THE EFFECT OF SURPRISE ON UPSET RECOVERY PERFORMANCE

Annemarie Landman
Delft University of Technology / TNO
Delft, The Netherlands / Soesterberg, The Netherlands

Eric L. Groen
TNO
Soesterberg, The Netherlands

M.M. (René) van Paassen
Delft University of Technology
Delft, The Netherlands

Adelbert W. Bronkhorst
TNO
Soesterberg, The Netherlands

Max Mulder
Delft University of Technology
Delft, The Netherlands

Introducing the element of surprise is one of the main challenges in simulator training of in-flight emergencies. In this simulator study, we investigated the differences in performance between predictable and surprising circumstances, in order to obtain insight into the transfer of training between predictable training settings and surprising circumstances in operational practice. This was done by testing twenty airline pilots who recovered from an aerodynamic stall in two conditions: one anticipation condition and one surprise condition. All pilots practiced beforehand using predictable, or non-surprising scenarios. The results show that pilots had significantly more difficulties in adhering to components of the FAA-commissioned recovery template in the surprise condition compared to the anticipation condition. These results suggest that predictable training may not be enough to prevent serious performance decrements under surprise.

As surprise and startle are considered to play an important role in a significant proportion of airplane safety events, aviation authorities have mandated the introduction of surprise and/or startle in upset prevention and recovery training (EASA, 2015; FAA, 2015). Both surprise and startle may occur in response to unexpected events, although the former relates specifically to a cognitive mismatch between new information and expectations (Meyer, Reisenzein & Schützwohl, 1997), while the latter refers to a highly physiological, sudden increase in stress (Martin et al., 2015). In the case of surprise, solving the cognitive mismatch (i.e., sensemaking) may be a mentally taxing and difficult task if one is unfamiliar with similar situations. It may require an adaptation, or switch, of one's cognitive "frame" (i.e., reframing; Rankin, Woltjer & Field, 2016). Frames are mental structures within which knowledge and procedures are grouped, and through which information is processed and understood (Klein, Phillips, Rall, & Peluso, 2007). Surprise is an indicator signaling that one's presently active frame is unable to fit with the emerging situation. This mismatch may cause the sensation of a loss of "grip" on the situation, and the desire to explain and understand it. If the surprising situation is also startling or threatening, the concomitant stress can be expected to impede the top-down, or goal-directed process of reframing (Eysenck, Derakshan, Santos & Calvo, 2007), which may further impair one's ability to respond quickly and appropriately (Landman, Groen, Van Paassen, Bronkhorst & Mulder, submitted). In contrast, when an upcoming event is anticipated, sensemaking can occur beforehand. The event is then immediately understood and less stress-evoking, which facilitates a quick and appropriate response.

It follows that if a procedure is only trained in highly anticipated conditions, the sensemaking activities that would be needed to identify and understand the situation before the procedure can be applied in operational practice, are never really practiced. The current simulator study aimed to test whether performance on a learned recovery procedure indeed suffers in surprising compared to anticipated conditions. In addition, the experiment aimed to test whether surprise can be used in simulator training to provide more challenging and realistic scenarios.

Several simulator studies have been published in which the effect of surprise was tested on pilot performance of learned procedures. One study (Schroeder, Bürki-Cohen, Shikany, Gingras & Desrochers, 2014) showed that adherence to a recovery template suffered when pilots were unexpectedly exposed to a previously practiced upset (aerodynamic stall). However, the experiment did not include a control condition, in which the pilots' performance was re-tested in a non-surprising scenario. Another relevant study showed that response times were longer when a stall was pilot-initiated versus when it was unannounced (Casner, Geven & Williams, 2013). However, this study did not include a detailed analysis of performance. The current study adds to these previous studies by comparing the effect of surprise to that of anticipation (manipulation check), while measuring several aspects of adherence to the recovery template.

Method

Participants

Twenty male airline pilots participated in the study (mean age = 36.3 years, SD = 7.88; mean flying experience: 12.4 years, SD = 5.05; 6986 flight hours, SD = 3804). Experience in operating medium-size twinjet aircraft types was required. Eight pilots had mainly experience with the A330, five with the B737, six with the E190 and one with the A320. All pilots were employed at the time of the experiment, and had been on duty at least once in the week prior to the experiment. Five were currently employed as captains, eleven as first officers and three as second officers. The pilots provided written informed consent prior to participation and the ethics committee of the TNO Soesterberg research institute approved the experiment.

Apparatus

The experiment was performed in the DESDEMONA flight simulator (manufactured by AMST Systemtechnik), located at TNO Soesterberg. The cockpit mockup was styled after the Boeing 737NG, and included the primary flight display, navigation display, engine-indicating and crew-alerting system, and a partial mode control with flight director and autopilot mode controls. There was no overhead panel or flight management system. Controls consisted of a yoke with control loading on pitch, rudder pedals with rudder limiter, throttles and a stabilizer with electric trim. Flaps and speed brake were not used. The aerodynamic model was derived from the SUPRA project (Groen et al., 2012), which extended the aerodynamic envelope of transport category aircraft (e.g., Boeing 737NG, Airbus A321, into high angles of attack.

Tasks and Conditions

Pilots were instructed beforehand that the simulator session would comprise two subsequent sections of circa 20 minutes: one upset recovery section and one spatial disorientation section. They were told that both sections were aimed at testing the simulator fidelity. In reality, the first section was used for practice of stall recoveries, while the second section did not take place as described: it was made up to manipulate the pilots' expectation before the test conditions. Figure 1 displays an overview of the experimental design. The order of the conditions was counterbalanced between pilots, and the two resulting groups were analyzed as one.

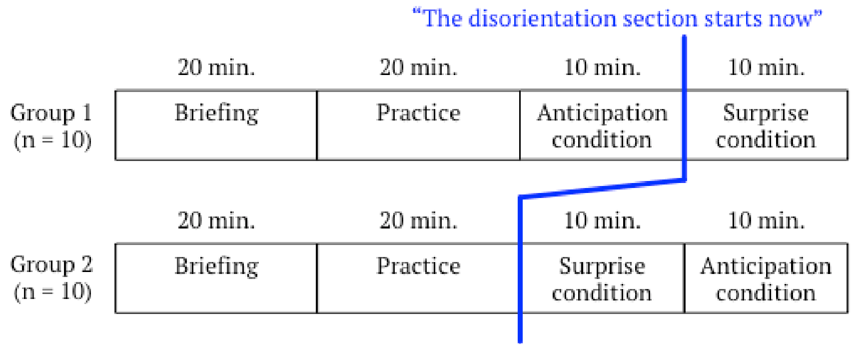

Figure 1. The experimental design, counterbalancing the two test conditions between two pilot groups. The groups were added together for analysis.

Pilots received verbal instructions about the simulated aircraft model and the stall recovery template as advised by the FAA (2015, p. 2. This involves the following steps: "1. Disconnect the autopilot and autothrottle/autothrust systems. 2a. Apply nose down pitch control until impending stall indications are eliminated. 2b. Use nose down pitch trim as needed. 3. Bank wings level. 4. Apply thrust as needed. 5. Retract speed brakes or spoilers. 6. Return the aircraft to the desired flight path." All pilots indicated that they were familiar with these steps. The practice session, aimed at decreasing inter-individual differences in skill level, consisted of the recovery of four different aircraft upsets and four different aerodynamic stall events. All scenarios in this training session were presented in a highly predictable and non-surprising manner, i.e., announced and explained beforehand. The final scenario of these was repeated until the pilot was able to push down quickly and forcefully enough to avoid stick shaker events, while also avoiding overspeed or excessive g-loads.

Unbeknownst to the pilot, the practice session transitioned into the testing section in which the same aerodynamic stall was presented in a surprise condition and in an anticipation condition. Each test condition was preceded and followed by three minutes of manual straight and level flight, with autothrottle on, at 5,000 ft. and with 220 knots. In the anticipation condition, pilots were told beforehand that the simulator operator would bring them into an aerodynamic stall when a certain landmark was crossed. The stall was induced by creating a strong "tailwind" (decreasing the calibrated airspeed (CAS) with 15 knots per second for five seconds), and by simultaneously adjusting the pitch trim up with 24 percent in 3 seconds time. None of the pilots reported afterwards that they were aware of any changes in pitch trim. In the surprise condition, exactly the same stall event was induced about five seconds before another landmark was crossed. In this case, however, pilots were instructed that the spatial disorientation section of the experiment had started, and that they would have to do a climb-out above the landmark while paying attention to any over-pitch sensation. In addition, their attention was taken away from the display at the moment the stall was initiated, as they were asked to give a rating on a sickness scale displayed to their lower right.

Outcome Variables

Flight parameters were logged from the simulator at a sample rate of 100 Hz. These flight parameters were twice (forth and back) low-pass filtered using a 2^{nd} order Butterworth filter with a cut-off frequency of 2 Hz. A script was used to determine the moment of tailwind onset and the start of several control actions in terms of autothrottle, trim, pitch, aileron and rudder control inputs. Since pitch was continuously adjusted before stall onset, a deviation above 5 SD from the mean (taken from the preceding period of straight and level flight) was determined to be a significant pitch control input. Using these data, we checked whether pilots did or did not meet each of the four performance criteria (see, Table 1).

Table 1.

Description of the four measured performance criteria, with the corresponding FAA (2015, p. 2) recovery template principles.

Variable	Corresponding FAA principle	Description
C1. Disengage autothrottle first	1	Disengage the autothrottle at least 2.0 s before significant yoke or pedal inputs.
C2. Start with pitch down control	2a, 3	Give priority to pitch down control by starting the recovery with pitch down control inputs. Strong aileron inputs (> 50% of max) may not occur at around the same moment (within 2.0 s) of pitch down control to meet this criterion.
C3. Unload sufficiently	2a, 6	Respond (within 2.0 s) to stick shaker events with significant pitch down control and maintain significant pitch down control during stick shaker activation. Or, apply sufficient pitch down control to avoid any stick shaker events. Keep the aircraft sufficiently unloaded until CAS increases in order to avoid secondary stick shaker events. Stick shaker events were defined as secondary if they occurred subsequent to an earlier stick shaker event, or if they occurred after the first unloading action, i.e., following the first peak of pitch down control.
C4. Apply pitch down trim	2b	Using the pitch trim to aid in pitch down control during the recovery.

Besides measuring adherence to the recovery template, we performed a manipulation check by asking the pilots to rate their level of surprise caused by the tailwind on a 0-10 point Likert type scale. This was done after both conditions had ended, so as to not cause suspicion about the goal of the experiment in the second condition.

Statistical Analysis

The effect of Condition (anticipation or surprise) on the binary performance variables, i.e. meeting the criteria, was tested using generalized estimating equations (GEE) models of logistic regression. To protect against an overestimation of significant differences, the outcomes were corrected using Holm's sequential Bonferroni correction for multiple comparisons.

The effect of Condition on the pilots' subjective level of surprise was tested with a paired-samples T-test.

Results

Table 2 provides an overview of the statistical differences between conditions for each of the four performance criteria that were measured. All differences are statistically significant, with effect sizes (d) varying from medium to large in strength, i.e., in or above the range of 0.5 to 0.8. Despite the verbal instructions beforehand, it seems that meeting the criteria in the anticipated condition was already quite difficult, as the proportion of pilots who adhered to the criteria was around 50-80%. Nevertheless, in the surprise condition the proportion of pilots who met the criteria decreased with 20-30%, and with 50% in the case of ‘start with pitch down control’. In sum, the surprise manipulation caused a significant decrease in adherence to several aspects of the recovery template.

The manipulation check confirmed that surprise was significantly higher in the surprise condition, 8.44 points, *SD* = 1.50, than in the anticipation condition, 1.39 points, *SD* = 2.00, Δ = 7.06, t = 12.35, $p < .001$. This difference constituted a large effect size, i.e., Cohen’s *d* = 3.99.

Table 2.
Criteria met in the two conditions. Effect sizes (d) are calculated by transforming the odds ratio, i.e. exp(B) from the GEE analysis, conform: Chinn (2000).

	Anticipation (met/unmet)	**Surprise (met/unmet)**	**N[a]**	**Δ[a]**	X^2	***p***	**Cohen’s *d***
C1: Disengage autothrottle early	11/9	6/14	20	-5*	5.10	.024	.69[b]
C2: Start with pitch down control	16/4	6/14	20	-10*	13.41	< .001	1.23[b]
C3: Unload sufficiently	10/10	5/15	20	-5*	3.94	.047	.61[b]
C4: Use trim	9/11	3/17	20	-6*	7.07	.008	.85[b]

* Significant after Holm-Bonferonni correction.

Discussion

The results of this simulator experiment show that the unexpectedness of an simulated aerodynamic stall effectively surprised the pilots, and negatively affected their recovery performance as measured using four criteria derived from the FAA stall recovery template. Our results suggest that significant decreases in adherence to learned procedures can be expected in operational practice compared to predictable training conditions, which is in line with previous similar studies. Our control condition shows that there is indeed a decrease in performance when pilots are surprised, thereby adding to the study of Schroeder et al., 2014. Also, adding to the study of Casner, Geven and Williams, 2013, our detailed measuring of performance indicates that several aspects of the recovery procedure were not followed in the surprise condition, while our manipulation check suggests that this was caused by our manipulation of surprise.

In the surprise condition, pilots were particularly more likely to incorrectly start their recovery with aileron control inputs. Perhaps this was influenced by an increase in lateral instability (wing drop) due to later responses in the surprise condition. Nevertheless, it suggests that ignoring a change in roll angle and giving priority to pitch control (step 1 in the FAA template) may be highly counter-intuitive and difficult to suppress, meaning that proper adherence to the FAA template in surprising situations may be very difficult.

Overall, the results suggest that skills trained under predictive conditions may not transfer to conditions containing an element of surprise. Hence, it may be useful to induce an element of surprise or unpredictability when designing training methods, which allows pilots to practice with sensemaking and switching frames and to increase their “cognitive flexibility” (Kochan, 2005). This conclusion is in line with ICAO’s recommendation for scenario-based recurrent training: “Wherever possible, consideration should be given towards variations in the types of scenario, times of occurrences and types of occurrence, so that pilots do not become overly familiar with repetitions of the same scenarios.” (2013, II-1-5). By practicing procedures under less predictable conditions, trainees learn to use the information presented by the situation itself to identify problems, and to apply solutions. Indeed, in the training domain it has been shown that experiencing examples of a concept in different contexts may strengthen the understanding (frame) relating to this concept and is thought to increase the trainee's ability to apply the concept in similar or novel situations (Van Merriënboer, 1997).

References

Casner, S. M., Geven, R. W., & Williams, K. T. (2013). The effectiveness of airline pilot training for abnormal events. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 55*(3), 477-485.

European Aviation Safety Authority. (2015). Loss of control prevention and recovery training (Notice of proposed amendment 2015-13). Retrieved from https://easa.europa.eu/system/files/dfu/NPA%202015-13.pdf.

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: attentional control theory. *Emotion, 7*(2), 336-353.

Federal Aviation Administration. (2015). Stall Prevention and Recovery Training. (advisory circular no. 120-109A). Retrieved from https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_120-109A.pdf.

Groen, E., Ledegang, W., Field, J., Smaili, H., Roza, M., Fucke, L., ... & Grigoryev, M. (2012). SUPRA-Enhanced upset recovery simulation. In *Proceedings of the AIAA Modeling and Simulation Technologies Conference,* (pp. 2-5).

International Civil Aviation Organization. (2013). Manual of Evidence-based Training. (Doc 9995 AN/497). Retrieved from http://www2010.icao.int/SAM/Documents/2014-AQP/EBT%20ICAO%20Manual%20Doc%209995.en.pdf.

Klein, G., Phillips, J. K., Rall, E. L., & Peluso, D. A. (2007). A data-frame theory of sensemaking. *Expertise out of context: Proceedings of the sixth international conference on naturalistic decision making,* (pp. 113-155). New York, NY, USA: Lawrence Erlbaum.

Kochan, J. A. (2005). *The role of domain expertise and judgment in dealing with unexpected events* (Doctoral dissertation, University of Central Florida Orlando, Florida).

Landman, A., Groen, E. L, Van Paassen, M. M., Bronkhorst, A. W., Mulder, M. (submitted). *Dealing With Unexpected Events on the Flight Deck: a Conceptual Model of Startle and Surprise.*

Martin, W. L., Murray, P. S., Bates, P. R., & Lee, P. S. (2016). A Flight Simulator Study of the Impairment Effects of Startle on Pilots During Unexpected Critical Events. *Aviation Psychology and Applied Human Factors, 6,* 24-32.

Merriënboer, van, J. J. G. (1997). *Training complex cognitive skills: A four-component instructional design model for technical training.* Educational Technology Publications, Englewood Cliffs, New Jersey.

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.

Rankin, A., Woltjer, R., & Field, J. (2016). Sensemaking following surprise in the cockpit—a re-framing problem. *Cognition, Technology & Work, 18*(4), 623-642.