2007

Skill Decay on Takeoffs as a Result of Varying Degrees of Expectancy

Susan M. Stevens

Timpthy E. Goldsmith

SKILL DECAY ON TAKEOFFS AS A RESULT OF VARYING DEGREES OF EXPECTANCY

Susan M. Stevens
University of New Mexico
Albuquerque, New Mexico

Timothy E. Goldsmith
Peder J. Johnson
John B. Moulton
University of New Mexico
Albuquerque, New Mexico

It is generally assumed that skills trained and assessed in a simulator will transfer to the line. However, there is a class of maneuvers that demand an immediate response to an unexpected event (e.g., rejected takeoffs) for which such transfer can be questioned and for which there is little or no empirical data to support a transfer assumption. Thus, we have completed a series of studies aimed at investigating the effects of expectancy on performance for unanticipated events in a laboratory situation with undergraduate college students and experienced pilots. Our participants were trained on both normal and rejected takeoffs and the expectancy for a rejected takeoff was manipulated in each study. There were two primary measures of performance on rejected takeoff trials: the amount of time it took the participant to close down the throttle after engine failure and the maximum deviation from center line achieved while bringing the aircraft to a stop. T-tests indicated that there was a significant degradation in throttle performance for both studies (all ps<.05) and in maximum deviation from center line performance for one of the studies (p<.001). Thus, it is questionable whether the assumption that performance on events that occur in high expectancy conditions will transfer to low expectancy conditions is valid.

Introduction

Skill decay refers to the loss of trained or acquired skills (or knowledge) after a period of nonuse (Arthur et al., 1998). The literature on skill decay has suggested that individuals in advanced fields requiring complex skills (such as aviation) may suffer skill decay due to various reasons including nonuse, heavy cognitive load, poor situational awareness and stress (Arthur et al., 1998; Childs & Spears, 1986; Lee & Liu, 2003; Wickens, 2002). Childs & Spears (1986) found that cognitive/procedural skills decayed more rapidly than did other types of skills in pilots. Lee & Liu (2003) found that pilots had a significantly higher workload for takeoff and landing phases of flight. In addition, they found that the temporal demands for takeoffs were rated significantly higher than any other phases of flight, which the researchers noted could impair pilot performance and lead to errors.

Skill decay and resulting errors may be especially likely to occur in unexpected situations. Wolfe et al (2005) found that their participants failed to detect key targets in a visual task when they only appeared infrequently but not otherwise. This finding has particular implications for pilots in emergency situations that require an immediate response (e.g., rejected takeoff; RTO). One type of RTO occurs when an engine fails prior to rotation speed. The pilot must abort the takeoff and quickly stop the plane on the runway. Although commercial pilots receive extensive training and evaluation on emergency maneuvers in simulators, they may fly their whole careers without actually experiencing one. How does their performance on an unexpected emergency event on the line compare with their performance on the same event during a simulator evaluation?

In a series of studies, we investigated the effects of expectancy on performance for an event (i.e. RTO) in a laboratory task with both undergraduate college students and experienced pilots. Participants were first trained to perform both normal and rejected takeoffs on a PC-based flight simulator. We then tested their performance on a RTO during a subsequent test session where we attempted to decrease the expectancy of a RTO by providing a long series of normal takeoffs (NTOs). We hypothesized that performance on the unexpected event would decrease relative to their session one performance. The first study tested inexperienced undergraduate students.

Study 1

Method

Participants. A total of one hundred forty seven undergraduate students from the University of New Mexico served as participants. The students received course credit for their participation. None of the students were pilots or had experience flying simulators.

Materials. Microsoft's Flight Simulator 2004 displayed on a personal computer was used to create a well controlled laboratory task to study performance on RTO's. The aircraft was a Boeing 737. The task achieved at least some level of realism. Flight Simulator was supplemented with (a) Adventure Basic Language Software Development Kit (ABL) to control and monitor the flights, (b) Flight Recorder for Microsoft Flight Simulator (FLTREC 9.0) to record the flight parameter data, and (c) Perl to run and coordinate the various simulation components. Participants used a yoke (CH Products USB LE Flight Simulator Yoke FSY208LE 200-608) and rudder pedals (CH Products PRO Pedals USB PP99USB 300-111) to control the aircraft. Instructions were played via Windows Media Player and were recorded (via Audacity) in a male voice. There were 4 different sets of instructions; each set described what the participant was to do in the various phases of the experiment.

Procedure. The participants were tested individually in a small room. The first set of instructions was played (via Media Player) informing the participant about the study and asking him to sign an informed consent. After signing the informed consent, the second set of instructions described the training session. These instructions explained the function of several cockpit displays and controls of the aircraft in Flight Simulator (i.e. yoke, throttle, rudder pedals, air speed indicator, attitude indicator and altimeter). The experimenter pointed out the different controls as the instructions described them. Next, the participant was informed that he would perform a series of NTOs. A normal takeoff consisted of taking off and then turning the aircraft. Each trial started with the plane on the runway, engines running and ready for takeoff. The altitude of the runway was 440 feet and the heading was 160 degrees. Once the participant released the brake and engaged the throttle, a text message was displayed indicating the takeoff speed (either 145 or 155 knots), the turn altitude (the minimum altitude at which the turn should be initiated; always 1000 feet) and the heading (140 degrees for a left turn or 180 degrees for a right turn). The takeoff speeds and turn direction were randomized and balanced across the 40 flights. The participant was asked to stay as close to the centerline as possible while traveling down the runway. After takeoff, the participant was told whether or not he took off at the correct speed. After the specified heading was reached, the program stopped and reset itself for another trial. The participant performed 10 NTOs.

Once the participant completed the NTOs, a third set of instructions described how to perform a RTO. A RTO trial began the same as a NTO trial but before the plane reached the designated rotation speed, one of the engines failed at 120, 125 or 135 knots. The speed at which an engine failed and which engine (left or right) failed was randomized and balanced across the trials. The participant was informed to close down the throttle as quickly as possible and then to steer and brake the plane to keep it as close to the centerline as possible. Once the plane came to a complete stop, the program stopped and reset itself for another trial. The participant performed 15 RTOs. One average, session one took forty five minutes to complete. After the participant completed these training trials, he was told he could take a break if he desired (most people declined).

A final set of instructions informed the participant about the next session. The participants were divided into three groups that received different instructions; this was the manipulation of expectancy. The first group (group A; N = 62) was simply told that the next set of flights would be similar to the previous. No other information was given. The second group (group B; N = 42) was told that they would be performing another set of takeoffs but would not know when an engine failure might occur. The third group (group C; N = 43) heard the same instructions as the second group but was also warned prior to each test RTO that a RTO would occur on that trial. All of the participants performed 15 trials in this last session; 13 NTOs followed by 2 RTOs.

Results

Several measures of the participants' performance on the RTOs were obtained (e.g. stopping time, stopping distance) but response time to shut down the throttle after engine failure and maximum distance off center line were used for statistical analysis. Extreme throttle and off center values were scaled such that throttle values above 6 seconds were eliminated and center line values over 65 feet were scaled to 65 feet. As a result of the throttle scaling, three participant's data from group A and one participant's data from group B were not included in the analysis. The last 5 RTOs during training were averaged for each participant to establish baseline performance measures. These baseline measures were compared to the participant's performance on first RTO of the test session. The results are shown in Tables 1 and 2. For group A, a paired t-test revealed a significant throttle effect $t(58) = 5.38$, $p<.001$ and a significant center line effect $t(61) = 3.05$, $p<.001$ indicating that the participants performed significantly worse during the unexpected RTO in the second session. Additional t-tests for the group A throttle effect using three alternate scaling methods (all latencies over 6 scaled to 6, removal of all participants with latencies over 6 and a log transform) produced similar results. For group B, a paired t-test revealed a significant throttle effect $t(40) = 4.91$, $p <.001$ but no significant center line effect $t(41) = .16$, $p>.05$. For group C, no significant throttle effects were found $t(42) = .92$, $p>.05$. Significant center line effects were found $t(42) = -3.13$, $p<.001$ but were in the opposite direction indicating that the participants' center line performance in the test session was significantly better than their baseline performance.

Table 1. Mean response time to shut down throttle after engine failure in seconds

Study	Baseline	Test RTO	Cohen's d	p
	Mean (std)			
1A	1.29 (.68)	2.07 (.95)	0.94	<.001
1B	1.07 (.47)	1.62 (.75)	0.88	<.001
1C	1.09 (.47)	1.11 (.47)	0.04	ns
2	.74 (.19)	.99 (.57)	0.58	<.05

Table 2. Maximum off center line distance in feet

Study	Baseline	Test RTO	Cohen's d	p
	Mean (std)			
1A	28.98 (12.59)	37.00 (20.54)	0.47	<.001
1B	26.07 (13.25)	25.59 (18.02)	0.03	ns
1C	24.47 (11.35)	19.47 (14.72)	0.38	<.001
2	13.79 (6.31)	15.81 (14.92)	0.12	ns

Study 1 showed a significant degradation in performance on an unexpected RTO in comparison to performance demonstrated during training. We wondered to what extent the results might be limited to naïve participants. Study 2 extended the work to experienced pilots.

Study 2

Method

Participants. Fourteen pilots from Double Eagle Airport in Albuquerque, NM served as participants. They were compensated for their participation. The pilots had an average of 720 hours of flying experience.

Materials. The materials were the same as those used for Study 1.

Procedure. The procedure was the same as in Study 1 except that the pilots received a shorter version of the first set of instructions due to their familiarity with the equipment. The pilots received the same final set of instructions as group A in Study 1.

Results

The data were scaled in the same way as Study 1. None of the pilots had extreme throttle scores so all were retained in the analysis. Performance was computed in the same way as Study 1 (see Tables 1 and 2). A paired t-test revealed a significant throttle effect $t(13) = 2.18$, $p<.05$ but an insignificant center line effect $t(13) = .46$, $p>.05$.

Discussion

The results from both studies indicated that performance can significantly degrade as a function of expectancy. Both actual pilots and students were slower to close down the throttle on an unexpected RTO. In addition, the students deviated more from center line in the test RTO condition. The pilots were clearly superior to the students in maneuvering the plane to a stop once an engine failed. However, once they recognized that the engine failed, there was little decrement in their maneuvering skills.

The results are at least suggestive that actual line pilots may not respond as well to emergency events on the line as they do in simulator sessions. Some corroboration of these results comes from studies of first-look maneuver performance of commercial pilots. Hendrickson et al (2006) found that pilots' performance on unexpected first-look maneuvers was slightly worse (but statistically significant) than performance on maneuvers for which they were briefed.

Acknowledgements

This research was funded by the Federal Aviation Administration to the University of New Mexico 00-G-014 and in part by NIMH COR grant T34-MH19101.

References

Arthur, W., Bennet, W., Stanush, P. L. & McNelly, T. L. Factors that influence skill decay and retention: A quantitative review and analysis. *Human Performance, 11*, 57-101.

Childs, J. M. & Spears, W. D. (1986). Flight-skill decay and recurrent training. *Perceptual and Motor Skills, 62*, 235-242.

Hendrickson, S. M. L., Goldsmith, T. E. & Johnson, P. J. (2006). Retention of airline pilots' knowledge and skill. *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*, 1973 – 1976.

Lee, Y-H. & Liu, B-S. (2003). Inflight workload assessment: Comparison of subjective and physiological measurements. *Aviation, Space and Environmental Science, 74*, 1078-1084.

Wickens, C. D. (2002). Situation awareness and workload in aviation. *Current Directions in Psychological Science, 11*, 128-133.

Wolfe, J. M., Horowitz, T. S. & Kenner, N. M. (2005). Rare items often missed in visual searches. *Nature, 435*, 439-440.