ARL-TR-9351 ● Nov 2021

The Effects of Task-Switching Within a Single Next Generation Combat Vehicle (NGCV) Role

by Thomas L Rohaly, Isadora M Fink, Katherine R Cox, Gregory M Gremillion, and Daniel E Forster

NOTICES

Disclaimers

ARL-TR-9351 ● Nov 2021

The Effects of Task-Switching Within a Single Next Generation Combat Vehicle (NGCV) Role

Thomas L Rohaly, Isadora M Fink, Katherine R Cox, Gregory M Gremillion, and Daniel E Forster
Human Research Engineering Directorate,
DEVCOM Army Research Laboratory

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED (From - To)
November 2021	Technical Report	16 February–13 October 2021

4. TITLE AND SUBTITLE
The Effects of Task-Switching Within a Single Next Generation Combat Vehicle (NGCV) Role

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Thomas L Rohaly, Isadora M Fink, Katherine R Cox, Gregory M Gremillion, and Daniel E Forster

5d. PROJECT NUMBER
ARL21-012

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
DEVCOM Army Research Laboratory
ATTN: FCDD-RLH-FC
Aberdeen Proving Ground, MD 21005

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-9351

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release: distribution unlimited.

13. SUPPLEMENTARY NOTES
ORCID IDs: Katherine Cox, 0000-0002-8351-751X; Gregory Gremillion, 0000-0002-0205-688X; Daniel Forster, 0000-0001-8351-2009

14. ABSTRACT
The Army's Next Generation Combat Vehicle (NGCV) concept will require that fewer Soldiers operate and manage a greater number of technologies and autonomous agents than in legacy armor platforms. This reduction in crew-to-asset ratio necessitates the ability for crewmembers to switch between various tasks in a rapid, efficient manner to meet the needs of the future operating environment. This work evaluates how task-switching affects performance in an NGCV-relevant paradigm, a high-fidelity simulation of gunning and target classification tasks. The timing of advanced preparation, which provided an alert of each pending task-switch, was examined as an intervention to mitigate the potential negative effects of task-switching, subject to variations in workload of both tasks. Results showed a task-switching cost in performance in the classification task, but the reverse was found for the gunning task, where performance was better after a task-switch. A manipulation check showed that the workload manipulation was successful, but contrary to the hypothesis, performance was better when workload was high for both the gunning (reaction time) and classification (accuracy) tasks. Finally, there were no effects of different levels of advanced preparation on performance following a task-switch in either task.

15. SUBJECT TERMS
Next Generation Combat Vehicle, NGCV, task-switching, task-switching cost, advanced preparation, workload

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					Thomas L Rohaly
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER (Include area code)
Unclassified	Unclassified	Unclassified	UU	32	(410) 278-7747

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

List of Tables

1. Introduction

Human–autonomy teaming (HAT) is a large focus in military research, seeking to determine how autonomy can supplement and improve team performance in a faster-paced and more complex future operating environment. The Human–Autonomy Teaming (HAT) Essential Research Program (ERP) in the US Army Combat Capabilities Development Command Army Research Laboratory is working to shift autonomous systems from tools to teammates, leveraging the strengths of humans and autonomy to allow them to perform better together than either could alone. The HAT ERP supports the Next Generation Combat Vehicle (NGCV) Army Modernization Priority and Cross-Functional Team to examine concepts and technologies to support the Manned–Unmanned Teaming (MUM-T) concept. The HAT ERP's Information for Mixed Squads laboratory conducts integrated simulation studies to examine human–autonomy teaming in the NGCV. This MUM-T Experimental Laboratory Simulation-in-the-Loop enables researchers to examine various concepts related to teaming and mission success, providing a testbed to examine new capabilities, such as task-switching, within military relevant scenarios. The HAT ERP is focused on developing concepts and technologies that will not only reduce the workload of the Soldier, but will allow autonomous systems to act as integrated teammates with Soldiers. For example, autonomy can be leveraged to support the Soldier's tasking, decision making, and overall performance.

The NGCV will come with a reduction in the crew-to-asset ratio, where armored vehicles will now be operated by two people, increasing the number of technologies and autonomous assets with which Soldiers will be interacting. This decreased ratio will likely result in Soldiers having to perform multiple tasks or switch between multiple tasks over the course of a mission. Some factors that could influence the efficiency with which Soldiers are able to complete tasks in this way are Soldiers' workload, task difficulty, and the amount and type of preparation they have before completing a given task.

2. Task-Switching

Task-switching is an executive function that involves shifting attention from one task to another (Demanet et al. 2011). When an individual switches between two tasks, whether voluntarily or involuntarily, there is often a switch-cost, or a decrease in performance as a result of the switch (Arrington and Logan 2004). This cost is thought to occur because when completing a given task, the rules for that task are primed to a certain degree such that the rules for a new task would be inhibited by this priming, affecting performance on the new task in a switch trial. The classic task-switching paradigm involves a task in which a participant is presented a letter/number pair, and before each stimulus presentation they are given a cue, either “Letter” or “Number.” When the cue “Letter” is presented, the individual’s task is to respond whether the letter in the target stimulus is a consonant or a vowel, and when the cue is “Number,” the task is to respond whether the target stimulus is even or odd (Cohen et al. 2008). Participants must use the cue information (prime) to prepare for the next trial, requiring them to switch between two sets of rules on a trial-by-trial basis. Switch costs are observed not only when people are required to complete a task-switch, but also when participants choose to complete a task-switch (Arrington and Logan 2004).

One main way in which the effects of task-switching are measured is through switching costs, or the decreases in performance seen following a switch trial compared with performing a single task (Koch and Allport 2006). Task-switching costs are assessed through performance decreases, measured through reaction time and accuracy (Samavatyan and Leth-Steensen 2009). The effects of task-switching on reaction time are measured by comparing reaction times for task repetitions, when the previous trial was the same task, with reaction time following task-switches, when the previous trial was a different task (Arrington and Logan 2004). A switch cost is said to occur if reaction time to complete a trial after a switch is longer than a trial in which the same task is repeated. Accuracy in task-switching is calculated by dividing the number of correct responses by the total number of responses for trials that do or do not follow a task-switch separately and then comparing accuracy between the two trial types (Koch and Allport 2006). While accuracy is often analyzed in task-switching paradigms, it is not focused on as much as reaction time, as it usually yields smaller effects.

Task-switching has largely been examined in individuals (Koch and Allport 2006; Cohen et al. 2008; Lin et al. 2017), but research on switching in teams has begun to expand recently. Task-switching in HAT has assessed how automation can be leveraged to assist in the completion of military-relevant missions (Squire et al. 2006). Specifically, Squire and colleagues have examined how automation might be used to assist human operators during high-workload conditions. Research has focused on determining what levels of automation lead to the smallest switch costs in two or more tasks (Squire et al. 2006; Squire and Parasuraman 2010). Results support the concept of flexible automation, where performance was improved when operators were able to decide when and how to use autonomy compared with persistent levels of low or high automation.

Various factors might affect the success of task-switching, such as mental workload, situational awareness (SA), stress, and anxiety (Derakshan et al. 2009;

Squire and Parasuraman 2010; Lin et al. 2017). In one study, participants controlled a team of robots to defend their own flag and capture their opponent's under varying workloads (Squire and Parasuraman 2010). Task load and level of autonomy were examined to determine how they affected performance when switching between offense and defense. Results showed that when participants switched between tactics, switch costs were greater when the robots had high levels of autonomy and participants took significantly fewer actions (i.e., overall number of commands given to the robots) in switch compared with repetition trials. Additionally, when participants controlled a greater number of robots, they won fewer games, had greater objective and subjective workload, and had longer mission times.

Stress and anxiety have also been shown to affect task-switching (Derakshan et al. 2009; Lin et al. 2017). One study compared how participants with low- or high-state anxiety performed switching between mathematical tasks of low (addition/subtraction) and high complexity (multiplication/division; Derakshan et al. 2009). Results showed that individuals with high levels of anxiety performed worse when switching between tasks than with repeating a task, whereas people with low anxiety performed slightly better when switching between tasks. Additionally, there was a three-way interaction between task complexity, task type, and state anxiety, where participants with high anxiety were slower on switch trials than repetition trials when complexity was high. Another study assessed the effects of acute stress on task-switching in a task in which participants had to switch between two sets of rules to describe a stimulus (Lin et al. 2017). Results showed that response times were faster following a task-switch when participants were stressed than when stress was not induced.

In addition to state effects (e.g., workload, anxiety, and stress) on task-switching, there are also task manipulations that can affect task-switching; specifically, we discuss advanced preparation (Meiran and Daichman 2005). Advanced preparation is a cue presented to participants in advance of a subsequent trial, which either provides information on rules of the next trial or whether it will be a switch or repetition trial (Meiran and Daichman 2005; Gruber et al. 2006; Karayanidis et al. 2010). Several studies have shown that switch costs can be greatly reduced by giving participants time to prepare for a task-switch (Meiran and Daichman 2005; Gruber et al. 2006; Karayanidis et al. 2010). While the exact mechanisms behind this improved performance are a matter of debate, the benefits of advanced preparation have been demonstrated across several different task-switching paradigms (Gruber et al. 2006).

3. Method

3.1 Participants

Twenty-one people were recruited from the DEVCOM Army Research Laboratory and Ground Vehicle System Center. During the recruitment process, participants completed the Motion Sickness Simulator Questionnaire (MSSQ) to screen for high susceptibility to motion sickness; people who were highly susceptible were not asked to participate in the study. Participants also provided details of their computer specifications in advance, as this study was conducted online at home and certain specifications were required for seamless streaming of the experiment. One participant was excluded from the study due to motion sickness susceptibility, and no one for computer specifications, resulting in 20 participants who took part in the study.

3.2 Stimuli

The main task in this study was developed and run in Unreal Engine 4 (https://www.unrealengine.com/en-US/). The simulated environment for this study was a 5- × 5-km region that contained both rural and urban areas, similar to what might be found in an Eastern European country. Two types of stimuli were used in the study. First, Unreal Tank and a van asset were placed throughout the environment and used as targets and foils (more detail follows on this gunning task). The second type of stimuli used in this study were images of different Unreal ground assets in the environment (e.g., vans, technical trucks, and tanks; more detail follows). A full list of assets can be found in Appendix A. Figures 1 and 2 show the interfaces participants used for the classification and gunning tasks, respectively.

Fig. 1 Interface for the classification task

Fig. 2 Interface for the gunning task

Over the course of the study, participants completed several questionnaires.

Motion Sickness Questionnaire

The Short form (MSSQ-Short; Golding 2006) is designed to determine how susceptible a person is to motion sickness. Here, sickness means feeling queasy, nauseated, or becoming physically ill. Because a 180° simulator can cause mild motion or simulation sickness in people who score in the 76th percentile of the MSSQ-Short, participants who had a score of 19 (75th percentile) or higher were thanked for their interest but not scheduled for participation. The MSSQ-Short was used in the present study, which has age and sex demographic information questions removed from the MSSQ. This questionnaire was only used to determine the eligibility of people interested in participating, and results were not used for any additional analyses.

Computer Specifications

The Computer Specifications questionnaire asked five questions about the specifications for the computer on which they intended to run this experiment from their home. This included questions about what central processing unit their computer used, how much RAM the computer had, and how much hard-drive space was left on the computer. Like the MSSQ-Short, this was only used to determine eligibility for participation. Everyone who was recruited met the requirements for the computer specifications.

Subjective Workload Questionnaire

The Subjective Workload questionnaire was designed to assess participants' subjective workload experienced during the experiment. Specifically, this questionnaire assessed workload experienced during the two tasks (gunning and classification) that participants completed, rating their workload on a seven-point Likert scale. This asked about workload in each specific task, for switches into each task type, and for task-switching overall. This questionnaire was given to participants immediately following the main task.

NASA-Task Load Index (TLX) Short

NASA-TLX Short is a questionnaire in which people self-report their perceived workload demands on six subscales: mental demand, physical demand, temporal demand, performance, effort, and frustration (Hart and Staveland 1988). The short version of the NASA-TLX only consists of the first section, where six items are used to gather an unweighted score for each subscale. For each subscale factor, participants responded on a scale ranging from 0 (Low) to 100 (High) in five-point increments, and all of the subscales were averaged to make up an overall workload score. This questionnaire was given to participants right after the Subjective Workload Questionnaire and again after the workload manipulation check.

In addition to the questionnaires, a short inventory of gaming experience was administered to evaluate participants' video game preferences and play characteristics.

3.3 Procedure

During the recruitment process, each potential participant was emailed the MSSQ-Short and the Computer Specification Questionnaire for prescreening. If a participant was susceptible to motion sickness or their computer did not meet the minimum required specifications to run the experiment, they were thanked for their interest but not scheduled to participate. Once the prescreening was complete, participants who were eligible to participate were sent a link to the main task, which was hosted on Amazon's Appstream Server. Participants completed the study in their homes using their own computer and mouse at a time that was convenient to them.

Once participants clicked on the link to begin the experiment, they read instructions on how to complete each task and completed 10 practice trials of each task type. Once training was complete, participants completed two blocks of trials, 100 trials each, with a forced break between blocks. Of the 200 total trials, 80 were switch trials and 120 were repetition trials. That is, after 80 trials, participants switched from the task they were completing to a new task (e.g., from task A to task B), and after 120 trials, participants continued the same task they were completing (e.g., from task A to task A). Due to the nature of one of the tasks (gunning) in which the sequence of trials was linear throughout the study, regardless of the number of trials of the other task (classification) that occurred between trials of the gunning task, it was not feasible to randomize trials in this study. There were two tasks in this study, gunning and classification, and each had high- and low-workload conditions.

In this experiment, participants played the role of a gunner in an NGCV Robotic Combat Vehicle. The vehicle was "driven" by someone else and moved through the environment without any guidance from the participant. Participants completed two tasks within this gunner role. The primary task was gunning, in which participants were to shoot all enemy tanks (targets) and not shoot civilian vans (foils) that they saw. Approximately 30% of all assets presented were foils, and 70% were targets. The computer mouse was used to aim, and firing was done using the left mouse key. In the low-workload condition, participants saw 6 assets over the course of a trial, and in the high-workload condition, they saw 12 assets over the course of a trial. Each gunning trial lasted for 25 s.

The second task was classification, in which participants were presented with a list containing a number of ground-based objects that they had to classify as either a threat (e.g., a tank) or a non-threat (e.g., a sedan). Participants classified each image by using a left mouse click to categorize it as either a "Threat" or "Non-threat" using one of two buttons next to each image. Up to four images were displayed in the list at any given time. When there were more than four images to be classified, a queue was shown to the right of the threat/non-threat buttons, and these images filled the four image slots as soon as a given image was classified and a slot open. Workload in classification trials was manipulated through the number of images presented for classification throughout a trial, including in the queue, as well as the rate at which new images were added. In low-workload trials, participants were initially presented with four images, and four new images were added every 4 s for a total of 24 images per trial. In the high-workload condition, participants were initially presented with six images, and an additional six images were added every 3 s for a total of 42 images per trial. Each classification trial lasted for 20 s.

To examine how varying levels of advanced preparation might affect task-switching in this study, participants were provided with advanced notice in trials before a switch took place. Advanced preparation was presented to participants as an alert bar across the top of the screen, telling them that they would soon be switching to the alternative task (i.e., "You will be switching to gunning soon"). To assess the optimal length of advanced preparation, five levels were used: 200 ms and 2, 4, 7, and 10 s. Advanced preparation was presented during each trial that preceded a task-switch; there was no inter-trial interval. Each value of advanced preparation was presented an equal number of times over the course of the experiment for each task. Participants were not told how much time they had before the switch occurred.

After participants completed the main task, they were asked to complete a set of questionnaires. Participants were first presented with the Subjective Workload Questionnaire and the NASA-TLX Short, asking about the two tasks they had just completed. A workload manipulation check was given next in which participants completed 20 trials across four blocks, with each block of 5 trials corresponding to the four different trial types in the main tasks: Gunning High Workload, Gunning Low Workload, Classification High Workload, and Classification Low Workload. At the end of each block, participants were given a single question asking them to rate their subjective workload for those five trials on a seven-point Likert scale. The experimental session lasted approximately 1 h and 45 min.

Several predictions were made about the results of the study. First, we hypothesized that reaction times for switch trials (trials preceded by a different task type) would be slower than reaction times for repetition trials (trials preceded by the same task type), as has been seen in previous studies (Arrington and Logan 2004; Meiran and Daichman 2005). Second, we hypothesized that accuracy for switch trials would be lower than accuracy for repetition trials, a switch cost seen previously (Logan 2003; Monsell 2003). Third, we hypothesized that performance in low-workload trials would be better (i.e., faster reaction times and higher accuracy) than performance in high-workload trials. Previous research has shown that performance decreases when in high-workload conditions compared with low-workload conditions (Squire and Parasuraman 2010). We hypothesized that as the length of advanced preparation increased, response times to the first stimulus in the new trial would increase in a logarithmic pattern. It was expected that there is an optimal amount of time that an individual needs to prepare for a task-switch, where too much or too little time compared with this optimal time would result in worse performance.

4. Results

4.1 Data Preparation

To prepare data for analysis, reaction times had to be filtered. For the gunning task, reaction time was calculated from one of three start times: 1) the trial start, 2) the time that a participant first saw a stimulus (line of sight [LOS]), or 3) the timestamp of the response to the previous stimulus. In most cases, reaction time was calculated by taking the timestamp for the response minus either the LOS timestamp for that asset or the timestamp of the previous response, whichever was more recent. Because repetition trials had a seamless transition in the environment, a subset of trials began in which a stimulus was in LOS before the previous trial ended and the trial in which a response was required. Thus, in these cases, when a participant responded to a stimulus before the start of the correct trial for that stimulus, responses were removed from further analyses. When participants responded to a stimulus in the correct trial but gained LOS in the previous trial, reaction time was calculated based on initial LOS rather than trial start time. In total, 8090 responses were removed after applying these exclusions, resulting in a total of 22,570 responses analyzed within these criteria. For the classification task, reaction time was calculated from stimulus onset, or when a target image first appeared in the image list.

Next, responses were filtered out for each task. In both tasks, responses faster than 100 ms were not recorded, so any response below this cut-off was removed. Average response times and standard deviations were then analyzed for gunning and classification separately, and all responses three standard deviations above or below the average were removed for each task. Thus, in the gunning task, any reaction time longer than 11 s was removed, and for the classification task, any response longer than 6 s was removed.

Response accuracy was coded as either a correct response (i.e., hits and correct rejections in gunning) or an incorrect response (i.e., misses and false alarms in gunning).

4.2 Workload Analyses

To evaluate the effectiveness of the workload manipulation, we analyzed responses to the Subjective Workload Questionnaire and the workload manipulation check. Data from the NASA-TLX short is not discussed in this report. We ran paired-samples t-tests on responses to the Subjective Workload Questionnaire to compare ratings of overall workload in the gunning task with the classification task and to compare switches to gunning with switches to classification. Results showed that there was no significant difference between ratings of perceived workload between the gunning task (M = 3.95, SE = 0.387) and the classification task [M = 3.35, standard error (SE) = 0.302; $t(19) = 1.878$, $p = 0.076$], and none between switches to the gunning task (M = 3.05, SE = 0.32) and switches to the classification task [M = 2.85, SE =0.244; $t(19) = –0.89$, $p = 0.385$; Fig. 3].

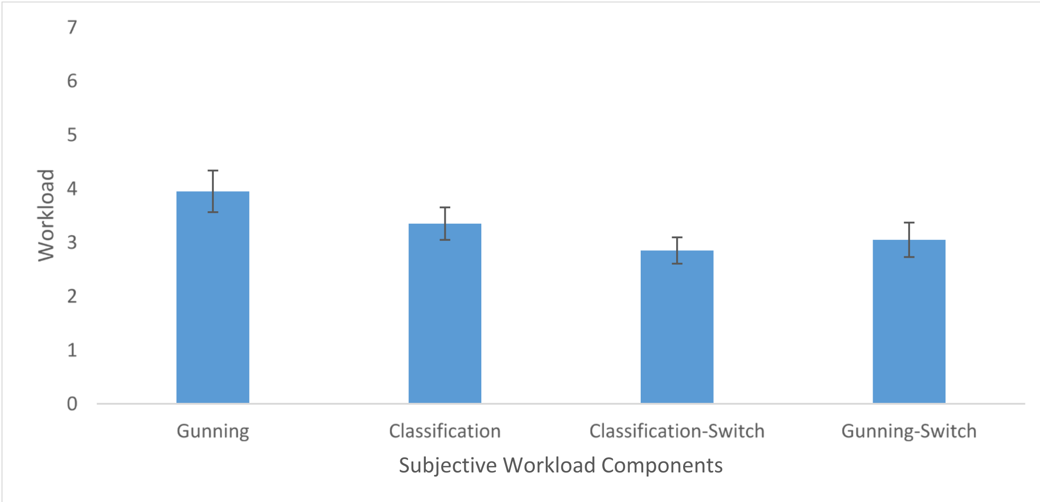

Fig. 3 Mean ratings of perceived workload on the Subjective Workload Questionnaire. Error bars represent ±1 SE of the mean.

Two paired-samples t-tests were conducted to examine workload ratings in response to the manipulation check blocks completed after the main task. Figure 4 shows the mean ratings of perceived workload for the low- and high-workload conditions for each task. For the gunning task, results showed that ratings of perceived workload for the low-workload condition (M = 3.1, SE = 0.24) were significantly lower than the high-workload condition [M = 4.35, SE = 0.31; $t(19) = -5.225$, $p < 0.001$]. The same was seen for the classification task, where perceived workload in the low-workload condition (M = 3.4, SE = 0.311) was rated significantly lower than in the high-workload condition [M = 4.32, SE = 0.316; $t(19) = -3.923$, $p < 0.001$].

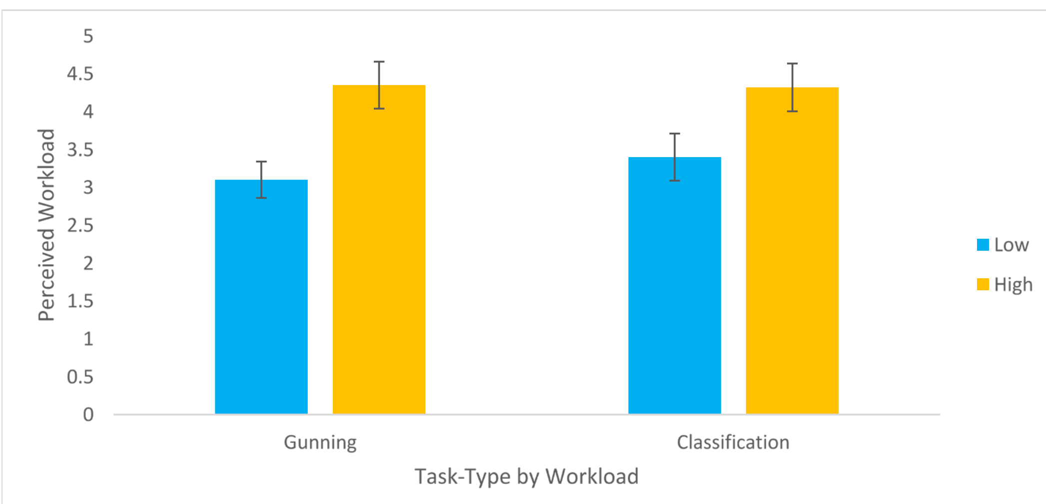

Fig. 4 Mean ratings of subjective workload for classification and gunning tasks at each level of workload. Error bars represent ±1 SE of the mean.

To evaluate the effects of workload on performance, we conducted a series of mixed effects models using the "lme4" package (Bates et al. 2015), with p-values computed using the "lmerTest" package (Kuznetsova et al. 2017), in R version 4.1.0 (R Core Team 2020). To test whether performance varied across levels of workload (low vs. high), we ran the following model:

$$Y_{ij} = \gamma_{00} + \gamma_{01}(Workload) + u_{oj}$$

where Y_{ij} represents response to stimulus i for person j (in terms of either Reaction Time or Accuracy), γ_{00} represents the expected value of the outcome for low-workload trials, γ_{01} represents the expected effect of changing from a low-workload trial to a high-workload trial, and u_{oj} allows individual intercepts to vary. For reaction times, we included responses within each trial; however, we averaged across accuracy responses within each trial to ease interpretation (that is, coefficients represent the changes in proportion of correct responses) and circumvent issues with estimating binary response outcomes with a logistic model. In the gunning task, results showed that reaction time was significantly faster in the low-workload condition (M = 3.55, 95% confidence interval [CI; 3.20, 3.91]) than the high-workload condition [M = 3.72, 95% CI (3.24, 4.19); $t(8951) = 2.737$, $p = 0.006$; Fig. 5]. In contrast, accuracy was significantly lower in the low-workload (M = 0.83, 95% CI [0.82, 0.85]) than the high-workload condition [M = 0.86, 95% CI (0.83, 0.88); $t(1968) = 3.097$, $p = 0.002$; Fig. 6]. In the classification task, results showed that reaction time was significantly faster in the low-workload condition (M = 2.037, 95% CI [1.83, 2.24]) than in the high-workload condition [M = 2.367, 95% CI (2.15, 2.59); $t(42440) = 34.8$, $p < 0.001$; Fig. 5]. There was no difference in accuracy between the low-workload condition (M = 0.98, 95% CI [0.95, 1.0]) and the high-workload condition [M = 0.98, 95% CI (0.946, 1.0); $t(1961) = –0.538$, $p = 0.591$; Fig. 6].

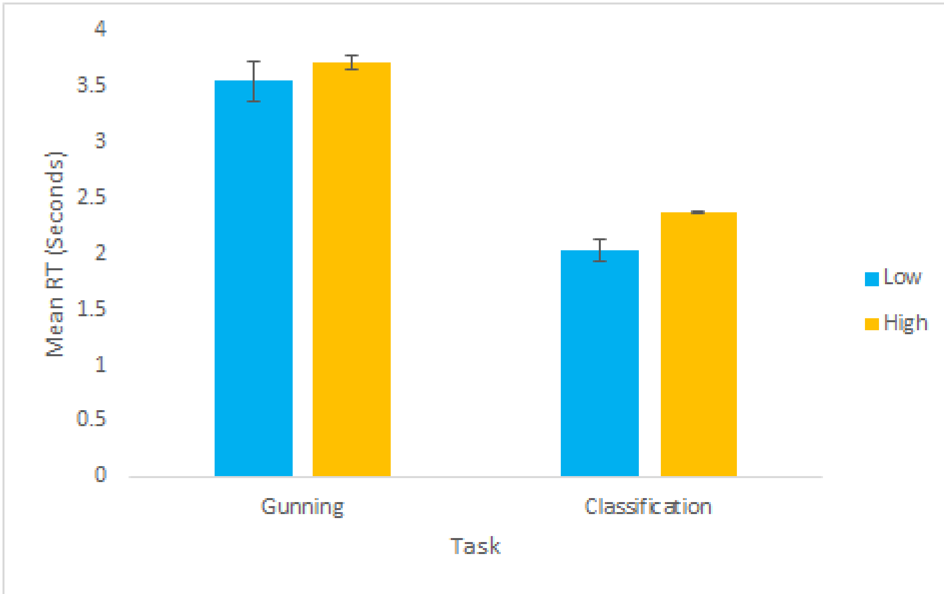

Fig. 5 Mean reaction time (seconds) at each level of workload. Error bars represent ±1 SE of the mean.

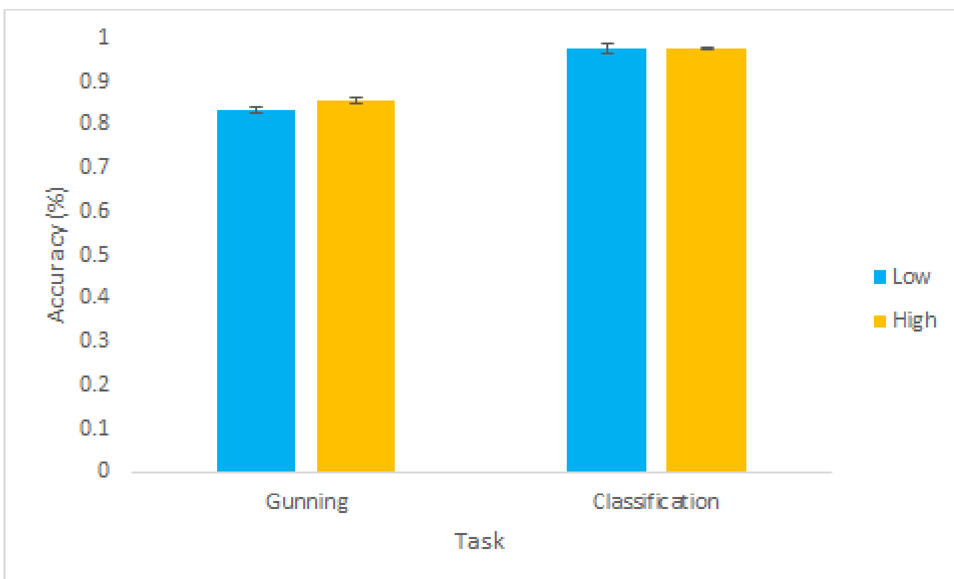

Fig. 6 Mean accuracy (%) at each level of workload. Error bars represent ±1 SE of the mean.

4.3 Task-Switching Analyses

To test whether reaction time and accuracy differed between switch and repetition trials, we averaged across participants' initial responses within a trial, separately for repetition and switch trials, and conducted paired-samples *t*-tests. In the gunning task, results showed that reaction time was significantly faster in trials following a switch (M = 1.62) than those following a repetition [M = 1.99, Mean Diff = 0.368; $t(19) = 3.189$, $p = 0.005$, 95% CI (0.126, 0.609); Fig. 7]. There was no difference in accuracy between trials following a switch (M = 0.984) and a repetition [M = 0.977; Mean Diff = –0.006, $t(19) = –0.729$, $p = 0.474$, 95% CI (–0.025, 0.012); Fig. 8]. In the classification task, reaction time was significantly slower in trials following a switch (M = 1.31) than in those following a repetition [M = 1.12, Mean Diff = –0.19; $t(19) = –3.817$, $p = 0.001$, 95% CI (–0.29, –0.09); Fig. 7]. Conversely, accuracy was significantly higher in trials following a switch (M = 0.985) than in those following a repetition [M = 0.929; Mean Diff = –0.056, $t(19) = –4.7652$, $p < 0.001$, 95% CI (–0.08, –0.03); Fig. 8].

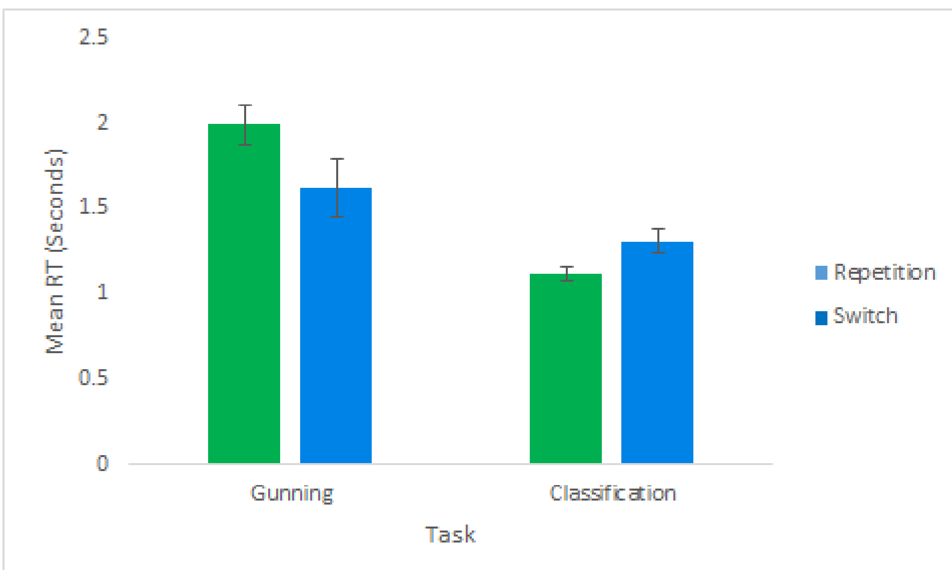

Fig. 7 Mean reaction time (seconds) at each level of task type. Error bars represent ±1 SE of the mean.

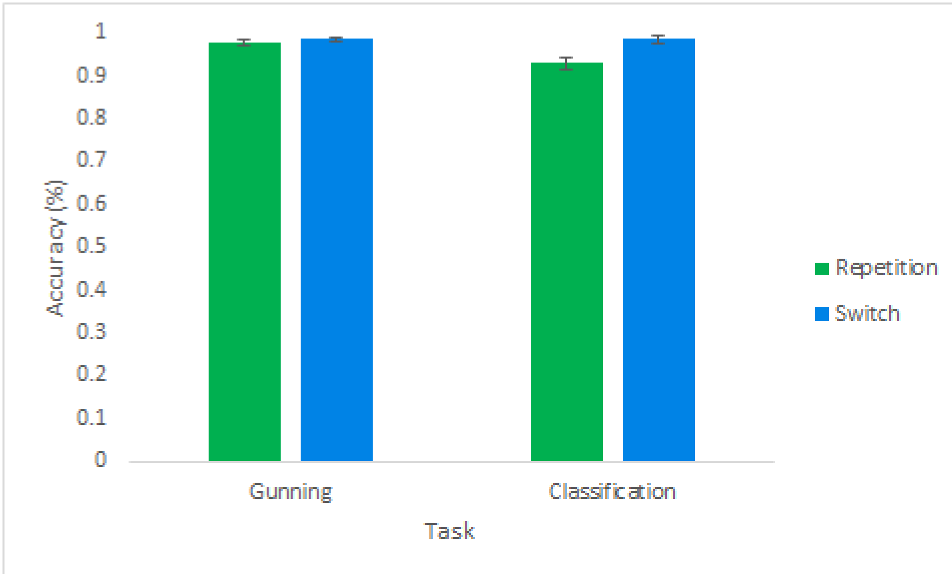

Fig. 8 Mean accuracy (%) at each level of task type. Error bars represent ±1 SE of the mean.

4.4 Advanced Preparation Analyses

To test whether reaction times differed across levels of advanced preparation ahead of a task-switch, we ran the following model:

$$Y_{ij} = \gamma_{00} + \gamma_{01}(AP_{2s}) + \gamma_{02}(AP_{4s}) + \gamma_{03}(AP_{7s}) + \gamma_{04}(AP_{10s}) + u_{oj}$$

where

- Y_{ij} represents the response to the first stimulus in trial i (whether reaction time or accuracy) for person
- γ_{00} represents the expected value of the outcome when participants were notified 0.2*s* prior to switching tasks
- γ_{01} represents the expected difference between 2*s* and 0.2*s* of preparation time
- γ_{02} represents the expected difference between 4*s* and 0.2*s* of preparation time
- γ_{03} represents the expected difference between 7*s* and 0.2*s* of preparation time
- γ_{04} represents the expected difference between 10*s* and 0.2*s* of preparation time
- u_{oj} allows individual intercepts to vary

As with workload analyses, accuracy on initial trial responses were averaged across trials separately for each advanced preparation time and each task, thus making this analysis akin to a single-level repeated measures analysis of variance. Tables 1 and 2 show the results of the mixed-effects model for the gunning and the classification tasks, respectively, at each level of advanced preparation. Results show that there were no differences in the effects of the different levels of advanced preparation on performance following a task-switch in either the gunning (p's ≥ 0.293) or the classification task (p's ≥ 0.245).

Table 1 Mixed-effects model summaries for the gunning task at each level of advanced preparation

Advanced preparation	Mean reaction time	SEM	df	t	sig	95% CI
0.2	1.57	0.21				[1.17, 1.98]
2	1.84	0.21	365.22	1.28	0.202	[1.03, 2.65]
4	1.44	0.23	364.19	–0.56	0.576	[0.59, 2.30]
7	1.37	0.21	363.15	–0.96	0.339	[0.55, 2.19]
10	1.85	0.28	362.40	1.27	0.205	[1.02, 2.70]
Advanced preparation	**Mean accuracy (%)**	**SEM**	**df**	**t**	**sig**	**95% CI**
0.2	0.98	0.012	75.99	…	…	[0.958, 1.0]
2	0.99	0.017	75.99	0.578	0.565	[0.949, 1.02]
4	0.98	0.002	75.99	0.096	0.924	[0.95, 1.02]
7	0.99	0.018	75.99	1.06	0.293	[0.949, 1.02]
10	0.98	0.017	75.99	–0.241	0.81	[0.94, 0.999]

Notes: SEM = standard error of the mean; df = degrees of freedom; t = t-score; sig = significance value

Table 2 Mixed-effects model summaries for the classification task at each level of advanced preparation

Advanced preparation	Mean reaction time	SEM	df	t	sig	95% CI
0.2	1.30	0.07	…	…	…	[1.16, 1.45]
2	1.30	0.04	756.10	0.01	0.99	[1.22, 1.378]
4	1.33	0.04	756.10	0.67	0.50	[1.28, 1.43]
7	1.30	0.04	756.10	–0.15	0.88	[1.21, 1.37]
10	1.27	0.04	756.10	–0.91	0.36	[1.15, 1.31]
Advanced preparation	**Mean accuracy (%)**	**SEM**	**df**	**t**	**sig**	**95% CI**
0.2	0.98	0.01	…	…	…	[0.957, 1.0]
2	0.99	0.01	76	0.651	0.517	[0.96, 1.0]
4	0.98	0.001	76	0.046	0.963	[0.96, 1.0]
7	0.99	0.01	76	1.171	0.245	[0.961, 1.0]
10	0.98	0.01	76	0.0	1.0	[0.959, 1.0]

5. Discussion and Conclusion

In this study, we investigated the effects of workload and advanced preparation on task-switching in a high-fidelity, military-relevant simulation. There were three main predictions in this study: 1) performance would be worse following a task-switch than a task repetition, 2) performance would be better when workload was low than when it was high, and 3) as the length of advanced preparation increased, performance after a task-switch would improve in a logarithmic pattern.

As hypothesized, a task-switching cost was seen in reaction time in the classification task, with a slower reaction time in trials following a switch than following a repetition. Accuracy in the gunning task did not show this same effect: There was no difference between switch and repetition trials due to accuracy being near ceiling. This ceiling effect may be partially due to stimulus presentation as an asset was always in view immediately following a switch to the gunning task, forcing participants to immediately orient to and respond in the new task. However, because this first asset could have been either a target or a foil, high accuracy suggests that participants did not simply fire at the first asset they saw in the gunning task, so we do not think that this is the cause of accuracy being at ceiling.

Contrary to our hypothesis, switch gains were seen in both gunning reaction time and classification accuracy. Switch gains have been shown previously, where performance was better following a switch than a repetition in participants with low anxiety and when switch trials were cued (Derakshan et al. 2009). Trends toward a switch gain have been seen in some studies (Meiran 2000; Brass et al. 2003; Meiran and Daichman 2005), and other studies have eliminated but not reversed a switch cost (Gruber et al. 2006). The amount of gaming experience a person has could also lead to switch gain, as previous literature has shown that gamers often show switch gains rather than switch costs (Colzato 2010; Strobach et al. 2012; Pallavicini et al. 2018; Steyvers et al. 2019). The present study, however, did not show enough variability in participant gaming experience to assess whether it may have affected task-switching performance and contributed to the switch gain. Asset placement in the gunning task may also have affected the switch gain that was seen with reaction time. That is, following the task-switch, there was always a target or foil in view, whereas in repetition trials, timing of and distance to the first asset when LOS was achieved was variable. While it is not clear how or if this confound of asset placement may have affected response time between switch and repetition trials, future studies should examine this possibility. When considering the switch gain seen in accuracy in the classification task, one possible explanation for worse accuracy in repetition trials may be the transition between repetition trials. When one classification trial ended and a new one began, any images that had not been classified were cleared from the screen and replaced with images for the new trial. To avoid inaccurate responses due to a trial change, any response in the first 100 ms of a trial were considered noise and not analyzed. However, any accidental responses to a previous trial that occurred more than 100 ms into the new trial were recorded as a response and therefore could have lowered accuracy for repetition trials. Results support this idea, as the average initial response accuracy for responses 200 ms or faster was 58.9%, whereas average accuracy for initial reaction times slower than 200 ms was at 97%.

As hypothesized, reaction time was faster in the low-workload condition than in the high-workload condition for both the gunning and classification tasks. This lower performance seen with higher workload has been seen in previous literature (Squire et al. 2006; Squire and Parasuraman 2010). The workload manipulation check showed that participants reported the high-workload condition to be more taxing than the low-workload condition, consistent with the slower response time when workload was higher. There was no difference between the workload conditions seen through accuracy in the classification task, but in the gunning task, accuracy was higher in the high-workload condition than in the low-workload condition. One possible explanation for this reverse effect is that participants may have prioritized accuracy over speed in this task. This effect is known as a speed–accuracy trade-off, where the two metrics are negatively correlated with one another (Samavatyan and Leth-Steensen 2009). To assess whether this effect was present in the high-workload condition of the gunning task, a mixed effects model was run where accuracy was evaluated using reaction time as a predictor and allowed individual intercepts to vary. Results showed a speed–accuracy trade-off in this condition ($b = 0.092$, $SE = 0.036$, Log Odds = 0.09, $z = 2.579$, $p = 0.0099$); that is, for every

1-s increase in reaction time, there was a 9% increase in the likelihood of the response being accurate. This result supports the idea that a speed–accuracy trade-off may have led to the switch gain, seen through higher accuracy in the high-workload conditions than the low-workload condition in the gunning task.

The hypothesis about advanced preparation was not supported, as no significant performance differences were found between the different levels of advanced preparation when switching into the gunning or classification tasks. One difference between this study and traditional task-switching studies is how advanced preparation was presented. Specifically, more-traditional task-switching studies provide advanced preparation in the form of cue-target-intervals (CTIs; Gruber et al. 2006; Karayanidis et al. 2010), which are intervals between different trial types during which the preparation cue is presented sometimes along with an additional interval of time (Meiran 1996; Gruber et al. 2006). In traditional task-switching paradigms, these intervals are normally up to 2 s long (Meiran 1996; Gruber et al. 2006; Kalanthroff and Henik 2014). Conversely, in the present study, the advanced preparation cue was presented to participants during each trial that preceded a switch, and a new trial began immediately following the switch with no interval between trials. Additionally, the current study examined advanced preparation times that ranged from 200 ms to 10 s, with most of the preparation times being longer than those in traditional task-switching paradigms. There are multiple possible reasons why no significant difference was found between these levels of advanced preparation. One possible reason is that participants had to prepare for the next task while simultaneously completing the current task. This could have made preparing for the next task, which involves priming the correct/next task and inhibiting the incorrect/current task, more difficult. However, it is not realistic for a Soldier to stop one task and do nothing for a period of time so that priming can occur in isolation, so even if this was a factor in the present study, adding a CTI is not realistic. A second possible factor is the advanced preparation levels assessed in this study. It is possible that the ideal amount of time to prepare for a task-switch between complex tasks in high-fidelity simulations is outside of the range tested here. However, because we tested a range of nearly 10 s between our shortest and longest advanced preparation times, it is unlikely that timing of this preparation was an issue.

More research should be conducted to further examine the effects of task-switching in high-fidelity, military-relevant simulations. Specifically, different tasks and situations that are common in military settings should be evaluated to determine what tasks may be affected by a task-switch and if there are tasks for which switching would result in a switch cost. The current study only examined task-switching between two tasks that are part of the role of a gunner of a Robotic

Combat Vehicle. These tasks may not have been different enough to show a task-switch cost because they were part of the same role and did not capture many of the persistent cognitive tasks that would be carried out by vehicle operators as they switch between tasks (e.g., communication with crew and maintenance of SA). Soldiers often perform more than two tasks within a role, and sometimes more than one role, so more-complex task-switches should be examined in the future. In the NGCV, there will be fewer crewmembers to operate and manage a greater number of technologies and intelligent agents, which will lead to increases in Soldier workload. This study also did not examine significant changes in perspective that may be expected for operators switching between tasks in physically distributed vehicles and which could induce greater workload, switch costs, and demonstrate larger effects of advanced preparation methods. Future research should also further examine advanced preparation to better understand how it is used to prepare for a task-switch.

It is possible that other methods or manipulations of advanced preparation in similar complex, military-relevant scenarios could lead to different results. For example, the current study only alerted participants that a task-switch was pending but did not provide the length of time before a switch; therefore, providing participants with the time before a switch occurs may allow for better preparation. Additional factors related to Soldier state or other internal factors may also affect performance in task-switching and should therefore be examined in these more-complex paradigms. Stress (Lin et al. 2017), heart-rate variability (Colzato et al. 2018), and cognitive flexibility (Koch et al. 2018) have all been shown to be related to task-switching performance, so future research should better understand how these factors may relate to Soldier task-switching. As more technologies and intelligent agents are used in the military, understanding how Soldiers switch between tasks will become increasingly important.

6. References

Arrington CM, Logan GD. The cost of a voluntary task-switch. Psychological Science. 2004;15(9):610–615. https://doi.org/10.1111/j.0956-7976.2004.00728.x.

Bates D, Maechler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. Journal of Statistical Software. 2015;67(1):1–48. doi:10.18637/jss.v067.i01.

Brass M, Ruge H, Meiran N, Rubin O, Koch I, Zysset S, Prinz W, von Cramon DY. When the same response has different meanings. NeuroImage. 2003;20(2):1026–1031. https://doi.org/10.1016/S1053-8119(03)00357-4.

Cohen AL, Bayer UC, Jaudas A, Gollwitzer PM. Self-regulatory strategy and executive control: implementation intentions modulate task-switching and Simon task performance. Psychological Research. 2008;72(1):12–26. https://doi.org/10.1007/s00426-006-0074-2.

Colzato LS, Jongkees BJ, de Wit M, van der Molen MJW, Steenbergen L. Variable heart rate and a flexible mind: higher resting-state heart rate variability predicts better task-switching. Cognitive, Affective, & Behavioral Neuroscience. 2018;18(4):730–738. https://doi.org/10.3758/s13415-018-0600-x.

Colzato LS. DOOM'd to switch: superior cognitive flexibility in players of first person shooter games. Frontiers in Psychology. 2010;1. https://doi.org/10.3389/fpsyg.2010.00008.

Demanet J, Liefooghe B, Verbruggen F. Valence, arousal, and cognitive control: a voluntary task-switching study. Frontiers in Psychology. 2011;2. https://doi.org/10.3389/fpsyg.2011.00336.

Derakshan N, Smyth S, Eysenck MW. Effects of state anxiety of performance using a task-switching paradigm: an investigation of attentional control theory. Psychonomic Bulletin & Review. 2009;16:1112–1117. https://doi.org/10.3758/PBR.16.6.1112.

Golding, J. F. (2006). Motion sickness susceptibility. Autonomic Neuroscience. 2006;129:67–76. doi:10.1016/j.autneu.2006.07.019.

Gruber O, Karch S, Schlueter EK, Falkai P, Goschke T. Neural mechanisms of advance preparation in task-switching. NeuroImage. 2006;31(2):887–895. https://doi.org/10.1016/j.neuroimage.2005.12.043.

Hart SG, Staveland LE. Development of NASA-TLX (task load index): results of empirical and theoretical research. Advances in Psychology. 1988;52:139–183.

Kalanthroff E, Henik A. Preparation time modulates proactive control and enhances task conflict in task-switching. Psychological Research. 2014;78(2):276–288. https://doi.org/10.1007/s00426-013-0495-7.

Karayanidis F, Jamadar S, Ruge H, Phillips N, Heathcote A, Forstmann UB. Advance preparation in task-switching: converging evidence from behavioral, brain activation, and model-based approaches. Frontiers in Psychology; 2010. https://doi.org/10.3389/fpsyg.2010.00025.

Koch I, Allport A. Cue-based preparation and stimulus-based priming of tasks in task-switching. Memory & Cognition. 2006;34(2):433–444. https://doi.org/10.3758/BF03193420.

Koch I, Poljac E, Müller H, Kiesel A. Cognitive structure, flexibility, and plasticity in human multitasking—An integrative review of dual-task and task-switching research. Psychological Bulletin. 2018;144(6):557–583. https://doi.org/10.1037/bul0000144.

Kuznetsova A, Brockhoff PB, Christensen RHB. ImerTest package: tests in linear mixed effects models. Journal of Statistical Software. 2017;82(13):1–26. doi: 10.18637/jss.v082.i13.

Lin C-T, King, J-T, Fan J-W, Appaji A, Prasad M. The influence of acute stress on brain dynamics during task-switching activities. IEEE Access. 2017;6:3249–3255. https://doi.org/10.1109/ACCESS.2017.2787673.

Logan G. Executive control of thought and action: in search of the wild homunculus. Current Directions in Psychological Science. 2003;12(2):45–48.

Meiran N. Reconfiguration of processing mode prior to task performance. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1996;22(6):1423–1442.

Meiran N. Reconfiguration of stimulus task sets and response task sets during task-switching. In Monsell, S, Driver J., Eds. Control of cognitive processes: attention and performance XVIII. MIT Press; 2000. p. 377–400.

Meiran N, Daichman A. Advance task preparation reduces task error rate in the cuing task-switching paradigm. Memory & Cognition. 2005;33(7):1272–1288.

Monsell S. Task switching. Trends in cognitive sciences. 2003;7(3):134–140.

Pallavicini F, Ferrari A, Mantovani, F. Video games for well-being: a systematic review on the application of computer games for cognitive and emotional training in the adult population. Frontiers in Psychology. 2018;9:2127. https://doi.org/10.3389/fpsyg.2018.02127.

R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing; 2020. https://www.R-project.org/.

Samavatyan H, Leth-Steensen C. The time course of task-switching: a speed–accuracy trade-off analysis. Memory & Cognition. 2009;37(7):1051–1058. https://doi.org/10.3758/MC.37.7.1051.

Squire PN, Parasuraman R. Effects of automation and task load on task-switching during human supervision of multiple semi-autonomous robots in a dynamic environment. Ergonomics. 2010;53(8):951–961. https://doi.org/10.1080/00140139.2010.489969.

Squire PN, Trafton G, Parasuraman R. Human control of multiple unmanned vehicles: effects of interface type on execution and task-switching times. AMC Press; 2006. p. 26–32. https://doi.org/10.1145/1121241.1121248.

Steyvers M, Hawkins GE, Karayanidis F, Brown SD. A large-scale analysis of task-switching practice effects across the lifespan. Proceedings of the National Academy of Sciences. 2019;116(36):17735–17740. https://doi.org/10.1073/pnas.1906788116.

Strobach T, Frensch PA, Schubert T. Video game practice optimizes executive control skills in dual-task and task-switching situations. Acta Psychologica. 2012;140(1):13–24. https://doi.org/10.1016/j.actpsy.2012.02.001.

Appendix A. List of Assets Used in the Classification Task

Threat Assets:

Infantry Fighting Vehicle

Large Tank

Manned Combat Vehicle

Robotic Combat Vehicle

Rocket Truck

Gun Truck

Small Tank

Non-Threat Assets:

Barricade

Dumpster

Gas Pump

Log Pile

Oil Drum

Rock

Sandbag Wall

Sedan

Destroyed Sedan

Stack of Crates

Pile of Tires

Tractor

Destroyed Truck

Utility Trailer

Van

Destroyed Van

Water Pump

List of Symbols, Abbreviations, and Acronyms

CI	confidence interval
CTI	cue-target-interval
DEVCOM	US Army Combat Capabilities Development Command
df	degrees of freedom
ERP	Essential Research Program
HAT	Human–Autonomy Teaming
LOS	line of sight
MSSQ	Motion Sickness Simulator Questionnaire
MUM-T	Manned–Unmanned Teaming
NASA-TLX	National Air and Space Administration – Task Load Index
NGCV	Next Generation Combat Vehicle
RAM	random access memory
SA	situational awareness
SE	standard error
SEM	standard error of the mean
sig	significance value
t	t-score

1 DEFENSE TECHNICAL
(PDF) INFORMATION CTR
DTIC OCA

1 DEVCOM ARL
(PDF) FCDD RLD DCI
TECH LIB

1 DEVCOM ARL
(PDF) FCDD RLH B
T DAVIS
BLDG 5400 RM C242
REDSTONE ARSENAL AL
35898-7290

1 DEVCOM ARL
(PDF) FCDD HSI
J THOMAS
6662 GUNNER CIRCLE
ABERDEEN PROVING
GROUND MD
21005-5201

1 USN ONR
(PDF) ONR CODE 341 J TANGNEY
875 N RANDOLPH STREET
BLDG 87
ARLINGTON VA 22203-1986

1 USA NSRDEC
(PDF) RDNS D D TAMILIO
10 GENERAL GREENE AVE
NATICK MA 01760-2642

1 OSD OUSD ATL
(PDF) HPT&B B PETRO
4800 MARK CENTER DRIVE
SUITE 17E08
ALEXANDRIA VA 22350

ABERDEEN PROVING GROUND

16 DEVCOM ARL
(PDF) FCDD RLH
J LANE
Y-S CHEN
P FRANASZCZUK
A MARATHE
K MCDOWELL
K OIE
FCDD RLH F
J GASTON (A)
FCDD RLH FA
A DECOSTANZA
DE FORSTER
FCDD RLH FB
D BOOTHE (A)
FCDD RLH FC
K COX (A)
TL ROHALY
FCDD RLH FD
A FOOTS (A)
KR COX
GM GREMILLION
FCDD RLH FE
D HEADLEY