhelm, J. A. (2017). The evolution of crew resource management training in commercial aviation. In *Human error in aviation* (pp. 275–288). Routledge.
- Hinkelmann, K., & Kempthorne, O. (2007). Latin square type designs. In *Design and analysis of experiments: Introduction to experimental design* (pp. 373–417): John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470191750.ch10>
- Holand, S., Girard, A., Laude, D., Meyer-Bisch, C., & Elghozi, J.-L. (1999). Effects of an auditory startle stimulus on blood pressure and heart rate in humans. *Journal of Hypertension*, 17(12), 1893–1897. <https://doi.org/10.1097/00004872-199917121-00018>
- Horstmann, G. (2006). Latency and duration of the action interruption in surprise. *Cognition & Emotion*, 20(2), 242–273. <https://doi.org/10.1080/02699930500262878>
- Hox, J., Moerbeek, M., & van de Schoot, R. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). Routledge. <https://doi.org/10.4324/9780203852279>
- Izard, C. E., Libero, D. Z., Putnam, P., & Haynes, O. (1993). Stability of emotion experiences and their relations to traits of personality. *Journal of Personality and Social Psychology*, 64(5), 847–860. <https://doi.org/10.1037//0022-3514.64.5.847>
- Jansen, A. S. P., van Nguyen, X., Karpitskiy, V., Mettenleiter, T. C., & Loewy, A. D. (1995). Central command neurons of the sympathetic nervous system: Basis of the fight-or-flight response. *Science*, 270(5236), 644–646. <https://doi.org/10.1126/science.270.5236.644>
- Kinney, L., & O'Hare, D. (2020). Responding to an unexpected in-flight event: Physiological arousal, information processing, and performance. *Human Factors*, 62(5), 737–750. <https://doi.org/10.1177/0018720819854830>
- Klein, G., Phillips, J. K., Rall, E. L., & Peluso, D. A. (2007). A data-frame theory of sensemaking. In R. R. Hoffman (Ed.), *Expertise out of context: Proceedings of the sixth international conference on naturalistic decision making*. Lawrence Erlbaum Associates Publishers.
- Koch, M. (1999). The neurobiology of startle. *Progress in Neurobiology*, 59(2), 107–128. [https://doi.org/10.1016/S0301-0082\(98\)00098-7](https://doi.org/10.1016/S0301-0082(98)00098-7)
- Ladd, C. O., Plotsky, P. M., & Davis, M. (2000). Startle response. In G. Fink (Ed.), *Encyclopedia of stress* (Vol. 3). Academic Press.
- Landman, A., Groen, E. L., van Paassen, M. M., Bronkhorst, A. W., & Mulder, M. (2017). Dealing with unexpected events on the flight deck: A conceptual model of startle and Surprise. *Human Factors*, 59(8), 1161–1172. <https://doi.org/10.1177/0018720817723428>
- Landman, A., van Middelaar, S. H., Groen, E. L., van Paassen, M. M., Bronkhorst, A. W., & Mulder, M. (2020). The effectiveness of a mnemonic-type startle and surprise management procedure for pilots. *The International Journal of Aerospace Psychology*, 30(3–4),

- 104–118. <https://doi.org/10.1080/24721840.2020.1763798>
- Landman, A., van Oorschot, P., van Paassen, M. M., Groen, E. L., Bronkhorst, A. W., & Mulder, M. (2018). Training pilots for unexpected events: A simulator study on the advantage of unpredictable and variable scenarios. *Human Factors*, 60(6), 793–805. <https://doi.org/10.1177/0018720818779928>
- Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1990). Emotion, attention, and the startle reflex. *Psychological Review*, 97(3), 377–395.
- LeDoux, J. (2012). Rethinking the emotional brain. *Neuron*, 73(4), 653–676. <https://doi.org/10.1016/j.neuron.2012.02.004>
- Leuchs, L., Schneider, M., & Spooemaker, V. I. (2019). Measuring the conditioned response: A comparison of pupillometry, skin conductance, and startle electromyography. *Psychophysiology*, 56(1), e13283. <https://doi.org/10.1111/psyp.13283>
- Martin, W. L., Murray, P. S., Bates, P. R., & Lee, P. S. Y. (2015). Fear-potentiated startle: A review from an aviation perspective. *The International Journal of Aviation Psychology*, 25(2), 97–107. <https://doi.org/10.1080/10508414.2015.1128293>
- Martin, W. L., Murray, P. S., Bates, P. R., & Lee, P. S. Y. (2016). A flight simulator study of the impairment effects of startle on pilots during unexpected critical events. *Aviation Psychology and Applied Human Factors*, 6(1), 24–32. <https://doi.org/10.1027/2192-0923/a000092>
- McDonald, R. P. (2013). *Test theory: A unified treatment*. Psychology Press.
- Meyer, W.-U., Reisenzein, R., & Schützwohl, A. (1997). Toward a process analysis of emotions: The case of surprise. *Motivation and Emotion*, 21(3), 251–274. <https://doi.org/10.1023/A:1024422330338>
- Papadimitriou, A., & Priftis, K. N. (2009). Regulation of the hypothalamic-pituitary-adrenal axis. *Neuroimmunomodulation*, 16(5), 265–271. <https://doi.org/10.1159/000216184>
- Parasuraman, R., & Caggiano, D. (2005). Neural and genetic assays of mental workload. In D. McBride & D. Schmorrow (Eds.), *Quantifying human information processing* (pp. 123–155). Rowman & Littlefield Publishers, Inc.
- Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 30(3), 286–297. <https://doi.org/10.1109/3468.844354>
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Sage Publications.
- Reisenzein, R., Horstmann, G., & Schützwohl, A. (2019). The cognitive-evolutionary model of surprise: A review of the evidence. *Topics in Cognitive Science*, 11(1), 50–74. <https://doi.org/10.1111/tops.12292>
- Rivera, J., Talone, A. B., Boesser, C. T., Jentsch, F., & Yeh, M. (2014). Startle and Surprise on the flight deck: Similarities, differences, and prevalence. In *Proceedings of the human factors and ergonomics society 58th annual meeting* (Vol. 58, pp. 1047–1051). Sage Publications. <https://doi.org/10.1177/15419312145812>
- Salas, E., Wilson, K. A., Burke, C. S., & Wightman, D. C. (2006). Does crew resource management training work? An update, an extension, and some critical needs. *Human Factors*, 48(2), 392–412. <https://doi.org/10.1518/001872006777724444>
- Sarter, N. B., Mumaw, R. J., & Wickens, C. D. (2017). Pilots' monitoring strategies and performance on automated flight decks: An empirical study combining behavioral and eye-tracking data. In *Human error in aviation* (pp. 205–215). Routledge.
- Schwartz, F., Deniel, J., & Causse, M. (2025). Effects of startle on cognitive performance and physiological activity revealed by FNIRS and thermal imaging. *Scientific Reports*, 15(6878), 6878. <https://doi.org/10.1038/s41598-025-90540-z>
- Sehlström, M., Ljungberg, J. K., Claeson, A., & Nyström, M. B. T. (2022). The relation of neuroticism to physiological and behavioral stress responses induced by auditory startle. *Brain and Behavior*, 12, e2554. <https://doi.org/10.1002/brb3.2554>
- Valls-Sole, J., Kumru, H., & Kofler, M. (2008). Interaction between startle and voluntary reactions in humans. *Experimental Brain Research*, 187(4), 497–507. <https://doi.org/10.1007/s00221-008-1402-0>

- Vlaskamp, D., Landman, A., van Rooij, J., Li, W.-C., & Blundell, J. (2024). Airline pilots' perceived operational benefit of a startle and surprise management method: A qualitative study. In *Proceedings of the 2nd international conference on cognitive aircraft systems* (pp. 29–34). SciTePress. <https://doi.org/10.5220/0012927800004562>
- Wessel, J. R., & Aron, A. R. (2017). On the globality of motor suppression: Unexpected events and their influence on behavior and cognition. *Neuron*, *93*(2), 259–280. <https://doi.org/10.1016/j.neuron.2016.12.013>
- Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, *3*(2), 159–177. <https://doi.org/10.1080/14639220210123806>
- Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors*, *50*(3), 449–455. <https://doi.org/10.1518/001872008X288394>
- Young, M. S., & Stanton, N. A. (2002). Malleable attentional resources theory: A new explanation for the effects of mental underload on performance. *Human Factors*, *44*(3), 365–375. <https://doi.org/10.1518/0018720024497709>

Author Biographies

Jiayu Chen received her PhD in aerospace engineering at Delft University of Technology in 2025. She is currently working as a researcher in the Intelligent Cockpit Evaluation Department at China Automotive Engineering Research Institute.

Annmarie Landman received her PhD (cum laude) in aerospace engineering at Delft University of Technology in 2019. She is currently working as a scientist in the Training and Performance Innovations department at TNO Human Factors and as an assistant professor at the section Control and Simulation.

Alexis Derumigny received his PhD in applied mathematics from the National School of Statistics and Economic Administration (ENSAE Paris) in 2019. He is currently an assistant professor in the Department of Applied Mathematics, Delft University of Technology.

Olaf Stroosma earned his MSc in aerospace engineering from Delft University of Technology in 1998. He is currently a senior researcher at the section Control and Simulation, Aerospace Engineering, Delft University of Technology, where he manages the SIMONA Research Simulator facility.

M. M. (René) van Paassen is an associate professor in the section Control and Simulation, Aerospace Engineering, Delft University of Technology, where he received his PhD in aerospace engineering in 1994.

Max Mulder is a full professor in the section Control and Simulation, Aerospace Engineering, Delft University of Technology, where he received his PhD (cum laude) in aerospace engineering in 1999.