2006 CCRTS

The State of the Art and the State of the Practice

Instant Messaging and Team Performance in a Simulated Command and Control Environment

Topics: Teamwork, Collaborative Technologies, Synthetic Task Environments

Gregory J. Funke*
Scott M. Galster
W. Todd Nelson
Allen W. Dukes

POC: Scott M Galster
Air Force Research Lab
2255 H Street
WPAFB, OH 45433
(937) 255-8737
(937) 255-8752
scott.galster@wpafb.af.mil

Instant Messaging and Team Performance in a Simulated Command and Control Environment

Gregory J. Funke
General Dynamics
Dayton, Ohio, USA

Scott M. Galster
W. Todd Nelson
Allen W. Dukes
Air Force Research Laboratory
Wright-Patterson Air Force Base, Ohio, USA

Recent trends in military acquisition have emphasized the desire to introduce collaboration technologies into the command and control environment (Kaufman, 2005). Personnel in this environment may be disparate in terms of rank, occupation, and even geographical location, yet are expected to rapidly coalesce into functioning teams in order to meet task requirements (Boiney, 2005). It has been proposed (e.g., Alberts & Hayes, 2003) that integration and performance may be facilitated through emerging collaborative technologies, such as email, instant messaging (IM), virtual whiteboards, and videoconferencing. Proponents of network-centric operations have argued that these technologies might engender a degree of command decentralization that would result in increased situational awareness and task flexibility (Alberts & Hayes, 2003). However, there are growing concerns about the potential negative impact on performance associated with the use of collaborative tools in distributed team environments.

For example, in a synthesis of the collaboration technology literature, Bordia (1997) examined the effects of text-based collaboration technologies (i.e., email and IM) on team processes. Specifically, Bordia (1997) examined 18 experimental studies which compared team decision making effectiveness, time to reach a decision, and team member satisfaction between teams restricted to collaboration technologies for communication and teams that employed face-to-face communication. Bordia (1997) concluded that teams restricted to collaboration technologies made poorer decisions, measured both objectively (e.g., meeting task goals) and subjectively (e.g., quality of solutions), and took more time to reach a decision; moreover, team members experienced less satisfaction with team processes. Interestingly, this pattern of results was observed across different task types, suggesting the ubiquity of the observed effects.

In addition, Baltes, Dickson, Sherman, Bauer, and LaGanke (2002) completed a meta-analysis exploring performance differences between collaboration technology and face-to-face teams. In their meta-analysis, Baltes et al. (2002) examined 27 experimental studies of text-based collaboration technologies, analyzing them on dimensions similar to those employed Bordia (1997). The results of the meta-analysis were similar to the conclusions reached by Bordia (1997), dramatically favoring face-to-face communication over communication mediated by collaboration technologies across many different tasks.

Though the experiments reviewed by Bordia (1997) and Baltes et al. (2002) cover an array of tasks, most of them use remarkably similar *task types* in their research. Referring to McGrath's (1984) circumplex model, task type can be said to fall within one of four quadrants corresponding to: (a) generating; (b) choosing; (c) negotiating; and (d) executing. Tasks from the generating quadrant require cooperative conception of plans and ideas. Tasks from the choosing quadrant require problem solving in situations with and without correct answers. Tasks within the negotiating quadrant involve participants resolving conflicts of viewpoint or interest. Finally, execution quadrant tasks involve competition (both inter- and intra-team) or performance measured against a standard of excellence. When the experimental tasks reviewed by Bordia (1997) and Baltes et al. (2002) are described in reference to the circumplex model, they overwhelmingly fall within the *choosing* quadrant. This is unfortunate in that, as McGrath (1984) has pointed out, tasks falling within the *execution* quadrant account for a larger fraction of the time and energies of real-world teams. This seems particularly relevant in command and control environments (Kaufman, 2005). McGrath (1984) argued that a potential reason for the task type representation disparity is that success on tasks within the choosing quadrant may be evaluated solely on dimensions of team performance. Performance on execution tasks, on the other hand, is simultaneously dependent upon in-team performance and opposing-team performance, making the situation more difficult to examine and evaluate.

Additional research based on execution type tasks is necessary, however, to capture the dynamic, adversarial nature of many real-world circumstances in which teams may be required to engage. For example, urban warfare, encountered frequently by U.S. forces in Iraq (e.g., Batiste & Daniels, 2005), features at least two adversarial teams with similar, but opposing goals (eliminate enemy forces, preserve allied troops) which take place in a timeframe dictated by task goals (i.e., combat continues until one side is victorious). Such scenarios may require team members to dramatically and abruptly shift strategies, possibly several times, in order to meet changing goals and objectives. In addition, due to their adversarial nature, tasks which place teams in opposition to other teams feature some element of information decay, as competing members evolve new strategies to defeat their opponents.

Further, execution type task research is needed to evaluate the potential impact of collaboration technologies on team situational awareness and workload. Bordia (1997) concluded that communication restricted to text-based collaborative technologies impairs comprehension of communications from teammates and overall task objectives. This effect may be further compounded by the adversarial, time-sensitive nature of many execution tasks. In order to meet task goals, teams must collaborate rapidly and effectively to implement coordinated team strategies. Due to such time pressures, teammates may be required to communicate through the briefest messages possible, which may negatively impact team strategy and synchronization, leading to poorer situational awareness and increased workload. In addition, usage of collaboration tools may distract users as they focus on composing messages and monitoring responses

instead of engaging in and attending to overall task goals, which may further erode team situational awareness and increase both individual and team workload.

Due to the increasing drive to implement collaboration technologies into command and control environments, it is vital to evaluate the impact of these technologies on individual and team performance in dynamic, adversarial mission scenarios. The present experiment represents an initial, exploratory attempt to do so using the RoboFlag simulation environment. RoboFlag has been used to examine delegation-type interfaces as well as factors surrounding human supervisory control of multiple autonomous vehicles (Parasuraman, Galster, Squire, Furukawa, & Miller, 2005). RoboFlag requires operators to control semi-autonomous robot agents developed to exhibit various forms of self-directed, cooperative behaviors within flexible experimenter-specified scenarios. The simulations employed in this investigation were designed to resemble the game 'capture the flag.' Two teams of participants controlled multiple robots with the mission goal of sending team robots from a home area into the opposing team's territory to access and obtain a specified target (the flag) and return home as quickly as possible. The opposing natures of the two teams' goals (guard the team's flag, capture the rival team's flag) required teammates to collaborate on strategies in an environment in which strategy would be dynamically affected by both teams' efforts.

The goal of this study was to evaluate the potential utility of one collaborative tool – instant messaging (IM) – and to examine its effects on team performance. To do so, two factors were manipulated experimentally. The first was team control environment, which restricted team communication modes. The second factor was a manipulation of robot command abstraction. The experimental results concerning the abstraction factor are being reported elsewhere so that a comprehensive evaluation of the control environment factor can be presented here. For this experiment, based on the findings of Bordia (1997) and Baltes et al. (2002), it was tentatively hypothesized that communication restricted to IM would result in lower mission success rates, longer mission completion times, and less coordinated team strategies compared to communication that was unrestricted. It was also hypothesized that restricted communication would result in higher participant ratings of workload and lower ratings of situational awareness due to the potential negative impact of text-based collaboration technologies on both constructs. Finally, it was predicted that teams restricted to IM would send more instant messages than teams whose communication was unrestricted.

Methods

Participants

Twenty-eight men and eight women between the ages of 17 and 39 (M = 24.31 years, SE = 0.90) served as paid participants in this study. Participants completed the experiment in teams of four, yielding a total of nine groups. All participants were recruited from local universities or from available Air Force personnel. Participants were free to withdraw from the study at any time without penalty.

Experimental Design

A within-subjects design was employed, with two control environments (remote, co-located) combined factorially with three levels of abstraction (manual, automated, mixed) yielding six experimental conditions. All participants completed six mission trials in each condition, for a total of 36 trials. Control environment was a block factor (with 12 trials in a block), and level of abstraction was randomized within each block. Participants filled out the NASA Task Load Index (TLX; Hart & Staveland, 1988), which is a standard measure of workload that is widely used in human performance research (Wickens & Hollands, 2000). The sum of the ratings given to all of the scales provides an overall or global index of workload, while the ratings for each subscale provide an index of the contribution of the dimensions represented in that subscale to the workload of the task. Participants also completed one item from the 3-D Situational Awareness Rating Technique (SART; Taylor, 1990) designed to elicit overall impressions of situational awareness during the task. Participants completed both questionnaires following completion of each mission trial. Objective performance measures included mission success rate, mission completion times, chat communication frequency, and number of mouse inputs. Subjective measures included participant ratings of workload and situational awareness, as well as frequency and content of team communications.

Apparatus

The RoboFlag simulation employed in the current experiment was accomplished by means of five separate personal computers (PCs) running RoboFlag version 2.1 and communicating via TCP/IP protocol over a wired 10/100/100 Megabits per second connection. Each PC was equipped with a 3.2 gigahertz Hyper-Threaded Intel processor, a 17-inch monitor with a screen resolution of 1280 × 1024 pixels, a web camera, and a microphone. Of the five computers employed, four were assigned to participants (two to each team), while another PC comprised a central processing executive (the ‘Arbiter’) and collected data. Mission data were recorded by means of MORAE software, version 1.1.1, operating in the background of each PC. Using the web cameras and microphones, MORAE recorded each participant’s behavior during each trial while simultaneously logging keyboard and mouse inputs and recording the game display presented to participants on their PC monitors. However, due to a technical error, audio recording of the participants could not be recovered and the results offered will be devoid of this measure. The five PCs were linked through a wireless network operating at 54 megabits per second. The wireless network enabled participants to collaborate with their teammates using Windows Messenger version 5.0 and served to limit network collisions between the bandwidth intensive RoboFlag traffic and the less intensive IM traffic.

As illustrated in Figure 1, the scenario objective employed in this study was modeled on ‘capture the flag,’ with the field of engagement divided equally into two quadrants, one for each team. The goal of the scenario was for operators to send some or all of their robots across the dividing line into their opponents’ field, capture their ‘flag’ (indicated by a white dot located at the center of a circular area in their opponents’ quadrant), and

return the robot with the flag safely to ‘home base’ (located at the upper left corner or lower right corner of the operators’ quadrant), while simultaneously protecting their own flag (also indicated by a white dot located at the center of a circle in the operators’ quadrant). Friendly robots (controlled by the operator or their teammate) appeared in blue on the screen, while opponent robots appeared in red.

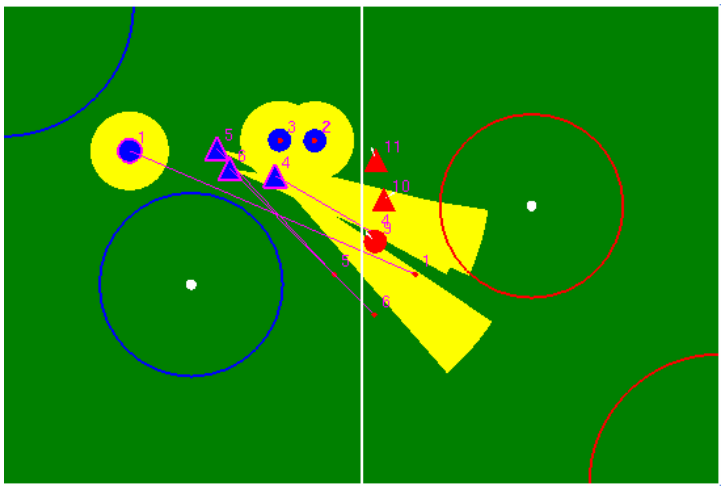

Figure 1. Screen shot from the RoboFlag simulated environment.

Within the simulation, two types of robots were employed, broadly corresponding to ground and air vehicles. The vehicles were differentiated by their simulated capabilities and on-screen shapes. Ground vehicles, which appeared on-screen as circles, were relatively slow moving, had a short visual range in a 360-degree visual arc, and were capable of capturing the opponents’ flag. Conversely, aerial vehicles, which appeared as triangles, moved at twice the speed of the ground vehicles, had a field of view that was 2.67 times longer than that of the ground vehicles in a conical 15-degree visual arc, and could not capture the flag. It should be noted that operators could not ‘see’ anything outside of the field of vision afforded by their team’s vehicles (e.g., operators could not see the movements of their opponents’ vehicles unless they were within sight range of a friendly vehicle).

In addition, as illustrated in Figure 2, each vehicle had a finite source of fuel that participants had to monitor using a graphic display located beneath the RoboFlag playing field. The monitor listed each robot and displayed its corresponding fuel level, both as a numerical value and as an analog bar which steadily decreased in size. As a further indicator of fuel consumption, the bar would change in color from blue, to yellow, to red. Robot fuel could be replenished by sending the vehicle back to home base. If a vehicle ran out of fuel, it would become inactive (essentially ‘dead’), in that it would cease functioning and the operator would be unable to command it until the beginning of the next mission trial.

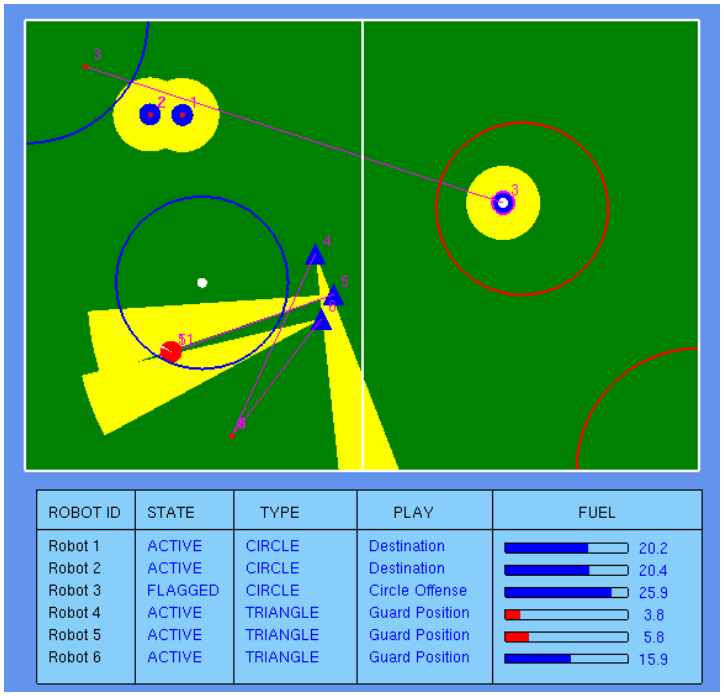

Figure 2. Participants had to monitor vehicle fuel levels, located in the bottom right of the display, in order to meet task goals.

Both teams in the RoboFlag simulation were equipped with six robots, three circles and three triangles. Each participant supervised three robots, all of the same type (i.e., one team member controlled all ground vehicles, and the other controlled all aerial vehicles). Based on the vehicles' simulated capabilities, ground vehicles were primarily responsible for offense, while aerial vehicles were reserved for defense and intelligence gathering. As each operator on a team was solely responsible for either ground or aerial forces, this created a natural division of labor between teammates that encouraged collaborative strategy development in order to meet mission goals. This division was not made explicit to participants leaving them free to develop and execute their own strategies.

While attempting to capture the opponents' flag, a vehicle could be 'tagged,' which resulted in that vehicle becoming disabled. Vehicles could become disabled by colliding with an opponent's vehicle while on their opponents' side of the field or on the operator's home side if the vehicle was carrying the flag. When tagged, a vehicle would return to home base by the shortest route possible, but their speed of movement was significantly reduced (ground and aerial vehicles had their movement rates reduced by factors of 2.5 and 5.0, respectively). A vehicle remained disabled and unavailable to respond to operator inputs until it had returned to home base.

Within the simulation, participants controlled their vehicles using combinations of manual point-and-click (waypoint) inputs and higher level automated 'commands,' which corresponded to complex aggregate behaviors (Parasuraman, Galster, & Miller, 2003). It should be noted that participants were only able to issue commands to vehicles for which

they were responsible, and that such commands had no direct effects upon vehicles controlled by their teammate or opponents.

Procedure

Participants completed the experiment in groups of four, with two participants assigned to each team. Once allocated to a team, participants were assigned to a control environment, either remote or co-located. In the remote condition, teammates were physically separated from each other by a partition. The partition prevented teammates from directly observing each other's display screens (though the display on both screens was identical) and participants were instructed to communicate with each other only through typed messages sent via Windows Messenger. In the co-located condition, teammates were stationed next to each other without a dividing partition, thereby enabling face-to-face collaboration. Teammates could observe each other's display screens directly and communication between them was unrestricted, allowing strategies to develop verbally or through Windows Messenger.

Following team assignments, all participants were given written and verbal instructions on the capabilities of each type of vehicle and on the automated commands available in the mission scenario. Participants were told that the experiment involved a game similar to 'capture the flag' and that they would each be controlling a set of three vehicles. Furthermore, they were instructed that each team would be in direct competition with the other, and that any single trial would continue until one team had successfully captured the other team's flag. Following the instructional period, participants were allowed to practice controlling the vehicles for five minutes. After this period had elapsed, participants were given additional practice time if they indicated that they were still uncomfortable controlling the vehicles. Prior to the start of each mission trial, participants were allotted 30 seconds for communication and strategy formation.

Participants were then assigned to control either the ground or aerial vehicles during the first half of the experiment, with the understanding that each participant would control the other type of vehicle during the second half. Next, participants were instructed on how to complete the NASA-TLX and the SART.

While individual trials varied in completion time, the entire experiment was generally concluded within a 3.5 hour time frame.

Results

Team Performance

During the 36 simulated trials, the RoboFlag software recorded which team successfully captured the flag (the winner) and the time elapsed during each mission trial. In addition, the RoboFlag software recorded the number of vehicle position changes initiated by each participant. In order to compare mission length and vehicle position changes, the winning

team's control environment and the value of each factor were recorded. Then, the values for the winning team on each trial for each factor were compared by control environment. Mean values for each factor can be seen in Table 1.

Table 1. Performance variable means as a function of control environment. Values in parentheses are standard errors.

Control Environment	Frequency of Wins	Length of Mission (sec)	Number of Vehicle Position Changes
Co-located	17.33 (1.41)	79.33 (9.26)	43.93 (7.27)
Remote	18.67 (1.41)	74.29 (5.04)	42.88 (6.95)

The data for each factor were tested for statistical significance by means of a 2 (control environment) × 3 (level of abstraction) repeated measures analysis of variance (ANOVA). However, as mentioned previously, only the main effects of control environment will be considered here and in all subsequently presented analyses. For the frequency of wins, mission length, and number of vehicle position changes no statistically significant differences were detected between the two conditions (F [1, 8] = 0.22, 0.49, 0.45 respectively, $p > .05$). Participants won as frequently in the remote condition as they did in the co-located condition, and the length of mission and number of position changes were similar regardless of the winning team's control environment.

One possible explanation for the observed results was that one team consistently won all mission trials, regardless of the manipulated experimental factors (i.e., teams were unevenly matched). In order to test this possibility, teams that began the experiment in the collocated condition were arbitrarily assigned to "team 1," and teams that began the experiment in the remote condition were assigned to "team 2." The number of mission trials each team won was then counted and compared by means of a two-sample *t*-test. The results of the *t*-test indicated that there was no statistically significant difference between the groups on number of wins, t (16) = 0.73, $p > .05$. Each team won approximately the same number of games (team 1 M = 18.89, SE = 1.84; team 2 M = 17.00, SE = 1.84).

The data were also examined to identify patterns of wins that were not due to the experimentally manipulated factors. For the purposes of these analyses a win 'streak' was defined as three or more serial wins by the same team. Using that criterion, and the previous "team 1, team 2" designations, a total of 38 win streaks were identified in the data. Streaks were found to occur across all experimental sessions. Overall, the mean number of win streaks per experimental session was 2.11 (SE = 0.32) and the mean number of trials in a streak was 4.05 (SE = 0.45). To determine if differences existed between teams in the frequency of streaks and the number of trials in each streak, teams were compared by means of two-sample *t*-tests. For these *t*-tests, and all subsequently reported *t*-tests, the Dunn-Sidak alpha correction was employed to control Type I error rates (Kirk, 1994). The results indicated that neither factor was significantly different between teams (t [16] = .68, .11 respectively, $p > .05$). As can be seen in Table 2, both teams displayed similar streak patterns.

Table 2. Mean streak frequency and mean number of trials in each streak as a function of team. Values in parentheses are standard errors.

Designation	Frequency of Streak	Number of Trials in Streak
Team 1	2.23 (.50)	4.10 (.60)
Team 2	1.89 (.42)	4.00 (.70)

Workload and Situational Awareness

To test the effects of the experimental conditions on participants' evaluation of task workload, the mean of the participants' ratings for the six TLX subscales on each trial were calculated to provide an overall or global index of workload. Mean global TLX and SART ratings were then calculated for each participant in each condition. These values were then tested for statistical significance by means of a 2 (control environment) × 3 (level of abstraction) repeated measures ANOVA. For workload and situational awareness, no statistically significant differences were detected between the two conditions (F [1, 35] = 0.30, 0.00 respectively, $p > .05$). Participants' mean workload ratings (co-located M = 36.02, SE = 2.11; remote M = 36.82, SE = 2.37) and mean situational awareness ratings (co-located M = 4.55, SE = .17; remote M = 4.56, SE = .15) were similar regardless of control environment.

Instant Messaging

From the instant messenger logs of communication between teammates, the total number of communications per experimental session was calculated. Messages were divided into three categories, depending on when they were sent: (1) pre-game messages were those sent during the 30-second strategy period immediately before the beginning of a mission trial; (2) in-game messages were those that were sent during the active portion of a mission trial; and (3) post-game messages were any that were sent after a mission trial was completed. It should be noted that participants were instructed *not* to send instant messages following the completion of a mission trial. Some teams, however, disregarded these instructions, leading to the inclusion of post-game messages in this analysis. The mean number of messages sent for each command environment in each messaging period can be seen in Figure 3.

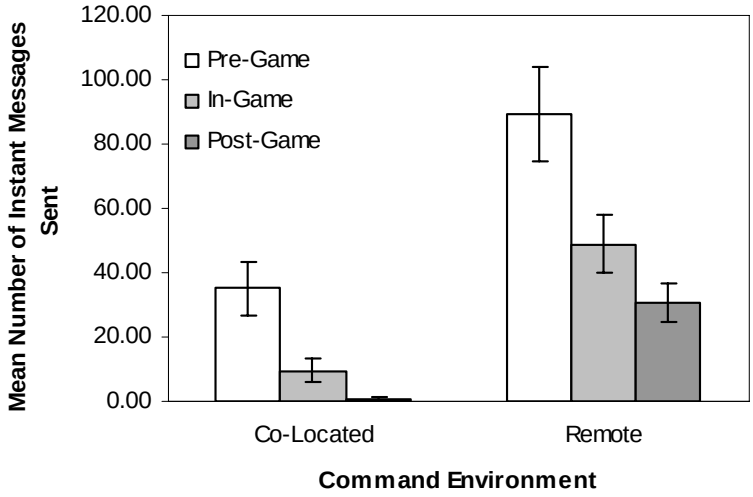

Figure 3. Mean number of instant messages sent as a function of command environment. Error bars represent standard errors.

In addition, the instant messages sent between teammates were analyzed to determine their content. Messages were coded as either 'irrelevant' (e.g., "I'm hungry," "I like this game") or 'strategy-relevant' (e.g., "go straight for their flag," "use more robots next time"). These categories were adapted from Bowers, Jentsch, Salas, and Braun (1998). Two coders separately classified each instant message into one of the two categories. Inter-coder reliability, assessed by Cohen's Kappa, was good (κ = 0.92). The mean number of irrelevant and strategy-relevant messages in each messaging period for each control environment can be seen in Figure 4.

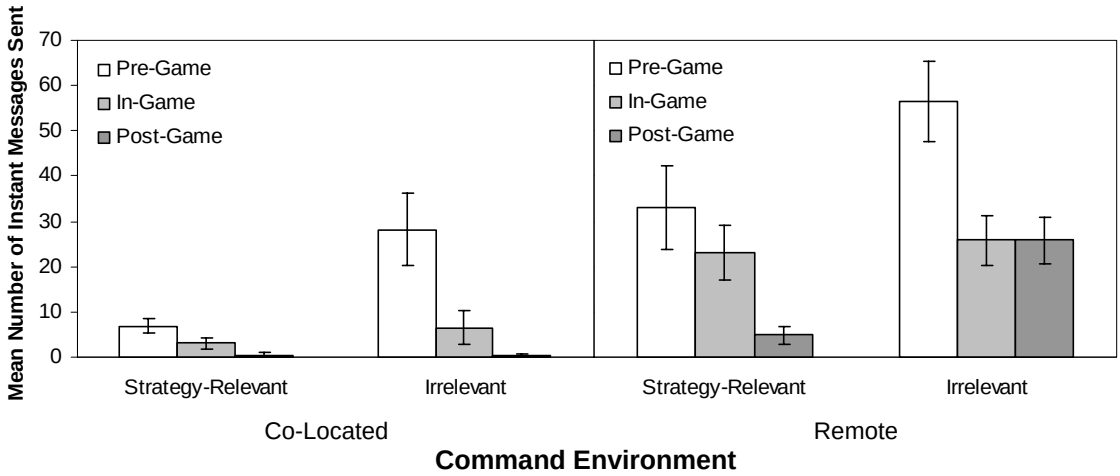

Figure 4. Mean number of strategy-relevant and irrelevant instant messages sent as a function of co-located (left panel) and remote (right panel) command environments. Error bars represent standard errors.

The mean number of strategy-relevant and irrelevant instant messages sent during each messaging period for the co-located and remote command environments were compared using a 2 (type of message) × 3 (messaging period) × 2 (command environment) repeated measures ANOVA. In this and all subsequent analyses involving repeated measures with more than two levels of the factor, Box's epsilon was employed to correct for violations of the sphericity assumption (Maxwell & Delaney, 2004). Statistically significant main effects were found for type of message (F [1, 8] = 8.81, $p < .05$), messaging period (F [1.69, 13.51] = 21.33, $p < .05$), and command environment (F [1, 8] = 27.17, $p < .05$), and for the interaction of message type and messaging period (F [1.70, 13.58] = 6.23, $p < .05$). All other sources of variance were not significant ($p > .05$). Post hoc t-tests of messaging period revealed that the difference in the mean number of messages sent was statistically significant comparing pre-game to in-game messages, t (16) = 2.98, $p < .05$, but not comparing in-game to post-game messages ($p > .05$). Overall, participants sent more irrelevant instant messages than strategy-relevant messages, sent more messages in the remote condition than the co-located condition, and sent the most messages during the pre-game strategy period.

To explore the interaction of message type and messaging period further, two repeated measures ANOVAs were calculated in which messaging period was analyzed separately for strategy-relevant and irrelevant messages. A statistically significant main effect of messaging period was found for both strategy-relevant messages, F (1.44, 11.53) = 14.71, $p < .05$, and irrelevant messages, F (1.77, 14.13) = 16.27, $p < .05$. For strategy-relevant messages, post hoc t-tests of messaging period revealed that the difference in the mean number of messages sent was not statistically significant comparing pre-game to in-game messages ($p > .05$), but was significant comparing in-game to post-game messages, t (16) = 3.32, $p < .05$. For irrelevant messages, post hoc t-tests of messaging period revealed that the difference in the mean number of messages sent was statistically significant comparing pre-game to in-game messages, t (16) = 3.12, $p < .05$, but not comparing in-game to post-game messages ($p > .05$). Participants sent more strategy-related messages during the pre- and in-game messaging periods and the most irrelevant messages during the pre-game period.

Discussion

The purpose of this experiment was to investigate the effects of communication restricted to instant messaging (IM) on team performance in a dynamic, adversarial task. Based on the research of Bordia (1997) and Baltes et al. (2002), it was hypothesized that restricted communication would impair team performance as evidenced by lower mission success rates, longer times to mission completion, and less coordinated team strategies compared to communication that was unrestricted. In addition, it was hypothesized that restricted communication would result in higher participant ratings of workload and lower ratings of situational awareness due to limitations in communication afforded by IM. However, initial predictions were not supported by the results of this experiment. No differences were detected between the co-located and remote conditions on frequency of wins,

mission length, or number of position changes. In addition, no differences were found in participant ratings of workload and situational awareness between the two conditions.

However, consistent with initial predictions, substantial differences were found between the two conditions in how frequently they communicated using IM. Participants sent more instant messages in the remote condition compared to co-located condition. Subsequent analysis of IM content indicated that participants in both conditions used IM overwhelmingly more frequently as a tool of socialization than for strategy building. Generally, participants sent the most strategy-relevant messages during the pre-game and in-game messaging periods, and the most irrelevant messages during the pre-game messaging period.

Overall, IM did not affect team performance. This is in stark contrast to the effects of collaboration technologies reported by Bordia (1997) and Baltes et al. (2002) who found that collaboration technologies exerted a negative influence on team performance. This difference may be explained by the difference in the task types, as specified by McGrath's (1984) circumplex task model. The reviews by Bordia (1997) and Baltes et al. (2002) were dominated by tasks from the *choosing* quadrant of the circumplex, which primarily involve team consensus building in tasks that may or may not have a correct answer. The RoboFlag simulation task used in the current experiment, on the other hand, falls within the *executing* quadrant of the circumplex because of the adversarial nature of each teams' goals (capture the flag and return it to home base before the other team does). It may be that the negative performance effects found in previous studies, which are largely associated with team discussion and consensus building, are mitigated by the time-critical need to anticipate and outperform the opposing team. The temporally-demanding, adversarial nature of the task may have forced participants to propose and accept strategies as rapidly as possible in order to meet task demands, thereby diminishing the negative effects of collaborative technologies previously reported by Bordia (1997) and Baltes et al. (2002).

In addition, participants received visual feedback during each trial as to the success or failure of their strategies as they tried to capture the flag. A string of successive failures using a particular strategy may have prompted participants to subtly shift strategies until they found a winning one. The development of a successful strategy by one team would then prompt the other team to change strategy to compensate. This dynamic interplay between team strategies would create a situation of continuous and rapidly evolving strategies as each team sought to stay ahead of the other. The winner-data analyses would seem to support this viewpoint. Across all experimental sessions, each team won approximately the same number of mission trials, indicating that teams were generally evenly matched in skill and strategy development. Win streaks were also generally short and infrequent, further supporting the idea of continuous forced strategy adaptation.

Strategy adaptation might also explain the relatively low ratio of strategy-related to irrelevant instant messages displayed by both teams. Due to the inherent uncertainty of the opposing team's strategy choices previous to a mission trial's commencement and the short time allotted for pre-game strategizing, rigidly scripted strategies would have been

difficult to communicate and implement regardless of command environment. Instead, participants may have devised weak or generalized strategies that could be substantially modified during play. By outlining strategies in more general terms, participants would have had the advantage of simultaneously communicating choices quickly and effectively, while still allowing for adaptation as the opposing team's strategy was encountered. In other words, the dynamic, adversarial task demands of the RoboFlag simulated environment, and perhaps other executing quadrant tasks, may favor the development and implementation of such strategies, which may in turn mitigate the negative impact of collaboration technologies on teams.

IM also did not seem to negatively impact participants' rating of workload and situational awareness. Participants did not express a difference in their perceived workload or situational awareness between the co-located and remote conditions. Again, this may be due to the dynamic, adversarial nature of executing quadrant tasks. The temporal demands of the task may favor short communications, both verbal and typed, and focused attention resulting in similar workload and situational awareness across command environments.

Additionally, the competition and game-related nature of the RoboFlag environment may motivate participation in ways that diminish the potential negative effects of collaboration technologies. It is possible that participants in the remote condition expend more effort and yet report no changes in workload or situational awareness because they are sufficiently engaged with the task. Support for this viewpoint comes from previous research linking task engagement to subjective energy and performance on difficult, attentionally demanding tasks (Matthews & Margetts, 1991; Matthews & Westerman, 1994). In other words, executing quadrant tasks may insulate participants from the negative effects of collaboration technologies on workload and situational awareness by motivation and engagement with the task.

Finally, as was initially predicted, participants sent more instant messages in the remote control environment than they did in the co-located condition, indicating that participants were using it for collaboration purposes. However, participants in both control environments largely used IM for socialization purposes, rather than using it exclusively for strategy development and coordination. This result may be cause for some concern. The dynamic nature of task goals in RoboFlag typically resulted in relatively short duration mission trials and active participation by all team members, which may have depressed the number of irrelevant messages participants were able to initiate. A task of longer duration, perhaps requiring less active involvement, may prompt participants to engage in off-task conversations more frequently, resulting in distraction, decreased situational awareness, and ultimately poor team performance.

Overall, the current experiment supports, in a limited fashion, the future successful integration of collaboration technologies into command and control environments. Team performance was essentially unchanged under both command environments, indicating that IM was at least as effective as face-to-face collaboration in this experiment. In addition, the results of this experiment underscore the need for continued research into

team performance and collaboration technologies in tasks from the executing quadrant of McGrath's (1984) circumplex model. Future research should focus on tracking strategy development and implementation in order to examine the factors that drive strategy evolution. Further, other factors that may mediate the use of collaborative tools, the types of tools used, and the consequences, either positive or negative, that stem from the utilization of these tools should be systematically manipulated and examined.

Acknowledgements

This work was supported, in part, by a grant from the Army Research Laboratory (GWSE047201202). Michael Barnes served as the Technical Monitor for this grant.

The authors would also like to thank Becky Brown and Judy Lee, from General Dynamics Advanced Information Systems, Dayton, OH, for their contributions to this research, and Robert Bolia, of the Air Force Research Laboratory, Wright-Patterson Air Force Base, OH, for his thorough and helpful comments.

References

Albert, D.S. & Hayes, R.E. (2003). *Power to the Edge*. Washington, DC: CCRP Publication Series.

Baltes, B.B., Dickson, M.W., Sherman, M.P., Bauer, C.C., & LaGanke, J.S. (2002). Computer-mediated communication and group decision making: A meta-analysis. *Organizational Behavior and Human Decision Processes, 87*, 156-179.

Batiste, J.R.S., & Daniels, P.R. (2005). The fight for Samarra: Full-spectrum operations in modern warfare. *Military Review, 85*, 13-21.

Boiney, L. (2005). Team decision making in time-sensitive environments. *Proceedings of the 10th International Command and Control Research and Technology Symposium: The Future of C2*. (#175). McLean, VA.

Bordia, P. (1997). Face-to-face versus computer-mediated communication: A synthesis of the experimental literature. *The Journal of Business Communication, 34*, 99-120.

Bowers, C.A., Jentsch, F., Salas, E., & Braun, C.C. (1998). Studying communication patterns among aircrews: Implications for team training needs assessment. *Human Factors, 40*, 672-679.

Hart, S.G., & Staveland, L.E. (1988). Development of a multidimensional workload scale: Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). Amsterdam: North-Holland.

Kaufman, A. (2005). Caught in the network: How the doctrine of network-centric warfare allows doctrine to dictate military strategy. *Armed Forces Journal*, February, 20-22.

Kirk, R.E. (1995). *Experimental design: Procedures for the behavioral sciences (3rd ed.)*. Pacific Grove, CA: Brooks/Cole.

Matthews, G., & Margetts, I. (1991). Self-report arousal and divided attention: A study of performance operating characteristics. *Human Performance, 4*, 107-125.

Matthews, G., & Westerman, S.J. (1994). Energy and tension as predictors of controlled visual and memory search. *Personality and Individual Differences, 17*, 617-626.

Maxwell, S.E., & Delaney, H.D. (2004). *Designing experiments and analyzing data (2nd ed.)*. New Jersey: Erlbaum.

McGrath, J.E. (1984). *Groups: Interaction and performance*. Englewood Cliffs, NJ: Prentice-Hall.

Parasuraman, R., Galster, S., & Miller, C. (2003). Human control of multiple robots in the RoboFlag simulation environment. *Proceedings of the Annual IEEE Conference, 33*, 3232-3237.

Parasuraman, R., Galster, S., Squire, P., Furukawa, H., & Miller, C. (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: Systems and Humans, Vol. 35, No. 4, July 2005.* (pp. 481-493).

Taylor, R.M. (1990). Situational awareness rating technique (SART): The development of a tool for aircrew systems design. In *Situational Awareness in Aerospace Operations* (AGARD-CP-478) (pp. 3/1 - 3/17). Neuilly Sur Seine, France: NATO - AGARD.