Ergonomics

ISSN: 0014-0139 (Print) 1366-5847 (Online) Journal homepage: http://www.tandfonline.com/loi/terg20

Effects of automation and task load on task switching during human supervision of multiple semi-autonomous robots in a dynamic environment

P. N. Squire & R. Parasuraman

To cite this article: P. N. Squire & R. Parasuraman (2010) Effects of automation and task load on task switching during human supervision of multiple semi-autonomous robots in a dynamic environment , Ergonomics, 53:8, 951-961, DOI: 10.1080/00140139.2010.489969

To link to this article: https://doi.org/10.1080/00140139.2010.489969

Published online: 22 Jul 2010.

Submit your article to this journal

Article views: 511

Citing articles: 21 View citing articles

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=terg20

Ergonomics
Vol. 53, No. 8, August 2010, 951–961

Effects of automation and task load on task switching during human supervision of multiple semi-autonomous robots in a dynamic environment

P.N. Squire[a,b]* and R. Parasuraman[b]

[a]*Naval Surface Warfare Center, Dahglren Division, Dahlgren, Virginia, USA;* [b]*Psychology Department, George Mason University, Fairfax, Virginia, USA*

(*Received 21 December 2009; final version received 20 April 2010*)

The present study assessed the impact of task load and level of automation (LOA) on task switching in participants supervising a team of four or eight semi-autonomous robots in a simulated 'capture the flag' game. Participants were faster to perform the same task than when they chose to switch between different task actions. They also took longer to switch between different tasks when supervising the robots at a high compared to a low LOA. Task load, as manipulated by the number of robots to be supervised, did not influence switch costs. The results suggest that the design of future unmanned vehicle (UV) systems should take into account not simply how many UVs an operator can supervise, but also the impact of LOA and task operations on task switching during supervision of multiple UVs.

The findings of this study are relevant for the ergonomics practice of UV systems. This research extends the cognitive theory of task switching to inform the design of UV systems and results show that switching between UVs is an important factor to consider.

Keywords: automation; human factors; human–robot interaction; multitasking; task switching; unmanned vehicles

1. Introduction

Unmanned vehicles (UVs) are rapidly being introduced in both military (Office of the Secretary of Defense 2005) and civilian (Casper and Murphy 2003) environments. Many factors are driving the increased use of UVs: a need for extending battlefield capabilities; increasing force lethality; rescuing survivors of natural or manmade disasters, such as fires and earthquakes; reducing human exposure to combat operations (Casbeer *et al.* 2006, Parasuraman *et al.* 2007). Personnel reduction is another major motivating factor. Labour costs are increasing in many work environments and robotic systems such as unmanned ground vehicles and unmanned air vehicles (UAVs) offer a cost-effective way to perform many of the tasks performed by humans without the associated limitations or expense. However, at present, most UV systems, such as the General Atomics Aeronautical Systems *Predator*, require several personnel to operate a single vehicle (Cooke *et al.* 2006).

To offset the increases in personnel costs associated with current human–UV operations, the military and other organisations have directed research and development efforts toward the goal of a system configuration, where one or more operators can oversee several UVs (Pew and Mavor 2007). If this few-personnel-and-many-UVs goal is realised, then future operators of these emerging systems will likely experience increased task load as they are required to concurrently monitor and manage multiple UVs. To offset the increased task load, it is likely that effective human supervision of multiple UVs will mandate the use of automated support systems for human operators (Parasuraman *et al.* 2007). The level of automation (LOA) is not likely to be 'all or none', but will vary in type and level (Parasuraman *et al.* 2000), depending on the system. Additionally, it is unlikely that UVs will all perform the same task. As such, operators will have to switch between different UVs to monitor or perform different tasks. Accordingly, a requirement for task switching under varying levels of automation will be imposed on operators.

The study of the effects of switching between two simple tasks has a long history in psychological research (Jersild 1927). In the last two decades, there has been a resurgence of new research on this topic, driven in part by the ongoing debate on the specific cognitive processes underlying task switching (for a review of task switching paradigms and theories, see Logan 2003, Monsell 2003). An example of a typical switching paradigm is as follows. A participant is given a stimulus (e.g. a number) and, for a given trial, asked

*Corresponding author. Email: peter.squire@navy.mil

ISSN 0014-0139 print/ISSN 1366-5847 online
This material is declared a work of the United States Government and is not subject to copyright protection: approved for public release, distribution in unlimited.
DOI: 10.1080/00140139.2010.489969
http://www.informaworld.com

to process it according to a specified task rule (e.g. classify the number as even/odd or high/low). Across experimental trials, the participant repeats the same task (e.g. even/odd followed by even/odd) or is given a cue to switch tasks (e.g. even/odd followed by high/low). The typical result is that reaction times are longer and accuracy is lower on trials when the participant has to switch between tasks compared to when the same task is repeated. This performance decrement is known as a switch cost (Logan 2003, Monsell 2003).

Recently, a new task-switching paradigm, known as voluntary task switching, has been developed (Arrington and Logan 2004, 2005). This paradigm extends the traditional task-switching paradigm—a somewhat artificial laboratory phenomenon—to the more naturalistic situation of the participant (rather than the experimenter) choosing when to switch between tasks. A participant is provided this freedom of choice when instructed to perform each task equally often and in a random sequence. By allowing participants to freely decide when to switch actions, Arrington and Logan (2004) demonstrated that switch costs are still present when participants (rather than the experimenter) decided when to switch tasks.

Time costs for switching between different tasks are therefore present regardless of the initiating source, i.e. participant or experimenter. Accordingly, one can predict that switch costs will also influence the timing and efficiency of switching between different UV tasks, irrespective of whether the switch was initiated by the operator or by some external event. These switch costs may markedly influence performance and thus influence how many UVs an operator can supervise and/or the type of tasks an operator can perform effectively while supervising UVs. For example, suppose the operator has to perform two tasks within 20 s approve UAV 1 action and confirm an image obtained from UAV 2. If each task takes approximately 10 s to complete, one can assume that these actions could be completed within the required time frame. But if additional time is necessary to switch from UAV 1 to UAV 2 action or vice versa, then performance might suffer (i.e. a critical target image could be missed or a hostile threat could be lost). Because such switch costs could alter the configuration of human–UV teams and operations, it is important to understand their impact on human–UV system performance.

There has been limited prior work on switch costs in the context of human–robot interaction. Goodrich *et al.* (2005) had participants tele-operate a simulated ground robot to different target (geometric shapes) locations within a simulated environment and classify each shape encountered. Participants performed this task with the use of a game pad to control the robot near a geometric shape and then button presses to identify the shape by cycling through a set of shapes (sphere, cube or tetrahedron). Occasionally, participants were interrupted from the primary classification task. When interrupted, participants performed one of four secondary tasks of varying task difficulty (in ascending order of task difficulty): blank screen; tone counting; watch video; count number of unique cars or play a game of Tetris. Participants were informed that the geometric shapes could change or disappear while they were engaged in the secondary task and, that if such a change occurred when they resumed the primary task, they were to report it once it was detected. The response time to report the shape change after resuming the primary tele-operation task indicated switch costs. Goodrich *et al.* (2005) found evidence of switch costs in this simulated environment and also reported that costs increased with task difficulty, consistent with other studies using simple laboratory tasks (e.g. Rubinstein *et al.* 2001).

Thus, there is some evidence that the difficulty of a single task that interrupts an ongoing task influences switch costs, but the effect of overall task load on switch costs is less clear. For example, the number of UVs needed to carry out a given mission could influence the task load and thus the operator's workload. Prior work has focused mainly on tele-operation, where the operator is involved in supervisory manual control of the UV (Di Nocera *et al.* 2005, Goodrich *et al.* 2005, Chen *et al.* 2007, Chen and Terrence 2009). However, UV technical developments, as well as the requirement to allow one or more operators to oversee several UVs, are changing control methods from tele-operation to supervision of UVs capable of more autonomous movements. These trends have fuelled discussions and metrics concerning the number of robots that a supervisor can effectively oversee (Crandall *et al.* 2005, Cummings and Guerlain 2007, Hancock *et al.* 2007).

Supervising multiple UVs will consequently increase the cognitive workload demands on operators and automation will be needed to support effective and safe human–system performance. Although the requirement for supervisory control of multiple UVs makes the use of automation mandatory, the level and type of automation (Parasuraman *et al.* 2000) that should be used to provide for effective support of the operator is less well understood.

Moreover, like task load, the influence that different LOA may have on switch costs has received little attention. One relevant study, conducted by Di Nocera *et al.* (2005), examined the performance of participants immersed in a simulated micro-world, who were instructed to oversee the exploration of an area on Mars. The primary task was to monitor the overall health of space station functions by detecting the occurrence of a fault and recovering from a fault. Station functions were monitored under four different LOA: (1) no automation; (2) system notification, where an automated notification was provided about events deviating from normal operating behaviour; (3) notification + suggestion, where, in addition to providing a notification, the automation also provided a possible solution; (4) system action, where automation took control of the task from the operator and sent a notification message reporting what was done. Each participant was randomly assigned a LOA condition, for which they received 40% of their support; the remaining support was provided equally across the other three LOA. When participants moved between the assigned and other LOA, a shift or switch in LOA occurred. In addition to switches in LOA, each participant's workload was manipulated. For low workload conditions (single task), participants only monitored station functions and for high workload (dual task), participants monitored station functions and tele-operated a robot. When operating within the assigned LOA (no-switch condition), response times to faults were modulated by a workload by station function type interaction rather than by LOA. Furthermore, during low workload conditions only, participants responded faster to detection faults during no-switch than during switch conditions, indicating that performance costs only occur when participants shifted (i.e. switched) between LOA. Although the LOA were intrinsically different, the authors argued that no costs were observed in no-switch conditions because participants were able to adjust their expectations according to their assigned LOA. In contrast, switch costs occurred because participants had to disengage from one cognitive-behavioural set and engage another set.

If, as the Di Nocera *et al.* (2005) study suggests, participants are able to switch their actions according to their assigned LOA, one would predict that different LOA—low vs. high—should not affect task-switching performance. Alternatively, different LOA might well affect task switching, given that previous research has suggested that LOA influences the type of information processing engaged in by human–machine system operators. In particular, the out-of-the-loop-unfamiliarity hypothesis (Wickens *et al.* 2003) suggests that with increased LOA, operators shift from an active information-processing mode to passive processing, such that they are impaired if they are required to return to manual performance (Endsley and Kiris 1995). This would predict a higher task-switching cost for a high rather than a low LOA.

The present study examined the effects of task load and automation on task switching during supervision of simulated robots in dynamic environment. The study used RoboFlag, which is a high-fidelity simulation of the actions and states of hardware associated with robots capable of autonomous cooperative behaviours in games such as soccer, tag and capture the flag (D'Andrea and Babish 2003). This simulation was previously used in a study examining the effects of adaptable automation on human supervision of multiple robots (Parasuraman *et al.* 2005). The present study examined task-switching performance under different levels of task load (number of robots supervised) and LOA. First, it was predicted that similar task-switching costs would be found in this more complex, dynamic task, as found in simpler laboratory tasks examined in previous studies of voluntary task switching (Arrington and Logan 2004). It was expected that the time to switch between tasks would be longer than when the same task was repeated. These switch costs should change with changes in LOA. If, as previously suggested, different LOA influence processing type (active vs. passive) then switch costs were predicted to increase with LOA. Finally, it was predicted that, as previously found with manipulation of task difficulty, increases in task load would also increase switch costs.

2. Methods

2.1. *Participants*

A total of 12 young adults (three males and nine females) between the ages of 18 and 27 (mean 21.1, SE 2.7) years served as paid participants. All participants had normal or corrected to normal vision.

2.2. *Apparatus*

The Roboflag simulation was run under two separate personal computers (PCs) communicating via transmission control protocol/internet protocol (TCP/IP) protocol. A participant used one PC; the other PC was used by the opposing team. A central processing executive (the 'arbiter') was responsible for passing all communication and control information for each team and collected the logging data. The RoboFlag simulation was used because it accurately captures the actions and states of hardware robots capable of several autonomous actions and cooperative behaviour (D'Andrea and Babish 2003). Thus, the simulated robots reflect the real-world characteristics of hardware robots possessing limited fuel, vision and speed. Each robot was controlled by a separate program and did not require a centralised entity to manage its interactions with other robots. Inter-robot communication and negotiation were used for collaboration and cooperative behaviours.

For the purposes of this study, the RoboFlag simulation was modified to allow a single participant to compete against an automated opponent in a game of ‘capture the flag’. The field of engagement was divided in half, one for each team (see Figure 1). Each participant was instructed to capture the other team’s flag (represented by a white dot and located in the centre of the circle area in each half) and bring it back to his or her own territory, while simultaneously protecting their own flag from capture. Thus, the game is a mix of offensive and defensive tasks, including securing the opponent’s flag and preventing the opponent from securing one’s own flag.

Participants viewed and tasked their robot team via a graphical user interface (GUI). The GUI shown in Figure 1 provides information on each robot’s location and field of view (light grey area surrounding each robot), status (active, flagged, tagged or inactive), fuel level and what (if any) automated play is executing. Depending on the LOA condition used, the operators tasked robots via manual point-and-click (waypoint) or by using automated ‘play’ commands (Miller and Parasuraman 2007).

For automated conditions only, three plays were provided to the participant: circle offence; circle defence; patrol border. A circle offence play instructed the robots to circle the opponent’s team flag area until the flag could safely be captured. Once the flag was captured, the robots were to return to midfield. In the circle defence play, the robots circled around the flag area and attempted to prevent opposing forces from entering and extracting the flag. Finally, in the patrol border play, robots were tasked to manoeuvre up and down the midfield line and engage any opponent robot that crossed the line.

Both manual and plays tasks could be applied to a group of robots of any size. Individual robot selection was accomplished by moving the mouse cursor to a specific robot and selecting that robot—selected robot(s) would then be highlighted. Group robot selection was accomplished in a manner similar to cropping an image, a box was drawn for a specific region of interest and those robots within that region were selected.

2.3. Procedure

Participants received no specific guidance on strategy, but instead received general instructions to enable them to achieve the overall objective of capturing the

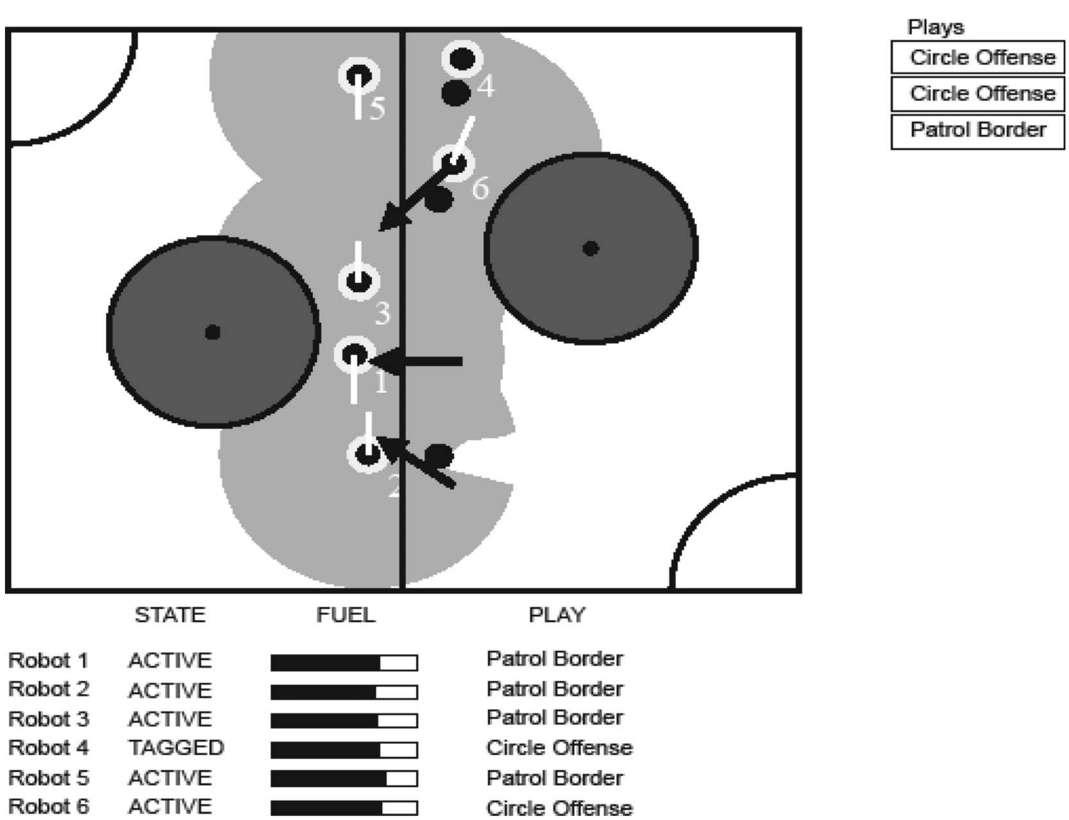

Figure 1. Roboflag simulation of participant interface at a single moment in time. Note: RoboFlag ‘play’ condition interface with opponent’s offensive routes shown. Play selection buttons are on the top right-hand corner and robot state and fuel status, with corresponding assigned plays, are at the bottom.

opponent's flag and preventing capture of their own flag. Participants were taught the general instructions for the simulation, which focused on how robots were selected and moved, the viewable area for robots and game rules. Furthermore, they were told they would supervise a team of four or eight robots in a game of 'capture the flag' and would play against an automated team with an equal number of robots and identical characteristics (e.g. fuel, speed, vision). The automated opponent used both offensive and defensive plays. For this scripted 'mixed' posture, half the available robots were selected for offensive plays and half for defensive plays. The attacking route used for offensive plays was random and varied between three different routes, as illustrated in Figure 1. For defensive plays, both patrol border and circle defensive plays were used. After these instructions, participants practised with each of the different conditions to ensure that they were comfortable with the simulation.

Participants received training for each condition prior to data collection and before each trial participants were informed about the conditions (e.g. number of robots available in that condition). An example of a typical Roboflag trial was as follows: (1) robots would start at the default home positions located at the corner of each team's half (see Figure 1); (2) robots would deploy. For the automated robot team, a script would assign plays to the robots; for the human robot team, depending on the supervision method, an operator would select waypoints for the robots (i.e. manual) or select automated plays; (3) robots would then engage each other. The teams attack and defend against each other. During the course of a game, robots are tagged and required to return to the default start position before they can become active again, are inactive if they run out of fuel and cannot be used for the rest of the trial or if they capture the flag. They are then flagged and if they successfully make it back to the mid-line, they win the game, ending the trial.

2.4. Design

In a within subjects-design, two LOA (low, high) were combined with two levels of task load (low, high). Each participant completed 10 trials for each condition for a total of 40 trials. LOA was a block factor, counterbalanced across participants, and referred to the method for supervising robots: waypoint control (low LOA) or automated play (high LOA). Task-load conditions were randomised within each block and referred to the number of robots for the participant and scripted opponent: four (low) or eight (high).

Performance data were collected for each trial. The critical data for this study were the tasks that participants assigned to robots and the time they took to do so. Recall that the game of 'capture the flag' is a mix of offensive (capture opponent's flag) and defensive goals (prevent capture of own flag). To achieve these goals, participants assigned robots a task, defined as either an offensive or defensive behaviour intention. For the low LOA condition, the endpoint location defined an offensive or defensive task. If a robot's end waypoint location (x and y coordinates) was on the offensive side of the playing field, the task was defined as offensive. If the location was on the defensive side of the playing field, the task was classified as defensive. For the high LOA condition, an offensive task was defined as the selection of the offensive play: circle offence. A defensive task was defined as the selection of one of two defensive plays: circle defence or patrol border. Thus, each robot task had an offensive or defensive classification and a time stamp as to when the task was assigned.

During the course of a single trial, participants performed a number of either offensive or defensive tasks. If a participant performed the same task, e.g. an offensive task followed by another offensive task, that sequence was defined as a no-switch action. However, if a participant executed a different task, e.g. offensive task followed by a defensive task, or vice versa, that sequence was defined as a switch action. Since the first task performed during each trial did not have a preceding task to which it could be referenced, these were not included in the analyses (see Table 1 for example computations). Based on these definitions, switch or no-switch effects were coded for all trials and conditions.

Other objective data measures were also collected; overall mission completion time and mission success (i.e. percentage of wins). Participants were also asked to provide mental workload (similar to the NASA-TLX; Hart and Staveland 1998) and situation awareness (similar to the SART: Taylor 1990) ratings. These ratings differed from the NASA-TLX and SART instruments. Specifically, the subscales/dimensions for the NASA-TLX and SART were combined into a single question for mental workload and situation awareness. After each trial, participants

Table 1. Example of switch and non-switch actions and times.

Task (No-switch/Switch)	Total time	Switch time	No-switch time
Defensive (first action)	2.17	–	–
Defensive (No-switch)	3.83	–	1.66
Defensive (No-switch)	4.5	–	0.67
Offensive (Switch)	6	1.5	–
Offensive (No-switch)	6.57	–	0.57
Defensive (Switch)	9.7	3.13	–

provided a single rating on a 100-point scale (0 = low to 100 = high). The modified instruments were used for convenience purposes and because there is previous evidence that the simpler 1-D measures are moderately correlated with the longer multi-dimensional measures (Parasuraman *et al*. 2009) and thereby reduce the time, as well as annoyance and frustration that participants might experience filling out numerous questionnaires items after each.

Table 2. Best- and worst-case movement time predictions.

Fitts' Law calculations	Total (s)	A value (inches)	W value (inches)
Best (high LOA)	0.43	0.80	0.63
Worst (high LOA)	1.30	9.8	0.63
Best (low LOA)	0.25	0.50	4.6
Worst (low LOA)	0.49	9.2	4.6

A value = distance of the movement from start to target centre; W value = width of the target; LOA = level of automation.

3. Results

Before switch times could be analysed and the main experimental hypotheses tested, two preliminary analyses had to be conducted: (1) the contribution of motor movement times to response times and to switch times; (2) the effectiveness of the task load manipulation on subjective and task performance measures. Following these analyses, the influence of low and high LOA and task load on the frequency and times associated with switch and no-switch actions were analysed.

3.1. *Motor movement evaluation*

Task-switching time included the time participants took to execute motor movements (i.e. using the mouse to click on the interface to select a waypoint or a play). To assess to what extent variation in such movement times contributed to the overall reaction times used to calculate task switching time, Fitts' Law was used to calculate upper and lower bounds for movement time for participants. The specific version of Fitts' Law, known as Shannon's formulation, was used:

$$MT = a + b\ log\,2(A/W + 1),$$

where MT is the movement time, a and b are the regression coefficients for a particular user and mouse, A is the distance of the movement from start to target centre and W is the width of the target (Soukoreff and MacKenzie 2004). Because a Fitts' test was not performed with the participants in this experiment, the regression coefficients used to estimate motor movement time were for derived pointing models for the mouse by MacKenzie (1992).

The movement times for the best- (lower bound) and worst-case (upper bound) conditions during the low and high LOA conditions are provided in Table 2. The best-case movement time for the high LOA (play) represented a situation where the participant made a very small motor movement, such as when the operator changed a play without deselecting a robot. For example, if the operator switched from a circle offence play (topmost play in Figure 1) to a circle defence play, the operator only needed to move 0.80 inches on the GUI. If, however, after selecting a play, the operator selected another robot (at the furthest point at the bottom left-hand corner of his or her area) and then selected a play for this robot, those actions required a total motor movement of 9.8 inches. This value represents a worst-case situation for the high LOA condition.

Under the low LOA condition (waypoint control), a best-case scenario occurred when a participant selected a waypoint very close to where the robot was currently located. This distance is similar to the best-case scenario for the high LOA condition but the width of the target is much larger for the low LOA condition and, therefore, movement time is faster. A worst-case scenario for the low LOA condition is similar to the worst-case scenario of the high LOA condition. Again, the movement time is faster for the low LOA condition because the width of the target is larger.

This analysis indicated that the average movement time to execute a task was relatively short. The predicted movement times for the high and low LOA conditions (best and worst cases) ranged from 0.25 to 1.3 s and when averaged together the mean value was 0.62 s. If the motor movements were the only factor associated with no-switch and switch actions then the observed (experimental conditions) and predicted (motor movement) times should be similar. However, as shown later, these predicted times are less than the observed no-switch and switch actions times. Intuitively, this should make sense, because no-switch and switch actions times are a combination of both physical and cognitive operations.

3.2. *Evaluation of task load manipulation*

To observe the effectiveness of the primary task load manipulation (i.e. number of robots supervised and pitted against), subjective and task performance measures were examined. These measures indicated that the number of robots (and thus level of imposed task load) influenced both objective and subjective measures. Subjective workload was higher when supervising and competing against a team of eight robots (mean 52.10, SE 5.67) than four robots (mean 43.46, SE 5.38), $F(1,11) = 25.61$, $p < 0.001$. Participants reported higher levels of situation awareness when supervising and competing against an opponent with four robots (mean 74.20, SE 3.71) rather than eight robots (mean 67.03, SE 4.18), $F(1,11) = 27.14$, $p < 0.001$.

In addition to the examination of subjective measures, task performance measures such as mission success (number of trials won) and mission completion times (length of a trial) were also examined. Regardless of LOA ($F < 1$), participants won fewer games with eight robots (mean 35%, SE 0.17) than with four robots (mean 55%, SE 0.25), $F(1,11) = 22.00$, $p < 0.001$ (as illustrated in Figure 2).

Mission times were also significantly affected by the number of robots and LOA (see Figure 3). For all trials, times were significantly longer when supervising and competing against an opponent with eight robots (mean 54.04 (s), SE 3.23) than against a team of four robots (mean 46.20 (s), SE 2.88), $F(1,11) = 15.59$, $p < 0.005$. In addition, supervising the robots under a high LOA resulted in significantly longer trial times (mean 54.48 (s), SE = 3.09) than for a low LOA (mean 45.77 (s), SE 3.46), $F(1,11) = 7.91$, $p < 0.05$.

In summary, the objective task-performance measures gave similar results to those for the subjective measures. Specifically, as the number of robots increased, both subjective and task performance measures were affected. Subjectively, participants reported increased mental workload and decreased situation awareness and, behaviourally, participants experienced decreased mission success and increased mission times when supervising a larger number of robots. Therefore, the task-load manipulation was effective in influencing a participant's subjective and behavioural performance.

3.3. *Task switching*

To examine how task load, LOA and type of action affected the frequency of actions, a repeated measure ANOVA was conducted with factors of task load, LOA and action type. Main effects for action type, $F(1,11) = 76.88$, $p < 0.001$ and task load $F(1,11) = 35.618$, $p < 0.001$ were obtained. In addition, there was an interaction for task load $\times$ action type, $F(1,11) = 15.02$, $p < 0.005$ (see Figure 4). Tukey *post-hoc* testing indicated that operators performed more no-switch actions with eight robots (mean 18.1(actions), SE 1.79) than with four robots (mean 11.87 (actions), SE 1.14), although no significant difference ($p = 0.72$) was observed for switch actions between eight (mean 2.83 (actions), SE 0.31) and four robots (mean 4.34 (actions), SE 0.35). These results indicate that, whereas LOA had no effect on the number of actions performed, the number of robots and type of action did influence the frequency of actions.

3.4. *Switch time*

Motor performance estimates indicated that variability in motor times should have a negligible effect on performance. To examine how task load, LOA and type of action influenced performance times, a repeated measure ANOVA was conducted on task load $\times$ LOA $\times$ action type. Mean times for switch and no-switch actions per trial were computed and examined (see Figure 5). There was a significant interaction between LOA and action type $F(1,11) = 15.968$, $p < 0.005$ and main effects for LOA $F(1,11) = 23.735$, $p < 0.001$ and for action type $F(1,11) = 24.12$, $p < 0.001$. No interaction or main effects were obtained for task load, i.e. number of robots.

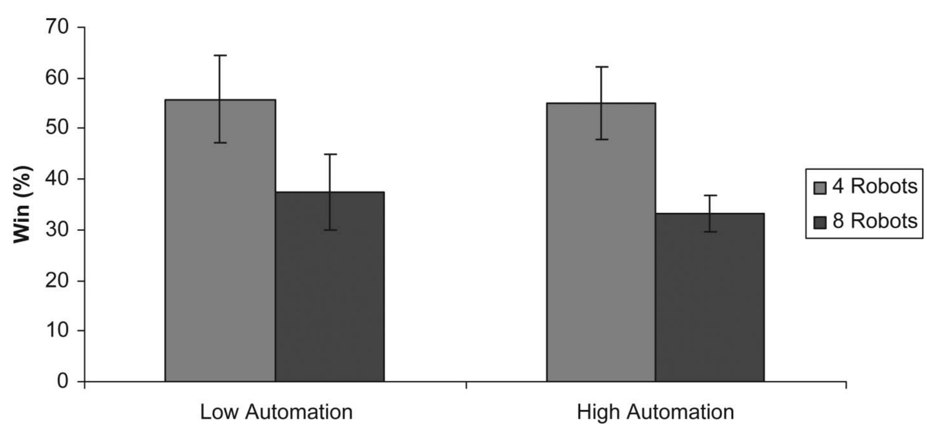

Figure 2. Mission success (percentage of wins) for level of automation by number of robots.

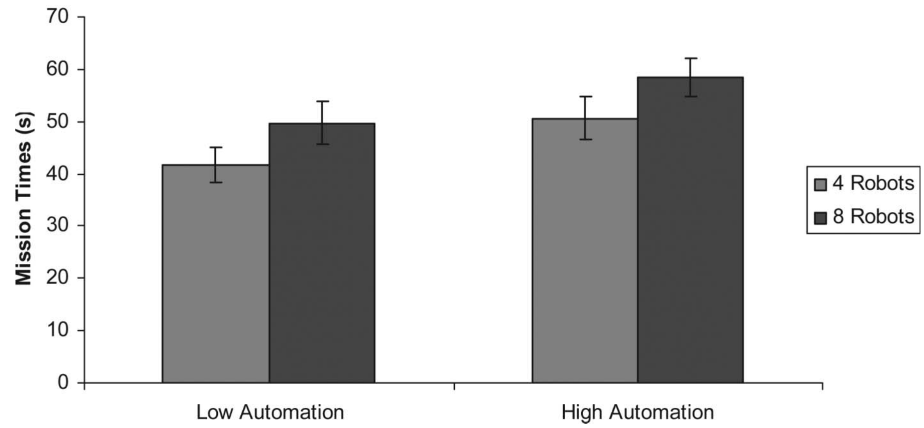

Figure 3. Mission times for level of automation by number of robots.

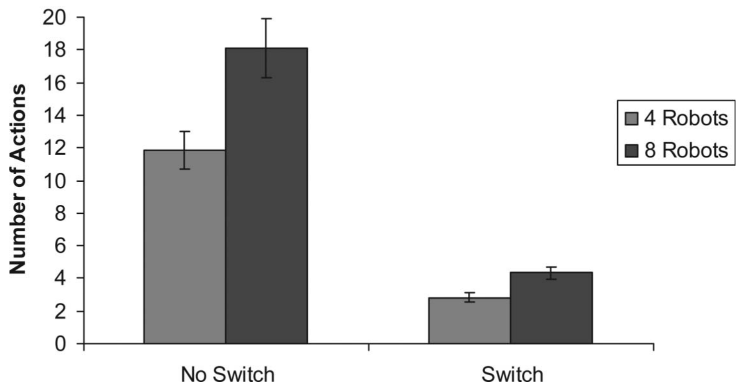

Figure 4. Number of actions for action type by number of robots.

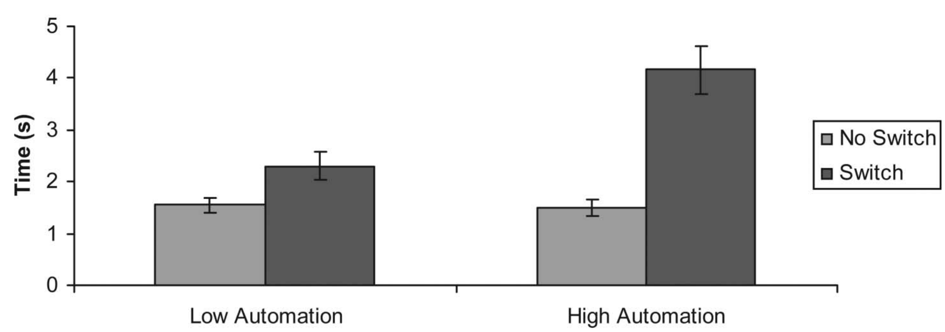

Figure 5. Switch or no-switch times for level of automation.

Main effects showed that: (1) when operators switched actions (mean 3.23 (s), SE 0.340) there was a time cost when compared to no-switch actions (mean 1.52 (s), SE 0.97); (2) action times were higher when operators managed robots at a high LOA (mean 2.83 (s), SE 0.240) than at a low LOA (mean 1.92 (s), SE 0.158). Tukey *post-hoc* testing for the LOA by action type interaction indicated two simple effect differences. Performance times for high (mean 1.496 (s), SE 0.17) and low automation conditions (mean 1.551(s), SE 0.15) were not significantly different for no-switch actions; but for switch actions, the high LOA condition (mean 4.160 (s), SE 0.47) increased performance times significantly more than the low LOA condition (mean 2.31(s), SE 0.266).

4. Discussion

The present study examined the effects of LOA and task load on task switching in a dynamic environment simulating human supervision of multiple semi-autonomous robots. Consistent with simpler laboratory experiments (Logan 2003, Monsell 2003), switch costs were also seen in the more complex and dynamic Roboflag environment. Participants took longer to switch between tasks (e.g. from offence to defence or from defence to offence), compared to when they repeated the same task (e.g. from offence to offence or from defence to defence). The results (see Figure 5) showed that the times for switch actions were relatively long—on the order of 2–5 s—and represent a substantial cost for participants, because the times were significantly longer than the times for no-switch actions ($\sim$ 1 s).

The switch costs found in this study are most likely not an artefact related to simple motor movements but represent a true cognitive cost of switching between different tasks. Several lines of evidence support this view. First, switch times were longer than the worse case movement time ($\sim$1.3 s) required to execute a task. Next, strategy development or planning should have minimal, if any, influence on switch cost values. The mean time for a trial was roughly less than 1 min and in a dynamic environment such as Roboflag, where the environment changes rapidly and where uncertainty is great because of the limited robotic vision, delaying tasking can have serious consequences for the outcome. Participants therefore did not have the luxury of time to develop elaborate strategies or plans, but needed to act quickly to ensure a successful outcome. Finally, the switch cost values obtained in the present study, though considerably larger than those in simple laboratory tasks involving voluntary task switching (e.g. Arrington and Logan 2004), are comparable to those reported in previous studies using complex multitask and robot simulations (Goodrich *et al.* 2005). Previous studies, however, experimentally manipulated when operators would switch between tasks by using interruptions (Goodrich *et al.* 2005) or with a blocked design (Di Nocera *et al.* 2005). In contrast to these studies, and consistent with what would be likely in real-world situations operations, in this study participants were given the freedom to decide when to switch. Thus, even though the timing of switching was not artificially controlled by the experimenter in this study, the present results indicate that switch costs occur in conditions more representative of naturalistic task environments.

The results also showed that participants preferred to repeat (e.g. offence to offence) than to switch tasks (e.g. offence to defence). Although participants were not explicitly instructed to split their time between offensive and defensive tasks, they were informed that each task was equally important. Experimental observation confirmed that all participants delegated robots to both a defensive and offensive task. Previous studies using laboratory tasks in which participants are directed to perform each task equally show a similar perseveration tendency—reuse rather than change a plan of action. These studies also suggest that the preservation tendency may be driven by an individual's bias or by environmental factors (Mayr and Bell 2006). Within the Roboflag simulation environment, an individual bias would influence whether an operator has a preference for focusing predominately on offensive or defensive tasking or for switching equally between the two tasks. Alternatively, the decision to switch tasks may be driven by an environmental factor, such as an interrupting event. For example, if an opponent was attacking the flag, such an event might interrupt the operator from offensive tasking. Given the desire to examine performance under conditions similar to naturalistic work performance, such factors were not directly manipulated in this study. Both factors may have contributed to participants' task switching decisions and are, therefore, likely to influence overall performance. Future research, however, is necessary to determine if the switch costs differ for self-initiated vs. interruption events. Such research could be performed with the use of a verbal protocol analysis to identify when participants were initiating the switch themselves, or responding to an interrupting event.

The results also supported the prediction that switch costs vary with LOA. First, the current findings were consistent with the results of Di Nocera *et al.* (2005). No-switch task times did not significantly differ at the low or high LOA. In contrast to these no-switch conditions, an effect on switch times was predicted because of the influence of LOA on processing type (active vs. passive). In support of this prediction, switch cost times were longer at a high than low LOA. Findings provide evidence for the out of the loop unfamiliarity hypothesis (OOTLUF; Wickens *et al.* 2003) and claim that more time is needed to intervene under automated than under manual control, because a person needs to first regain awareness of the state of the system.. Alternatively, these findings might result from 'clumsy automation', which is known to increase physical workload when automation is difficult to engage, understand or bypass (Wiener 1988). Previous research with Roboflag suggests that participants sometimes perceived the automation to be limited or clumsy (Parasuraman *et al.* 2007). Results from this study appear to support that claim, as the number of interactions did not differ between high and low LOA, and suggest that participants were continually 'tinkering' with the automation as much as the manual condition. Additional research is needed, however, to resolve the question of whether switch costs are due to clumsy automation, OOTLUF or both.

In addition to LOA, it was predicted that increases in task load would also increase switch costs. In contrast to the prediction, task load, as manipulated by the number of robots, did not influence switch costs. This was not due to possible insensitivity of the task load manipulation: Other performance indicators showed that an increase in task load led to lower situation awareness and higher mental workload, reduced mission success and increased mission times. These findings diverge from previous research examining the influence of task difficulty on switch costs, where increases in task difficulty have been found to increase switch costs (Rubinstein *et al.* 2001, Goodrich *et al.* 2005).

The task-switching measure used in the present study extends the more naturalistic laboratory studies recently performed (Arrington and Logan 2004) to a more applied domain. Consider a scenario in which a military unit has left a ship to go ashore and obtain a high-value asset. A lone UAV operator supervising two UAVs is responsible for providing surveillance to the military unit and the ship, which is located near a shoreline of a known threat. The operator must switch back and forth between those UAVs, providing surveillance for the forward deployed unit and the UAVs that are providing surveillance for the ship. In this instance, the operator must perform tasking akin to the offence and defence tasking provided in the Roboflag simulation: using the UAV to infiltrate enemy units and defend the ship from possible threats.

The findings of this study have implications for the design of future UV systems. One conclusion is that designers should be less concerned with how many UVs an operator can supervise and focus instead on the required LOA and number of different tasks an operator needs to perform with the UVs. For example, an operator controlling UAV 1 and UAV 2 can perform tasks more quickly if the tasks are the same rather than different, regardless of LOA. In contrast, performance time is slowed if different tasks are required for UAV 1 and UAV 2, and if those UVs are operating at a high rather than low LOA. Many of the operations for which UVs are deployed are critical and require rapid response to address potential threats, e.g. military, fire or biohazard. Therefore, designers of these future systems should consider the implications of these factors on the configuration of human–UV teams and operations.

Acknowledgements

The authors wish to thank the George Mason University Interruption Group, Archlab Writing Group and Owen Seely for feedback and comments on this manuscript. This research was supported in part by the Naval Research Independent Laboratory In-House Research (IHR) and Independent Applied Research (IAR) Programs to P.S.; and the Army Research Laboratory under Contract 6004.005.01-C and Air Force Office of Sponsored Research MURI Grant to R.P.

References

Arrington, C. and Logan, G., 2004. The cost of a voluntary task switch. *Psychological Science*, 15 (9), 610.

Arrington, C. and Logan, G., 2005. Voluntary task switching: chasing the elusive homunculus. *Journal of Experimental Psychology, Learning, Memory and Cognition*, 31 (4), 683–702.

Casbeer, D., *et al.*, 2006. Cooperative forest fire surveillance using a team of small unmanned air vehicles. *International Journal of Systems Science*, 37 (6), 351–360.

Casper, J. and Murphy, R., 2003. Human-robot interactions during the robot-assisted urban search and rescue response at the World Trade Center. *IEEE Transactions on Systems, Man and Cybernetics-Part B*, 33 (3), 367–385.

Chen, J.Y.C., Haas, E., and Barnes, M., 2007. Human performance issues and user interface design for teleoperated robots. *IEEE Transactions on Systems, Man and Cybernetics-Part B*, 37 (6), 1231–1245.

Chen, J.Y.C. and Terrence, P.I., 2009. Effects of imperfect automation and individual differences on concurrent performance of military and robotics tasks in a simulated multitasking environment. *Ergonomics*, 52 (8), 907–920.

Cooke, N., *et al.*, eds., 2006. *Human factors of remotely operated vehicles*, vol 7. Stamford, CT: JAI Press.

Crandall, J., *et al.*, 2005. Validating human-robot interaction schemes in multitasking environments. *IEEE Transactions on Systems, Man and Cybernetics-Part A*, 35 (4), 438–449.

Cummings, M. and Guerlain, S., 2007. Developing operator capacity estimates for supervisory control of autonomous vehicles. *Human Factors*, 49 (1), 1–15.

D'Andrea, R. and Babish, M., 2003. The RoboFlag testbed. *Proceedings of the American Control Conference*, 1, 656–660.

Di Nocera, F., Lorenz, B., and Parasuraman, R., 2005. Consequences of shifting from one level of automation to another: main effects and their stability. *Human Factors in Design, Safety, and Management*. Maastricht, The Netherlands: Shaker Publishing, 363–376.

Endsley, M. and Kiris, E., 1995. The out-of-the-loop performance problem and level of control in automation. *Human Factors*, 37 (2), 381–394.

Goodrich, M.A., Quigley, M., and Cosenzo, K., 2005. Task switching and multi-robot teams. *Proceedings of the third international multi-robot systems workshop*. 14–16 March, Washington, DC.

Hancock, P., *et al.*, 2007. Provocation: Is the UAV control ratio the right question? *Ergonomics in Design*, 15 (1), 7–31.

Hart, S.G. and Staveland, L.E., 1998. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. *In*: P.A. Hancock and N. Meshkati, eds. *Human mental workload*. Amsterdam: Elsevier Science/North Holland, 139–183.

Jersild, A.T., 1927. Mental set and shift. *Archives of Psychology*, 89, 5–82.

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

MacKenzie, I., 1992. Fitts' Law as a research and design tool in human-computer interaction. *Human-Computer Interaction*, 7 (1), 91–139.

Mayr, U. and Bell, T., 2006. On how to be unpredictable: evidence from the voluntary task-switching paradigm. *Psychological Science*, 17, 774–780.

Miller, C. and Parasuraman, R., 2007. Designing for flexible interaction between humans and automation: Delegation interfaces for supervisory control. *Human Factors*, 49, 57–75.

Monsell, S., 2003. Task switching. *Trends in Cognitive Sciences*, 7 (3), 134–140.

Office of the Secretary of Defense, 2005. *Unmanned aircraft systems (UAS) roadmap, 2005–2030*. Washington, DC: Department of Defense.

Parasuraman, R., Barnes, M., and Cosenzo, K., 2007. Adaptive automation for human-robot teaming in future command and control systems. *International Journal of Command and Control*, 1 (2), 43–68.

Parasuraman, R., Cosenzo, K., and de Visser, E., 2009. Adaptive automation for human supervision of multiple uninhabited vehicles: Effects on change detection, situation awareness, and mental workload. *Military Psychology*, 21 (2), 270–297.

Parasuraman, R., Sheridan, T.B., and Wickens, C.D., 2000. A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics-Part A*, 30, 286–297.

Parasuraman, R., *et al*., 2005. A flexible delegation-type interface enhances system performance in human supervision of multiple robots: empirical studies with RoboFlag. *IEEE Transactions on Systems, Man and Cybernetics-Part A*, 35 (4), 481–493.

Pew, R.W. and Mavor, A.S., eds., 2007. *Human-system integration in the system development process: A new look*. Washington, DC: National Academy of Sciences.

Rubinstein, J., Meyer, D., and Evans, J., 2001. Executive control of cognitive processes in task switching. *Journal of Experimental Psychology*, 27 (4), 763–797.

Soukoreff, R. and MacKenzie, I., 2004. Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI. *International Journal of Human-Computer Studies*, 61 (6), 751–789.

Taylor, R.M., 1990. *The development of a tool for aircrew systems design*. No. AGARD-CP-478. Neuilly sur Seine, France: NATO-AGARD.

Wickens, C., *et al*., 2003. *Introduction to human factors engineering*. Upper Saddle River, NJ: Prentice-Hall, Inc.

Wiener, E.L., 1988. Cockpit automation. *In*: E.L. Wiener and D.C. Nagel, eds. *Human factors in aviation*. New York: Academic, 433–461.