*Human Factors in Transportation,* Vol. 95, 2023, 609–619
https://doi.org/10.54941/ahfe1003846

# Why the Stability-Flexibility-Dilemma Should Be Taken Into Consideration When Studying Pilots Multitasking Behaviour

**Sophie-Marie Stasch and Wolfgang Mack**

Universität der Bundeswehr München, Munich, Germany

## ABSTRACT

The ability to execute multiple flight tasks simultaneously is a basic requirement for safe aircraft operation. To the present time, there is no consensus about the degree to which simultaneous task execution is actually possible without performance decrements. The flexibility perspective on multitasking explains how cognitive control enables task sets to be flexibly activated and shielded from interference. However, cognitive control is subject to the stability-flexibility dilemma. This dilemma describes the conflicting demands on cognitive control that influence goal-directed behaviour in multitasking situations. On the one hand, cognitive stability has the advantage of minimizing task interference, while not facilitating flexible goal updating. On the other hand, cognitive flexibility allows for constant background monitoring and facilitates task switching. In addition, it has been demonstrated that overlearned action sequences reduce multitasking costs, but are also accompanied by mitigated behavioural flexibility. However, behavioural flexibility is particularly necessary in novel and complex flight scenarios to ensure a pilot's rapid operational readiness. This issue raises two questions: How does the stability-flexibility-dilemma affect multitasking performance in flight environments? And which control mode is strategically beneficial in which flight scenarios? To answer these questions, the cognitive control mode of 34 subjects was experimentally manipulated in a multitasking flight environment. A gamification method shifted the participants control mode in a more stable and more flexible control mode respectively. Results show not only differences in the performance of the individual flight tasks, but also in the subjective workload and various eye tracking metrics. The latter could be taken into account by a cognitive assistance system to detect the control mode of pilots in real time. It enables appropriate assistance to be provided, taking into account the control mode and situational demands. Ultimately, this leads to the provision of situation-specific assistance with the potential to enhance the overall safety in the cockpit.

**Keywords:** Matb, Eye-tracking, Cognitive control, Pilots, Cognitive assistance systems, Workload

## MULTITASKING IN AVIATION

Handling multiple flight tasks at the same time is one of the most important abilities of a skilled pilot. Multiple systems have to be monitored, communication with air traffic control must be ensured at any given time and the aircraft must be operated safely. In case of sudden interruptions, like system failures, attention must be switched and prioritized to avoid safety-critical errors. While these situations make multitasking behaviour seemingly unavoidable, potentially harmful errors of omission may be the consequence of misdirected attention (Loukopoulos et al., 2009).

But what exactly is multitasking? Salvucci and Taatgen (2008) characterize multitasking in their theory *Threaded Cognition* as a continuum varying between concurrent (parallel) and sequential (serial) multitasking. The distinction between these two states lies within the temporal dimension. While concurrent multitasking refers to task-switching within seconds, sequential multitasking means task switching within hours. Both kinds of multitasking occur in the cockpit. For instance, pilots have to aviate and manage the navigation at any time (concurrent multitasking), while only occasionally responding to radio messages or operating the landing gear (sequential multitasking).

Koch et al. (2018) organize the multitasking research into three perspectives: a structural view, relating multitasking to the cognitive bottleneck, a plasticity view, meaning the influence of training on multitasking abilities and a flexibility view that relates cognitive control processes to multitasking behaviour. While multitasking in aviation has been extensively studies from a structural view (Antosko and Lipovsky, 2022; Barron and Rose, 2017; Morgan et al., 2013) and a plasticity view (Koglbauer, 2015; McLean et al., 2016), the following study takes a flexibility view on multitasking. The flexibility perspective explains how cognitive control, a core component of executive functions (Diamond, 2013), enables task sets to be flexibly activated and shielded from task interference in multitasking scenarios. This ability is especially important in emergency and high workload situations. Pilots need to prioritize the most safety-critical tasks and adapt efficiently to the rapidly changing circumstances. The famous A-N-C-axiom (Aviate – Navigate - Communicate) indicates one task prioritization strategy of handling highly demanding multitasking scenarios. One would assume that cognitive flexibility has only a beneficial impact on multitasking performance in these instances, because it's the ability to adapt cognitive processing strategies to novel conditions in the environment (Canas et al., 2003). However, cognitive flexibility is also subject to the stability-flexibility-dilemma of cognitive control.

## The Stability-Flexibility-Dilemma

The stability-flexibility-dilemma describes the antagonistic demands of cognitive control on multitasking behaviour (Musslick et al., 2018). While cognitive flexibility is associated with flexible task switching and facilitated monitoring of potentially relevant signals in the cockpit, this state is also linked to distractibility and impulsivity. Complementary, cognitive stability is accompanied by enhanced goal shielding and target-orientated stimulus selection. However, cognitive stability is also linked to reduced background monitoring, reducing the ability to detect potential harmful incidents in time.

**Table 1.** Advantages and disadvantages of cognitive control (Goschke, 2017).

| Attributes | Cognitive Flexibility | Cognitive Stability |
|---|---|---|
| Advantages | Flexible task switching<br>Monitoring of potentially relevant stimuli | Goal shielding<br>Target-oriented stimulus selection |
| Disadvantages | Distractibility<br>Impulsivity | Reduced background monitoring<br>Difficult task switching |

Both states should not be conceived of as distinct entities. Instead, they lie on a continuum. With the advantages and disadvantages of each control mode in mind (see Table 1), it arises the question how these two states are regulated.

Eppinger et al. (2021) propose that the updating threshold is a key factor of meta-control, the process regulating the stability-flexibility-dilemma. A low threshold leads to an easy passage of information into working memory. Hereby, task switching is facilitated at the expense of distractibility. Therefore, a low threshold is associated with high cognitive flexibility. On the contrary, a high updating threshold shields distracting information efficiently from entering the working memory. Consequently, new information can pass this threshold only with difficulty. It leads to a reduced ability to switch tasks and monitor background information. A high updating threshold is therefore linked to cognitive stability.

Other meta-control parameter have been proposed, such as contextual demands (Siqi-Liu and Egner, 2020), the immediate reward history (Dreisbach and Fröber, 2019) and positive affect (Goschke and Bolte, 2014), but will not be discussed in this article.

While these insights originate from basic research in cognitive psychology, it remains to be clarified to what extent the stability-flexibility dilemma can be transferred to multitasking scenarios in the cockpit. With a better understanding of how the stability-flexibility-dilemma influences flight performance, safety-critical errors like errors of omission or distraction could be prevented more easily. From a technical point of view, a cognitive assistance system could take the control mode of the pilot into account and compensate for the associated disadvantages of each control mode.

To accomplish this goal, two fundamental objectives need to be addressed initially: First of all, the influence of the stability-flexibility-dilemma on performance must be determined. Secondly, correlates of each control mode must be identified to be able to diagnose the control mode in the cockpit in real time.

## CURRENT STUDY

The current study examines the influences of the stability-flexibility-dilemma on flight task performance and subjective workload. In addition, eye-tracking metrics are explored as correlates of the cognitive control mode. Eye-tracking is highly suitable for this purpose, because task switches and the distribution of visual attention can easily be determined by fixation rates. The distribution of ambient/focal visual attention is computed by the Coefficient K

(Krejtz et al., 2016), whereas positive values indicates focal visual processing and negative values relate to ambient visual processing.

### Method

The openMATB (open Multi-Attribute Task Battery) served as the experimental flight environment (Cegarra et al., 2020), consisting of a tracking task, a system monitoring task, a communication task and a resource management task. The control mode of each participant was manipulated via a gamification method in a counterbalanced within-subject design. The control mode was primed via instruction, either by requiring all tasks to be performed with equal priority (flexible condition) or by prioritizing the tracking task (stable condition). After each trial participants received a feedback score between 0 and 100. This feedback score was calculated based on the number of fixations on each MATB task. The constant feedback after each trial served as a reinforcement mechanism for the respective control mode. Participants were instructed to achieve a score as high as possible. A high score table additionally addressed the subjects' motivation to pursue the corresponding control mode. Every participant completed two different scenarios in each control mode to exclude scenario related effects. This resulted in four experimental conditions (Scenario A – flexible, Scenario A – stable, Scenario B – flexible, Scenario B - stable). Each condition consisted of five trials (each takes 90 s). After each condition, participants filled in the NASA-TLX (Hart and Staveland, 1988). The Eyelink 1000 Plus (SR Research Ltd., Ottawa, ON, Canada) in head-fixed mode recorded the participants eye movements.

### Participants

Thirty-four students and employees from the Universität der Bundeswehr München were tested ($M_{Age} = 26.35$, $SD_{Age} = 4.68$). The sample consists of 41% female participants and 59% male participants. Sample size was calculated based on a G*power analysis (Faul et al., 2007) assuming a medium effect size (Cohen's d = 0.49) of van Steenbergen et al. (2009) for the effect of reward on a cognitive control task.

## DATA ANALYSIS AND RESULTS

Data analysis was conducted by performing Bayesian Paired Samples T-Test with the software JASP (JASP Team, 2022). Bayes Factor interpretation is based on Andraszewicz et al. (2015). Results were averaged across conditions with the same cognitive control mode.

### Feedback Score

Participants received better feedback scores in the flexible condition (M = 65.66, SD = 9.29) than in the stable condition (M = 35.74, SD = 18.00) with extreme evidence for this result (BF = 331000). Moreover, the improvement of scores from trial to trial was higher in the stable condition (M = 8.35, SD = 11.33) than in the flexible condition (M = 0.83, SD = 3.45) with extreme evidence (BF = 205).

## Performance

Performance of the MATB sub-tasks was operationalized by assessing the respective task errors. For the tracking task, the root-mean-square error was calculated. For the system monitoring task and the communication task, the number of misses were counted respectively. The deviation from the optimal fuel level was computed for the resource management task. Each performance metric was z-standardized to allow comparison between the single task performances. Overall MATB performance was calculated by summation of the single task z-scores. For better interpretation, the z-scores were reversed by a sign change, so that a higher z-value corresponds to a higher task performance.

Results indicate improved task performance in the system monitoring task in the flexible condition compared to the stable condition ($M_{Diff} = -0.27$, $SD_{Diff} = 0.61$) with moderate evidence ($BF_{10} = 3.39$). Performance of the communication task was also improved in the flexible condition compared to the stable condition ($M_{Diff} = -0.33$, $SD_{Diff} = 0.62$) with a Bayes Factor indicating strong evidence ($BF_{10} = 10.09$). Participants showed better task performance in the resource management task in the flexible condition compared to the stable condition ($M_{Diff} = -0.21$, $SD_{Diff} = 0.41$) with moderate evidence ($BF_{10} = 7.60$). For the tracking task, participants showed better performance in the stable condition compared to the flexible condition ($M_{Diff} = 0.40$, $SD_{Diff} = 0.52$) with extreme evidence ($BF_{10} = 365.72$). The overall task performance was slightly better in the flexible condition ($M_{Diff} = -0.41$, $SD_{Diff} = 1.16$) than in the stable condition. The Bayes Factor ($BF_{10} = 1.18$) shows anecdotal evidence for this finding.

## Workload

Overall workload was lower in the flexible condition (M = 12.52, SD = 1.58) than in the stable condition (M = 13.07, SD = 1.46). Bayes factor analysis revealed moderate evidence for this finding ($BF_{10} = 6.77$). Additional analyses of the single NASA-TLX-Dimensions were conducted. Results indicate that participants perceive the mental demand as lower in flexible condition (M = 15.38, SD = 2.54) than in the stable condition (M = 16.40, SD = 2.74) with anecdotal evidence ($BF_{10} = 2.80$). Physical demand was rated lower in the flexible (M = 8.54, SD = 4.53) condition compared to the stable condition (M = 9.56, SD = 4.80) with moderate evidence ($BF_{10} = 8.23$). Furthermore, temporal demand was perceived as lower in the flexible (M = 14.12, SD = 3.17) than in the stable condition (M = 15.12, SD = 3.55) with moderate evidence ($BF_{10} = 3.36$). Participants perceived the own performance as lower in the stable condition (M = 8.62, SD = 4.57) than in the flexible condition (M = 11.40, SD =3.65) with extreme evidence ($BF_{10} = 109.82$). Moreover, effort was rated as lower in the flexible condition (M = 14.90, SD = 2.93) than in the stable condition (M = 15.96, SD = 2.48) with moderate evidence ($BF_{10} = 9.87$). Participants were less frustrated in the flexible condition (M = 10.76, SD = 3.03) compared to the stable condition (M = 12.79, SD = 3.74), with a Bayes Factor showing strong evidence for that finding ($BF_{10} = 17.24$).

### Eye-Tracking Metrics

Different eye-tracking metrics were calculated to investigate visual correlates of each control mode. Areas of Interest (AOI) were defined according to the respective MATB tasks. Fixations and saccades were computed using the EyeLink Data Viewer software package (SR Research Ltd., version 4.3).

### AOI Specific Metrics

Participants looked on average per fixation longer on the communication task in the flexible condition (M = 212.30, SD = 65.55) than in the stable condition (M = 196.82, SD = 60.33) with strong evidence for this finding ($BF_{10}$ = 29.72). The number of fixations was higher in the flexible condition (M = 426.06, SD = 110.08) compared to the stable condition (M =369.79, SD = 137.24) with very strong evidence (BF10 = 63.96).

Mean fixation duration was increased in the flexible condition (M = 209.43, SD = 47.75) compared to the stable condition (M = 201.22, SD = 48.26) for the system monitoring task with anecdotal evidence (BF10 = 2.88). Number of fixations was also increased in the flexible condition (M = 450.50, SD = 165.83) compared to the stable condition (M = 323.38, SD = 159.73) with extreme evidence ($BF_{10}$ = 771752.62).

Concerning the resource management task, the mean fixation duration was higher in the flexible condition (M = 175.66, SD = 36.12) than in the stable condition (M = 175.25, SD = 34.59) with strong evidence for this result ($BF_{10}$ = 12.51). The number of fixations was also higher in the flexible condition (M = 787.97, SD = 219.16) compared to the stable condition (M = 622.91, SD = 257.83). Bayes Factor analysis shows extreme evidence for this outcome ($BF_{10}$ = 15262.94).

For the tracking task, the mean fixation duration was lower in the flexible condition (M = 314.13, SD = 75.60) compared to the stable condition (M = 393.47, SD = 138.04) with extreme evidence ($BF_{10}$ = 144.02). The number of fixations was lower in the flexible condition (M = 1024.12, SD = 198.59) compared to the stable condition (M = 1095.88, SD = 215.12) with anecdotal evidence for this result ($BF_{10}$ = 2.78).

### Overall Metrics

Participants switched tasks more often in the flexible condition (M = 1002.32, SD = 231.50) than in the stable condition (M = 871.74, SD = 247.07). Bayes Factor analysis report extreme evidence for this finding ($BF_{10}$ = 879.45). The Coefficient K was increased in the stable condition (M = 0.15, SD = 0.35) compared to the flexible condition (M = −0.08, SD = 0.16) with extreme evidence for this result ($BF_{10}$ = 762.45). Difference between conditions in transition probabilities of fixations shows that it is more likely in the stable condition that the next fixation lands on the tracking task, independently of the previous task fixation.

## DISCUSSION

The presented study manipulated the cognitive control mode of participants in either a stable or flexible direction via a gamification method in the multitasking flight environment MATB. Results indicate differences in each sub-task performance, the way how participants distributed their visual attention and switched between sub-tasks tasks as well as the perceived mental workload.

Participants received on average a higher feedback score in the flexible condition compared to the stable condition. However, the improvement in feedback score from trial to trial was increased in the stable condition compared to the flexible condition. Moreover, participants demonstrated a better task performance for the system monitoring task, the communications task and the resource management task in the flexible condition (see Figure 1). Performance for the tracking task was better in the stable condition, which is in line with the stable control mode instruction. Although the number of fixations did not differ for the tracking task between conditions, participants looked on average longer per fixation on this task (see Figure 3). One explanation for this could be that the task has already attracted a considerable amount of attention due to the continuous control of the joystick. The increased mean fixation duration indicates that participants prioritized this task as expected. It is also noticeable that the mean fixation duration does only differ slightly between conditions for the system monitoring task. However, the number of fixations was distinctly increased in the flexible condition. The fact that the mean fixation duration did not differ between conditions may be related to the fact that the change in scales/buttons acted as an exogenous stimulus requiring the same time to be visually processed. The additional number of fixations in the flexible condition may signify a distraction by the moving scales.

The scenarios in the experiment were designed in a way that steady input of the participant was required. This created a concurrent multitasking situation, requiring constant task switching of the participant. Overall task performance was slightly increased in the flexible condition compared to

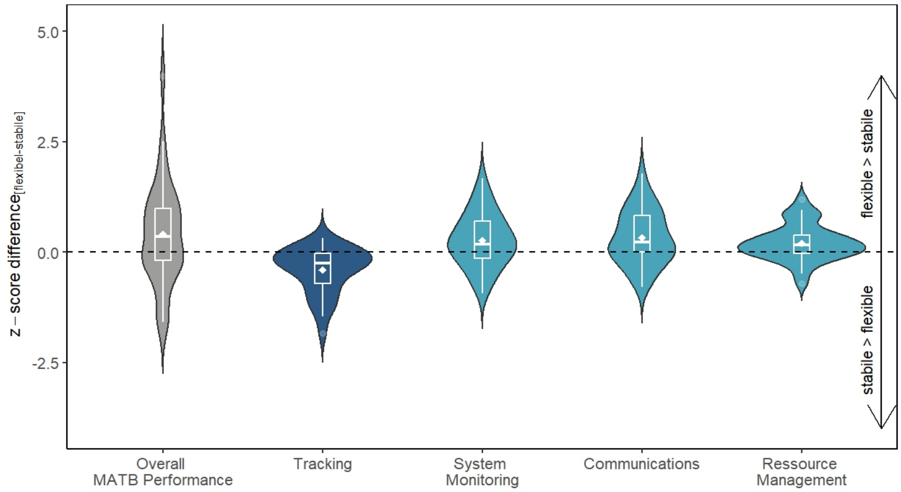


**Figure 1:** Results of the sign changed and z-standardized performance difference between the flexible and the stable condition for the four MATB subtasks.

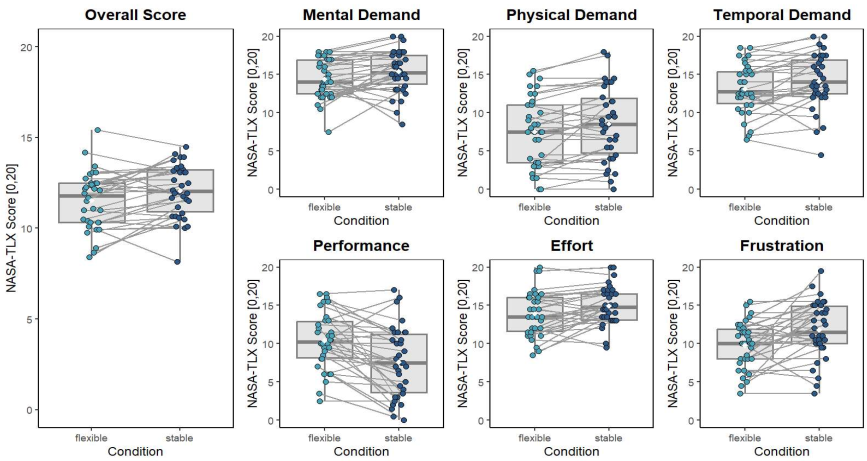


**Figure 2:** Results for the mental workload assessed by the NASA-TLX (Hart and Staveland, 1988).

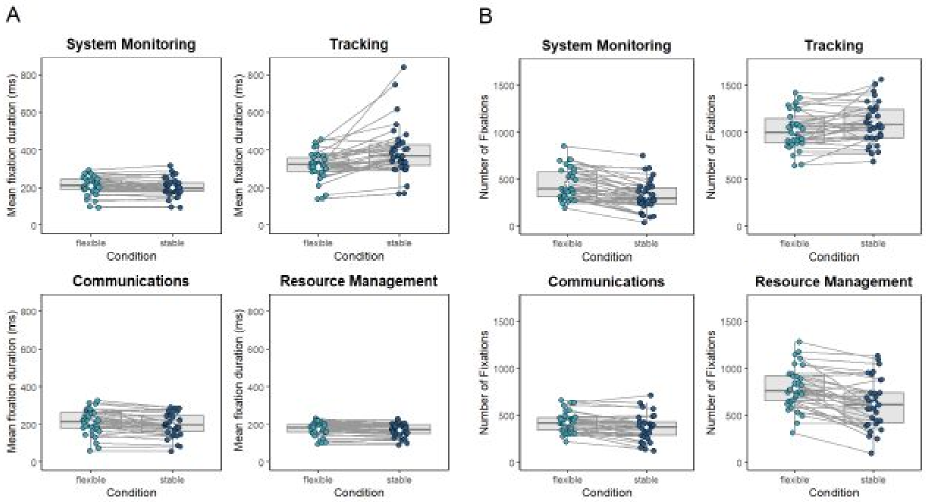


**Figure 3:** Comparison between the flexible and the stable condition regarding (A) the number of fixations and (B) the mean fixation for the four MATB subtasks.

the stable condition, which could be explained by individual differences in a preference for either a stable or flexible control mode (Brüning et al., 2021). Taking into account that participants perceived the mental workload as higher in the stable condition compared to the flexible condition (see Figure 2), one can conclude that the flexible control mode is preferred in concurrent multitasking situations. This conclusion is supported by the fact that participants received substantially better feedback scores in the flexible condition than in the stable condition. The observation that participants hardly improved their feedback score in the flexible condition indicates that the multitasking scenario itself triggered an optimal distribution of attention. However, in case of emergency situations requiring the pilot to follow task prioritization (A-N-C axiom) a stable control mode might be better suited to shield the safety-critical aviate task from interference with distracting tasks.

## Consideration for a Cognitive Assistance System

A prerequisite for an adaptive cognitive assistance system to provide situation-specific support is the real time diagnosis (Schwarz and Fuchs, 2017) of the control mode. The current study has demonstrated that besides the mean fixation duration and number of fixations per task, the number of task switches and the Coefficient K operate as correlates of each control mode. Participants switched tasks more in the flexible condition and employed ambient visual processing. The transition probabilities of fixations between all four tasks (see Figure 4C) indicates that participants are more likely to fixate the prioritized task than any other task in the stable condition. A cognitive assistance system could provide adaptive assistance according to the current employed control mode. For instance, if the pilot is in a stable control mode, but the situation requires that several flight tasks are processed with equal importance, the system could compensate for the disadvantages of the stable control mode. Exogenous cues, such as a stronger blinking of buttons on a display or increased volume of an auditory warning signal, would increase the salience of the non-prioritized task. The increase in salience could facilitate that the cue lowers the threshold for updating the working memory and reduces focal visual processing. Consequently, the likelihood of task switches and the detection of significant background information could be enhanced. Contrary, a pilot in a situation-inappropriate flexible control mode could be assisted in task prioritization by reducing the salience of background tasks. Less distraction would make it less likely that stimuli of competing tasks pass the working memory threshold. The enhanced goal-shielding could result in enhanced performance of the primary task, making errors of omission less likely.

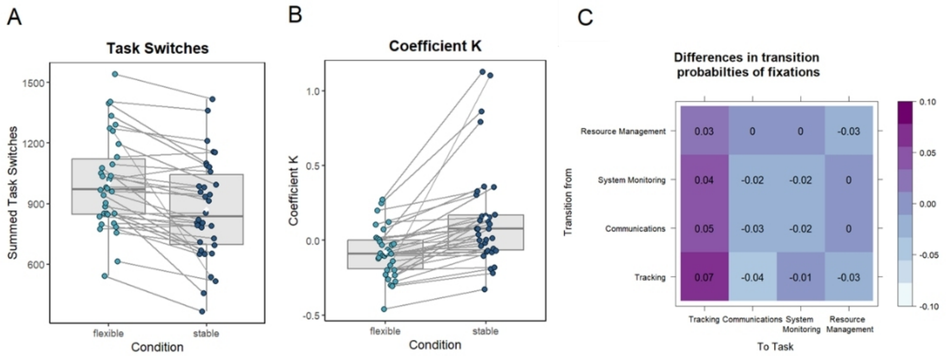


**Figure 4**: Comparison between the flexible and the stable condition regarding (A) the number of task switches, (B) the Coefficient K and (C) the difference in transition probabilities of fixations between the four MATB tasks. Difference was computed by subtracting the transition matrix of the stable condition from the flexible condition. Positive values indicate a higher transition probability in the stable condition, negative values indicate a higher transition probability in the flexible condition.

## CONCLUSION

The presented study shows how the stability-flexibility-dilemma of cognitive control influences multitasking behaviour in a flight environment. A flexible control mode resulted in lower mental workload and slightly enhanced overall task performance and is therefore better suited for concurrent multitasking situations. Different eye-tracking metrics, such as the number of fixations, the mean fixation duration, the number of task switches, the Coefficient K and transition probabilities of fixations serve as correlate of each control mode. A cognitive assistance system could take these correlates into account to provide situation-adaptive assistance by compensating for the associated disadvantages of each control mode.

## ACKNOWLEDGMENT

This research is funded by dtec.bw – Digitalization and Technology Research Center of the Bundeswehr project MissionLab which the authors gratefully acknowledge.

## REFERENCES

Andraszewicz, S., Scheibehenne, B., Rieskamp, J., Grasman, R., Verhagen, J. and Wagenmakers, E.-J. (2015) 'An introduction to Bayesian hypothesis testing for management research', Journal of Management, vol.41, no.2, pp. 521–543.

Antosko, M. and Lipovsky, P. (2022) 'A Technical Device for Determining the Predispositions of Students—Air Traffic Controllers and Pilots during Multitasking Training', Applied Sciences, vol. 12, no. 21, p. 11171.

Barron, L. G. and Rose, M. R. (2017) 'Multitasking as a predictor of pilot performance: Validity beyond serial single-task assessments', Military Psychology, vol. 29, no. 4, pp. 316–326.

Brüning, J., Reissland, J. and Manzey, D. (2021) 'Individual preferences for task coordination strategies in multitasking: exploring the link between preferred modes of processing and strategies of response organization', Psychological Research, vol. 85, no. 2, pp. 577–591.

Canas, J., Quesada, J., Antolí, A. and Fajardo, I. (2003) 'Cognitive flexibility and adaptability to environmental changes in dynamic complex problem-solving tasks', Ergonomics, vol. 46, no. 5, pp. 482–501.

Cegarra, J., Valéry, B., Avril, E., Calmettes, C. and Navarro, J. (2020) 'OpenMATB: A Multi-Attribute Task Battery promoting task customization, software extensibility and experiment replicability', Behavior research methods, vol. 52, no. 5, pp. 1980–1990.

Diamond, A. (2013) 'Executive functions', Annual review of psychology, vol. 64, pp. 135–168.

Eppinger, B., Goschke, T. and Musslick, S. (2021) 'Meta-control: From psychology to computational neuroscience', Cognitive, Affective, & Behavioral Neuroscience, vol. 21, no. 3, pp. 447–452.

Faul, F., Erdfelder, E., Lang, A.-G. and Buchner, A. (2007) 'G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences', Behavior research methods, vol. 39, no. 2, pp. 175–191.

Goschke, T. (2017) 'Volition und kognitive Kontrolle', in Allgemeine Psychologie, Springer, pp. 251–315.

Hart, S. G. and Staveland, L. E. (1988) 'Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research', in Hancock, P. A. and Meshkati, N. (eds) Advances in Psychology: Human Mental Workload, North-Holland, pp. 139–183.

JASP Team (2022) JASP (Version 0.16.4) [Computer software] [Online]. Available at https://jasp-stats.org/.

Koch, I., Poljac, E., Müller, H. and Kiesel, A. (2018) 'Cognitive structure, flexibility, and plasticity in human multitasking—An integrative review of dual-task and task-switching research', Psychological bulletin, vol. 144, no. 6, p. 557.

Koglbauer, I. (2015) 'Training for prediction and management of complex and dynamic flight situations', Procedia-Social and Behavioral Sciences, vol.209, pp. 268–276.

Krejtz, K., Duchowski, A., Krejtz, I., Szarkowska, A. and Kopacz, A. (2016) 'Discerning ambient/focal attention with coefficient K', ACM Transactions on Applied Perception (TAP), vol. 13, no. 3, pp. 1–20.

Loukopoulos, L. D., Dismukes, R. K. and Barshi, I. (2009) 'The perils of multitasking', AeroSafety World, vol. 4, no. 8, pp. 18–23.

McLean, G. M. T., Lambeth, S. and Mavin, T. (2016) 'The use of simulation in ab initio pilot training', The International Journal of Aviation Psychology, vol. 26, 1–2, pp. 36–45.

Morgan, B., D'Mello, S., Abbott, R., Radvansky, G., Haass, M. and Tamplin, A. (2013) 'Individual differences in multitasking ability and adaptability', Human factors, vol. 55, no. 4, pp. 776–788.

Musslick, S., Jang, S. J., Shvartsman, M., Shenhav, A. and Cohen, J. (2018) Constraints associated with cognitive control and the stability-flexibility dilemma.

Salvucci, D. D. and Taatgen, N. A. (2008) 'Threaded cognition: An integrated theory of concurrent multitasking', Psychological Review, vol. 115, pp. 101–130.

Schwarz, J. and Fuchs, S. (2017) 'Multidimensional Real-Time Assessment of User State and Performance to Trigger Dynamic System Adaptation', in Schmorrow, D. D. and Fidopiastis, C. M. (eds) Augmented Cognition. Neurocognition and Machine Learning, Cham, Springer International Publishing, pp. 383–398.

van Steenbergen, H., Band, G. P. H. and Hommel, B. (2009) 'Reward counteracts conflict adaptation: Evidence for a role of affect in executive control', Psychological Science, vol. 20, no. 12, pp. 1473–1477.