5-7-2019

The NASA MATB-II Predicts Prospective Memory Performance During Complex Simulated Flight

Kathleen Van Benthem

Caitlin Shanahan

Chunyun Ma

Adam Fraser

Chris M. Herman

THE NASA MATB-II PREDICTS PROSPECTIVE MEMORY PERFORMANCE DURING COMPLEX SIMULATED FLIGHT

Kathleen Van Benthem, Caitlin Shanahan, Chunyun Ma, Adam Fraser, Chris M. Herdman
Carleton University, Ottawa, Canada

Prospective memory is essential for flight, where failures can result in incorrect flight control settings, leading to loss of life and equipment. Furthermore, prospective memory is highly-sensitive to pilot age, cognition, and experience. This research reports on the relation of the NASA Multi-Attribute Test Battery-II (MATB-II) to prospective memory during simulated VFR flight (N=51). Prospective memory was indexed with specialized radio calls that were associated with non-focal visual cues. Linear regression models examined the relative association of MATB-II variables to prospective memory in low and high workloads. System monitoring, psychomotor tracking, and resource management, generally at higher difficulty levels, were the variables most predictive of prospective memory, r^2 =0.41. Pilot experience improved the model in the high-workload condition. Estimating risk for prospective memory failures via multitasking ability, with a focus on monitoring tasks, may inform cognitive assessment approaches to enhance aviation safety.

Prospective memory is a cognitive construct associated with remembering to perform critical tasks in the future and is a skill relevant to successful daily living, including aviation outcomes (Dismukes & Nowinski, 2007). A retrospective look at aviation incidents and accidents found that 74 of the 75 reports associated with memory errors were, in fact, prospective memory failures (Nowinski, Holbrook, & Dismukes, 2003). Failures of prospective memory in the cockpit, such as forgetting to adjust the flaps or to lower the landing gear, can have disastrous results and can occur with pilots at any level of expertise (Dismukes & Nowinski, 2007). Research has shown that prospective memory is sensitive to mental workload demands during flight and a pilot's ability to detect relevant memory cues in the environment (Van Benthem, Herdman, Tolton & LeFevre, 2015).

Despite its relevance to aviation safety, cognitive assessments designed for pilots have yet to explicitly measure prospective memory. CogScreen-AE (Kay, 1995) and CogState (CogState Ltd., Melbourne, Australia) are comprehensive neuropsychological tests with links to pilot performance, however neither test has demonstrated an association with the risk of prospective memory failures.

We selected the MATB-II as a possible predictor of prospective memory during complex simulated flight due to the similar demands for cue detection inherent in both the MATB-II and prospective memory. The NASA Multi-Attribute Test Battery-II (Santiago-Espada, Myer, Latorella, & Comstock, 2011) is a cognitive screening tool designed for aviators that incorporates planning, vigilance, and monitoring, with varying levels of multitasking requirements.

Method

Participants

Participants were licensed and medically certified aeroplane pilots (or those with permits) who had flown within the last 18 months (N=51). Pilots were recruited from local flying clubs, aviation interest groups, and flight schools via newsletters, posters, and social media. All research activities took place at a university flight simulation research laboratory and were part of a larger research agenda investigating the cognitive health screening and intervention for general aviation pilots.

Table 1. Description of Full Pilot Sample (N=51)

	Minimum	Maximum	Mean	Std. Deviation
Age	17	71	46.3	17.4
PilotLevel	1	6	4.1	1.3
Flighthours	2	12000	1311.0	2592.3
Recent Pilot-in-Command hours	0	582	49.1	96.3
YearsLicensed	1	70	14.4	14.5
Simulator Hours	0	1000	88.7	192.1

Procedure

The experiment took place over two sessions. At the first session, pilots were provided with a briefing of the flight tasks, which were based on the premise of a search and rescue familiarization flight. The flights were conducted in two segments, where pilots had access to an electronic navigational aid for one of the legs. Pilots were provided with a briefing regarding the flight simulator and completed three warm-up flights where they practised performing the required tasks, including the prospective memory task. After the warm-up session, pilots were given a second briefing on the flight plan, and material required to complete the flight and the tasks. Pilots wore a wristband, which collected biometric information such as heart rate, and a lightweight wireless 14-channel electroencephalography headset.

An ongoing peripheral detection task (PDT) was also undertaken by participants. The PDT was designed to measure mental workload. Pilots were provided with one break time to hydrate or rest briefly, if required. During this break pilots were queried regarding their situation awareness. The biometric, PDT, and situation awareness data were not analyzed for this report. The second session included completion of the three difficulty-levels of the MATB-II (counterbalanced presentation order) and a virtual reality flight simulation task using a custom flight control unit and graphics displayed using the Oculus Rift headset (Microsoft) (also part of the larger research agenda). The order of MATB-II versus flight simulation was alternated so that half the participants completed the virtual reality flight first.

Flight Simulation Apparatus

Pilots flew a converted Cessna 172 aircraft simulator. The cockpit displays flight information via a virtual '6-pack' (i.e., the primary flight controls) and was equipped with a yoke, throttle and flaps. The graphics were produced by Prepar3D (Lockheed Martin) "on the fly" and were modeled after Canadian aerodromes and their surrounding terrains. The flight

graphics were displayed on a broad-angle display system consisting of eight theater-quality 1080p projectors and a 14-foot tall, 180-degree curved screen that provided 45 degrees of vertical field of view and 120 degrees of horizontal field of view. The time and the pilot's location, airspeed, heading, bank, pitch, and altitude were recorded at one hertz.

Measures

Prospective memory. The prospective memory task assessed participant's ability to perform radio calls in response to non-focal visual cues. Participants were instructed to alternate between two scripted calls each time they detected the appearance of the cue – a right-facing arrow presented on a screen mounted to the cockpit. This task required that participants remembered to check for the visual cue, and their previous radio call to form appropriate intentions. Prospective memory was measured as the ratio of the number of calls made over the time spent in each leg (leg 1=low workload and leg 2=high workload).

MATB-II multitasking measures. The Multi-Attribute Task Battery II (MATB-II) is a neuropsychological test designed to measure multitasking and mental load (Santiago-Espada et al., 2011). The MATB-II tasks were configured to provide three difficulty levels (low, medium, and high). As illustrated in Figure 1, the MATB-II subtests included system monitoring, psychomotor tracking, communication, and resource management. In the system monitoring module, participants were tasked with clicking on the green light when it went out, clicking on the red light when it appeared, and clicking on the scales when the central indicators deviated significantly from centre. For the tracking module, when manual mode was engaged, the participants used a joystick to keep the reticle inside the central square.

Participants were expected to monitor a task scheduling module to become aware of upcoming tracking tasks. During predefined periods, the communications module intermittently broadcasted messages. Each message was prefixed by a call sign. If the call sign matched the ownship call sign, then participants were to change the frequency of a specific radio. The resource management task required participants to route fuel into two main tanks (top), maintaining the fuel level at the target level indicated by the shaded blue region while circumventing blocked fuel-lines (in red).

For each level of difficulty, the outcome measures from the MATB-II included total occurrences where errors in the system monitoring displays were not detected and responded to (lapses in vigilance), root mean square tracking error, and for resource management, the average amount of time pilots did not maintain adequate fuel levels, and average units of fuel deficit. After the completion of the tasks at each difficulty level, the MATB-II presents a modified NASA Task Load Index (TLX) screen to measure subjective mental workload (Cao, Chintamani, Pandya, & Ellis, 2009). Communication scores were not used as most pilots scored 100% on this task.
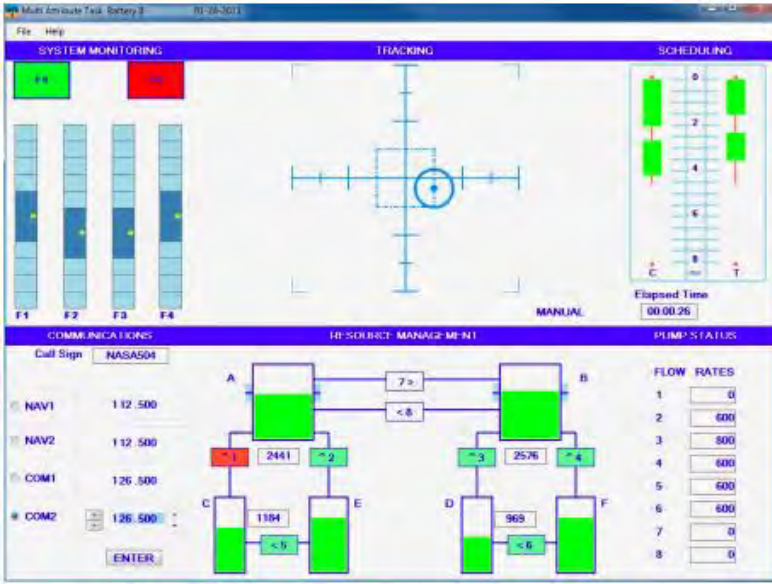
Figure 1. The MATB-II interface. Clockwise from the top-left: system monitoring (lights and scales), psychomotor tracking, scheduling (non-task), pump status (non-task), resource management, and communications.

Results

Pilot Attributes as Predictors of Prospective Memory and the MATB-II

As shown in Table 2, age was significantly associated with prospective memory in the low- but not high-workload condition. In contrast, pilot expertise positively correlated with prospective memory in the high- but not low-workload. The exception to the relationship of expertise and prospective memory was found for number of years licensed, where more years licensed was associated with lower scores. A post hoc analysis showed that the trend towards a deleterious effect of expertise on low workload prospective memory may be due to its conflation with age (e.g., age and years license were significantly correlated, $r = 0.634$, $p < .01$).

Table 2.
Pearson Correlations between Prospective Memory Scores and Pilot Attributes

PM	Age	Years Licensed	Hours Flown	Recent Hours	License Level
Low Workload	**-0.381***	-0.242	-0.231	-0.067	0.099
High Workload	-0.121	0.167	**0.282***	**0.329***	**0.321***

Note. * = $p < .05$ and ** = $p < .01$, two-tailed.

As shown in Table 3, older age was positively correlated with the error scores for all subtests (other than resource management time measure). Pilot level and hours flown did not correlate with any of the MATB-II subtests. However, recent pilot-in-command hours was negatively correlated with system monitoring errors in the medium- and high-difficulty levels.

Years licensed was positively correlated with errors in system monitoring, although this was most likely an artefact of the negative effect that age showed on performance.

Table 3.
Pearson Correlations between MATB-II Subtests, and Pilot Attributes and Prospective Memory

MATB-II Variable	Difficulty Level	Age	Years Licensed	Recent Hours	Low-Workload PM	High-Workload PM
System Monitoring Errors	Low	.293[*]	0.165	-0.173	-.307[*]	-0.266
	Medium	.247[*]	0.108	-.280[*]	-.346[*]	-.483[**]
	High	.393[**]	0.252	-.255[*]	-.514[**]	-.401[**]
Tracking Deviation	Low	0.227	0.122	-0.201	-0.042	-.328[*]
	Medium	.288[*]	0.157	-0.138	-0.143	-.331[*]
	High	.274[*]	0.335*	-0.152	-.379[*]	-0.161
Resource Management (Average Units Under)	Low	.538[**]	0.217	-0.170	-.527[**]	-.314[*]
	Medium	.478[**]	0.223	-0.171	-.463[**]	-.434[**]
	High	.474[**]	0.189	-0.061	-.437[**]	-0.238

Note. PM = prospective memory. * = $p < .05$ and ** = $p < .01$, two-tailed.

Models of Prospective Memory Performance

We developed a hierarchical linear regression model for each level of workload. Each model consisted of five blocks (using the stepwise feature to account for the large number of potential predictors). The first block was age, then experience factors, then each of the three difficulty levels of tracking, system monitoring, and resource management. As shown in Table 4, the final model for low-workload prospective memory was comprised of the high-difficulty scores for tracking, system monitoring, and resource management, $F(4, 39)=7.53$, $p< .001$, $r^2 = 0.38$. The strongest effect was found with system monitoring.

Table 4.
Summary of Multiple Regression Analysis for Low-Workload Prospective Memory

MATB-II Subtest (difficulty level)	B	SE(B)	β	t	Sig. (p)
Tracking Deviations (3)	-0.030	0.015	-0.251	-1.971	0.056
System Monitoring Errors (3)	-0.191	0.074	-0.352	-2.573	0.014
Resource Management Deviation (3)	-0.002	0.001	-0.261	-1.941	0.059

The final model for prospective memory (high-workload) was comprised of similar-sized effects, ranging from $\beta= 0.279$ to -0.295, from the medium-difficulty scores for system monitoring and resource management and pilot license level, $F(3, 41) =7.54$, $p< .001$, $r^2 = 0.31$.

Discussion and Implications

In this work, we quantified the effects of pilot attributes and cognitive functions on prospective memory. Our results support the theory that prospective memory is strongly tied to executive functions, such as multitasking and cue detection (Dismukes & Nowinski, 2007; Kliegel, Martin, McDaniel, & Einstein, 2002; Van Benthem, Herdman, Tolton & LeFevre, 2015). We also found that pilot attributes, such as lower license level and older age, were associated with lower prospective memory and MATB-II subtests. In light of these findings, future tests designed to predict pilot prospective memory will benefit from a design that features executive cognitive functions and multitasking ability where monitoring the environment is integral to the task.

Acknowledgements

References

Cao, A., Chintamani, K. K., Pandya, A. K., & Ellis, R. D. (2009). NASA TLX: Software for assessing subjective mental workload. *Behavior Research Methods, 41* (1), 113–117. doi: 10.3758/BRM.41.1.113.

Dismukes, R., & Nowinski, J. (2007). Prospective memory, concurrent task management, and pilot error. In A. Kramer, D. Wiegmann, & A. Kirlik (Eds.) *Attention: From Theory to Practice*. (pp. 225-236). New York: Oxford.

Kay, G. (1995). *CogScreen Aeromedical Edition Professional Manual (Professional Manual)*. Florida: Psychological Assessment Resources, Inc.

Kliegel, M., Martin, M., McDaniel, M. A., & Einstein, G. O. (2002). Complex prospective memory and executive control of working memory: A process model. *Psychologische Beitrage*, *44*(2), 303–318.

Nowinski, J. L., Holbrook, J. B., & Dismukes, R. K. (2003). Human memory and cockpit operations: An ASRS study. *In Proceedings of the 12th International Symposium on Aviation Psychology* (pp. 888-893). Dayton, OH: The Wright State University.

Santiago-Espada, Y., Myer, R. R., Latorella, K. A., & Comstock, J. R. (2011). *The Multi-Attribute Task Battery II (MATB-II) Software for Human Performance and Workload Research: A User's Guide*, 269. NASA/TM–2011-217164.

Van Benthem, K. D., Herdman, C. M., Tolton, R. G., & LeFevre, J.-A. (2015). Prospective memory failures in aviation: effects of cue salience, workload, and Individual differences. *Aerospace Medicine and Human Performance*, *86*(4), 366–373. doi.org/10.3357/AMHP.3428.2015