Support of Teamwork in Human-Agent Teams

Terri Lenox, Michael Lewis, Emilie Roth, Rande Shern, Linda Roberts,
Tom Rafalski, and Jeff Jacobson
Department of Information Sciences and Telecommunications
University of Pittsburgh
135 N. Bellefield Avenue
Pittsburgh, PA 15260

ABSTRACT

Our research program is investigating how to effectively incorporate intelligent agents into human teams, specifically to support team performance. There are several ways to incorporate agents into teams; our approach is to use agents to support the team as a whole (facilitating communication, allocation of tasks, coordination among the human agents, and improving attention focus). In our initial experiments, we investigated the effects of trust and calibration of trust that allows human decision-makers to assess the reliability and meaning of communications from software and human agents. Currently, we are focusing on the team and the different ways of deploying agents to support multi-person teams: 1) supporting the individual by tracking the collected information; 2) supporting communication among team members by automatically passing information to the relevant person; and 3) supporting task prioritization and coordination by providing a shared checklist. Only performance data have been reported for this paper. Aiding with agents was found to be different from the control condition. Learning effects and effects for the target difficulty were found.

1. INTRODUCTION

Due to the complex and dynamic nature of our world it is increasingly common for decisions to be made by teams rather than individuals. Much of the early focus on decision aids has been on supporting the individual [1]. Under a Multi-University Research Initiative (MURI) project we are engaged in determining how to effectively incorporate intelligent agents into human teams for complex team tasks such as joint operations planning. Our research program focuses on using intelligent agents to support team performance.

We have developed a framework for examining the different ways that machine agents can be deployed in support of team performance. One option is to support the individual team members in completion of their own tasks. Another option is to allocate to the machine agent its own subtask as if we were introducing another member into the team. In this case all the issues associated with communication and coordination among team members become relevant [1], [2], [3].

The third option is to support the team as a whole (facilitating communication, allocation of tasks, coordination among the human agents, and attention focus.) The focus here is on how to support interactions among team members using machine agents [4]. Specifically the focus is on how machine agents can be used to support and promote teamwork dimensions [3]. Jentsch et al. have identified important dimensions of team work to include: 1) Team situation awareness (exploit all available information sources; disseminate information; provide situation updates); 2) Supporting behavior (prompt correction of team errors; provide and request backup when necessary); 3) Communication (proper terminology; complete internal and external reports; brevity and clarity); and 4) Team initiative/leadership (provide feedback to team members; state clear and appropriate priorities).

A basic tenet of this model is that teamwork skills exist independent of individual competencies. The performance of teams, especially in tightly coupled tasks, is believed to be highly dependent on these interpersonal skills. We are currently engaged in a research program that systematically explores these alternative ways of using machine agents to support individuals and teams (Table 1).

The initial set of studies were conducted using a low fidelity simulation, TANDEM, that was jointly

Teamwork Dimension	*Ways to Provide Support*
Situation assessment	❑ act as an information provider ❑ ensure the exchange of relevant knowledge among all members ❑ provide a shared representation of the situation for common knowledge ❑ notify people when all available information sources are not being used ❑ notify people when information is currently disable or unreliable
Supporting behaviors	❑ make other's actions and decisions visible to detect potential errors ❑ alert members to potential errors, constraints and conflicts ❑ support self-reflection of the team for self-correcting teams
Communication	❑ alert members to ambiguity ❑ support translation of terminology among various subgroups ❑ serve as a repository for messages between team members
Team leadership/initiative	❑ communicate intent behind commands to allow lower levels of command to meet intent under evolving conditions ❑ communicate priorities and notify members when priorities change

Table 1: Using Agents to Support Teamwork

developed at the Naval Air Warfare Center - - Training Systems Division and the University of Central Florida. In TANDEM, subjects must identify targets and choose an appropriate response to a series of air, surface or submarine targets. Subjects perform a sequence of time critical information gathering and communications tasks in order to decide whether to shoot or clear each target. The data necessary to make a correct decision is distributed among three team members. Subjects must communicate with one another to amass the information needed by the coordinating node to make the final decision.

2. THE ROLE OF TRUST IN HUMAN-AGENT INTERACTION

In the standard TANDEM paradigm (Figure 1), three member teams perform the target identification task. In this first experiment we substituted machine agents for the two non-coordinating team members to investigate the effects of agent communication protocols on human decision making [6]. This is an example of the second way to assist teams with agents by introducing machine agents as part of the team. Trust was treated as an intervening variable enabling the human decision-maker to assess the reliability and meaning of communications from software agents. Following Muir [5], trust is considered to be multidimensional and varying in character from verifiable consistency to blind faith and teleology. We hypothesize that effective human-agent performance requires a precise calibration of trust so that the decision-maker can accurately interpret an agent’s communications. This calibration depends on experience with an agent and “explanations” the agent may provide to support its messages.

Three different agent levels were tested, each type roughly paralleling the level of trust required for interpretation. The first agent aggregated information into lists of parameters and values. The second agent provided inferential information by classifying values in the form of a table. The third agent provided decision information in the form of a classification message (e.g., target 45 has a .81 probability of being a submarine). To manipulate subjects’ trust of the agents, presentation errors were introduced in all three levels. These errors included data and classification errors and bad decision rules.

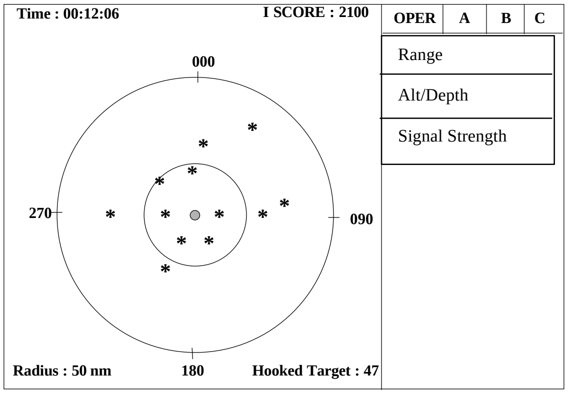

Figure 1: Standard TANDEM environment

There were over 70 subjects in this study. Overall, the second agent (table) seems to provide the best support for the target identification task. Although subjects consulted the level 3 agent more than either of the other two agents, their scores were lower than subjects using the other agents. Regardless of their source, errors affected subjects' performance, reliance on agents and ratings of trust [6].

3. SUPPORTING TEAM WORK DIMENSIONS

This second study focuses on the team. In this study we are looking at different ways of deploying machine agents to support multi-person teams: 1) supporting the individual (within a team context) by keeping track of the information he has collected and in sense, helping the individual with his task and with passing information to team mates (Individual Clipboard); 2) supporting communication among team members by automatically passing information to the relevant person which should reduce communication errors and facilitate individual classification (Team Clipboard); and 3) supporting task prioritization and coordination by providing a shared checklist (Team Checklist).

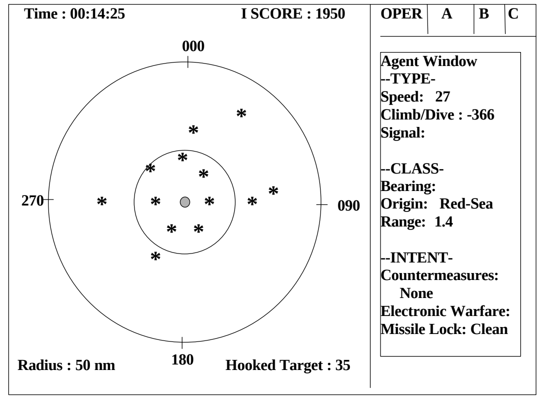

Figure 2: Individual Agent

We hypothesized that the Individual Agent should aid the individual task and aid communication among team members (Figure 2). This agent shows all data items available to an individual team member (in this case, ALPHA) and fills in the values for the data items as the subject selects them from the menu. The values under the TYPE heading assist the individual with their task while the other team members may need the remaining values.

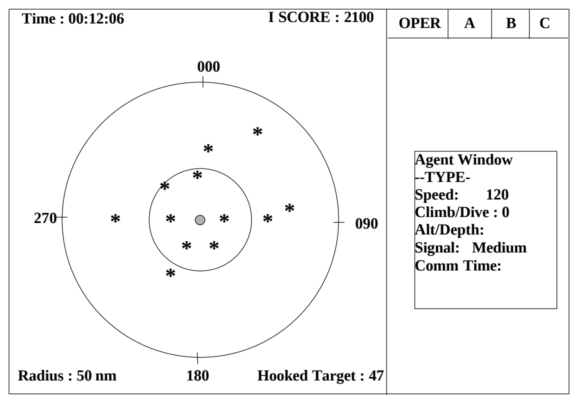

Figure 3: Team Clipboard Agent

The Team Clipboard Agent should also aid the individual task and aid team communication to a greater degree than the Individual Agent (Figure 3) should. This agent aggregates values from all members of the team to help the individual with his/her task. It automatically passes values as they are selected from the menu to the appropriate team member. Thus, when altitude/depth is selected from a menu, it is passed to an individual team member (ALPHA) who can use it to make the type identification. We hypothesized that this agent should reduce verbal communication among team members and reduce communication errors.

The third agent, Team Checklist, should aid team coordination (Figure 4). This agent shows who has access to what data. For example, all three team members (ALPHA, BRAVO, CHARLIE) have access to speed, but only BRAVO has access to

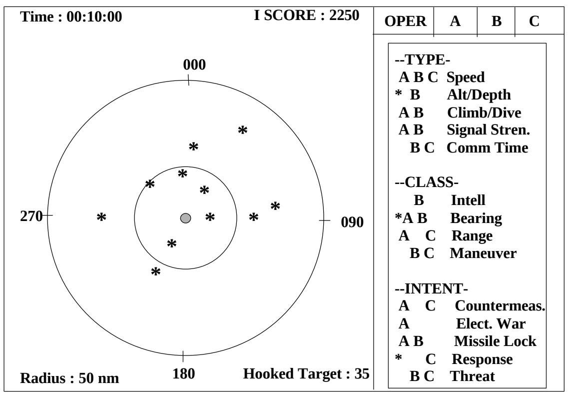

Figure 4: Team Checklist Agent

altitude/depth. The asterisk (*) indicates that the data item has been accessed from the menus.

The final condition is a control where we observed team performance without the aid of any machine agent. This is the standard TANDEM paradigm used by Jentsch, et al [3].

The goal of the study is to examine the impact of the alternative agents on: 1) communication patterns, 2) data gathering strategies, 3) reliance (i.e., use of) on the intelligent agents, and 4) performance. We are collecting both process measures (e.g., communication and data gathering strategies) as well as outcome measures of performance (e.g., number of targets processed and number of errors).

4. METHOD

Teams of three subjects were recruited for this study. Each team was assigned to one of four conditions: 1) control, 2) individual agent, 3) team clipboard agent, or 4) team checklist agent.

TANDEM was used with three-person teams, each member with a different identification task to perform (air/surface/submarine, military/civilian, and peaceful/hostile). One person was assigned to ALPHA, one to BRAVO and one to CHARLIE. ALPHA, BRAVO and CHARLIE had different items on their menus and different tasks during the trials. ALPHA identified the type of target (air, surface or submarine); BRAVO determined whether the target was civilian or military; CHARLIE determined whether the target was peaceful or hostile. In addition, CHARLIE acted as the leader by indicating the type, classification and intent of each target to the system and taking the final action (shoot or clear).

There were five pieces of information for each identification task, three of which must agree in order to make a positive identification. These pieces of information were distributed among the three team members. Each team member saw different data items on the menus and had three data items required for his/her identification task and several other items that the other team members might need to complete their tasks. Thus, the subjects needed to communicate with one another to perform their tasks for roughly two-thirds of the targets. All five pieces of information might agree for a particular target, however, in many cases, the ambiguity of the data was manipulated such that only three pieces agreed.

The targets were divided into three groups: 1) easy -- all three pertinent items on the individual's menu agree; 2) medium -- only two items on the menu agree, a team member must ask one or both teammates for data; and 3) hard -- two items on the menu agree, but do not provide the correct solution. For example, ALPHA's task was to identify the type of target. If the target was easy, all three items on ALPHA's menu indicated the same type (e.g., air). If the target was of medium difficulty, one or two values would indicate air and the other indicate submarine. If the target was hard, both of ALPHA's menu items indicate air, but the remaining three items from ALPHA's menu and the other team members indicate surface. Thus, the target is a surface vessel. Subjects had no way of knowing the difficulty level of the targets.

Each team participated in a 90-minute sessions which began with a 15-minute training session in which the TANDEM software and team goals were explained. The team was told to identify as many targets as possible, as accurately as possible during the 15-minute trial. After the training session, the team participated in three 15-minute trials. At the conclusion, subjects were asked to complete a brief questionnaire.

Several forms of data were collected during the trials: 1) *performance data* from Tandem logs including the type and number of targets hooked and classified, the percentage of targets correctly identified, and the number of times the agents were activated; 2) *communication data* encoded from observers or audio tapes including the number of requests for data (e.g., does anyone have initial range?), the number of responses (e.g., range is 5.6 nm), the number of target identifications (e.g., it's civilian), and the number of confirmations (e.g., target is sub, civilian); 3) *observer data* including ratings on team communication, situation assessment, leadership ad supporting behaviors; and 4) *questionnaires c*ompleted by the subjects before they leave. Only performance data have been reported for this paper. Measures included: proportion of targets correct, time per target, and total targets hooked.

5. RESULTS

The performance data reported in this paper are based on five teams per condition. One team from each condition was dropped from this analysis due to performance problems. That is, these teams had less than 64% accuracy in target identification.

Subjects hooked equal numbers of easy, medium and hard targets during each session.

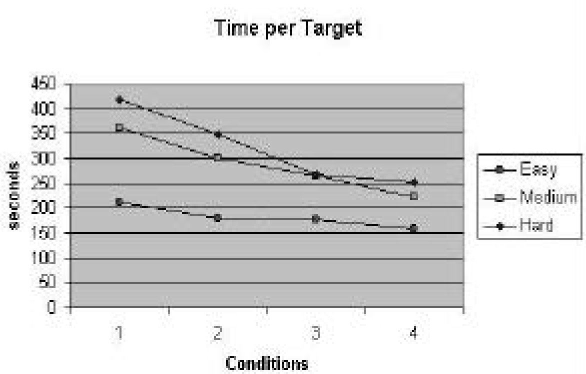

Figure 5: Time per Target across Target Difficulties

Figure 5 shows how the time per target varies for both the target difficulty and across conditions. For example, teams took approximately 450 seconds per target to process hard targets in the control condition, 350 seconds in the individual agent condition, 250 seconds in the team clipboard condition, and 150 seconds in the team checklist condition.

	Within Subjects Effects	
Measure	**Trial**	**Difficulty**
proportion correct	p<.009	p<.0001
time per target	p<.0001	p<.0001
total targets hooked	p<.0001	-

Table 2: Within Subjects Effects for 4 Conditions

Using a repeated measures design with four conditions and N = 5 (Table 2), there were significant within subject effects over the three trials for the proportion of correct targets ($p<.009$), the time per target ($p<.0001$), and the total targets hooked by a team ($p<.0001$). Significant within subjects effects were also found across the target difficulties (easy, medium, and hard) for the proportion of correct targets ($p<.0001$) and the time per target ($p<.0001$).

For time per target, effects were also found for three interactions: difficulty by condition ($p<.048$), difficulty by trial ($p<.0001$) and difficulty by trial by condition ($p<.030$). Looking at the proportion of correct targets, significance was also found for three interactions: difficulty by condition ($p=.048$), difficulty by trial ($p<.0001$), and difficulty by trial by condition ($p<.028$).

In pairwise comparisons for time per target, the control condition differed significantly from the team clipboard agent ($p<.036$) and the control condition differed significantly from the team checklist agent ($p<.021$). Target difficulties showed effects for the time per target measure also; easy versus medium difficulty ($p<.0001$), easy versus hard difficulty ($p<.0001$), and medium versus hard ($<=.004$). Finally, all pairs across trials were significant ($p<.0001$).

In pairwise comparisons for the proportion of correct targets, target difficulties showed effects for easy versus medium difficulty (p<.010), easy versus hard (p<.000) and medium versus hard (p<.022). The first trial differed from both the second trial (p<.015) and the third trial (p<.008).

Grouping all agent conditions (individual agent, team clipboard agent and team checklist agent) into one condition (N = 15, conditions = 2), showed that agent aiding is different from the no aiding condition (control) over the three trials on the proportion of correct targets (p<.0001), time per target (p<.0001) and total targets hooked (p<.0001). (See Table 3.)

	Within Subjects Effects	
Measure	**Trial**	**Difficulty**
proportion correct	p<.0001	p<.0001
time per target	p<.0001	p<.0001
total targets hooked	p<.0001	-

Table 3: Within Subjects Effects for All Agents and Control Condition

Significant within subjects effects were also found across the target difficulties (easy, medium, and hard) for the proportion of correct targets (p<.0001) and the time per target (p<.0001).

Significant within subject effects were found for the proportion of correct targets for the interaction of difficulty and trial (p<.019). Also, for the time per target, significant effects were found for the interaction of difficulty and trail (p<.0001).

In pairwise comparisons for the proportion of correct targets showed effects for target difficulties and trials. Specifically, easy targets differed from medium targets (p<.012); easy targets differed from hard (p<.0001); and medium targets differed from hard (p<.018). Across trials, the three combinations of trials differed : trial 1 from trial 2 (p<.051), trial 1 from trial 3 (p<.008), and trial 2 from trial 3 (p<.039).

In pairwise comparisons for time per target showed significant effects for all agents versus the control condition (p<.025), for all three possible target difficulty crosses, and for all three trial crosses.

6. DISCUSSION

Subjects learned across the trials, hooking more targets, spending less time on any particular target and getting more targets correct. And in general, the difficulty levels built into the targets held true. That is, easy difficulty targets took less time to process and teams got more of them correct than they did with medium difficulty targets than they did with hard targets.

Teams using the clipboard agent spent less time per target than teams in the control condition.

In general, teams using agents differed from teams in the control condition (no agents) on the performance data. However, it is not clear from this performance data exactly how the agents aided the teams and to what degree this aiding helped performance. Individual differences may have obscured the differences among the various agents. In addition, the question remains whether or not the other data points collected will shed any light on the differences among the aiding conditions. Specifically, we have collected communication data encoded from observers (number of requests for data, responses, target identifications, and confirmations, observer data (ratings on team communication, situation assessment, leadership ad supporting behaviors), and exit questionnaires.

Further work needs to be done to understand the degree to which the various agent-aiding strategies differ and on which measures. To accomplish this, an additional study will be undertaken.

ACKNOWLEDGMENTS

This research was supported by an Office of Naval Research grant N-00014-96-1-1222.

7. REFERENCES

[1] E. M. Roth, J. T. Malin, J. T & D. L. Schreckenghost (1997). Paradigms for Intelligent Interface Design. In M. Helander, T. K. Landauer, and P. Prabhu (Eds) Handbook of Human-Computer Interaction, Second Edition, pp. 1177 -1201.

[2] P. M. Jones & C. M. Mitchell (1995). Human-computer cooperative problem solving: Theory,

design, and evaluation of an intelligent associate system. IEEE Transactions on Systems, Man, and Cybernetics. SMC-25(7), 1039-1053.

[3] K. Smith-Jentsch, J. H. Johnston, S. C. Payne (in press). Measuring team-related expertise in complex environments. To appear in J. A. Cannon-Bowers and E. Salas (eds.), Decision Making Under Stress: Implications for Individual and Team Training. Wash., DC: American Psychological Association.

[4] J.T. Malin, D.L. Schreckenghost, D.D. Woods, S.S. Potter, L. Johannesen, M. Holloway, & K. D. Forbus (1991). Making Intelligent Systems Team Players: Case Studies & Design Issues. Human-Computer Interaction Design (NASA Technical Memorandum 104738). Houston, TX: NASA Johnson Space Center.

[5] B. M. Muir. (1994). Trust in Automation, Part I. Theoretical issues in the study of trust and human intervention in automated systems, Ergonomics, 37(11), 1905-1922.

[6] T. L. Lenox, L. L. Roberts, and M. Lewis (1997). Human-Agent Interaction in a Target Identification Task, 1997 IEEE International Conference on Systems, Man, and Cybernetics. 2702-2706