**Abstract**

This paper explores the key characteristics of an intelligent advisory system. A central feature is that human-machine cooperation should be based on a metaphor of human-to-human cooperation. ALLY, a computer-based operator's associate is discussed which is based on a preliminary theory of human-to-human cooperation. ALLY assists the operator in carrying out supervisory control functions for a simulated NASA ground control system. Experimental evaluation of ALLY indicates that operators using ALLY performed as well as they did when using a human associate, and in some cases they performed even better.

**INTRODUCTION**

Command and control (C2) systems have undergone dramatic changes within the last twenty years. Operators are faced with monitoring and controlling large, complex systems which rely heavily on the use of automation. Often, the system is too large and complex for a single operator to monitor.

This paper presents the results of a research effort to explore the issues associated with human-machine cooperation in complex, dynamic supervisory control situations and to develop a theory of human-machine cooperation which can be used to design the architecture for a computer-based operator's associate. The research focused on the development of a computer-based associate that is capable of cooperating with a human operator in monitoring and controlling a complex, dynamic system.

**OPERATOR'S ASSOCIATE**

As systems become more automated, the human operator performs fewer tasks on a routine basis. In complex dynamic systems, however, safety requires staffing at a level that can meet the most challenging or threatening abnormal conditions (Wickens, 1984). Normally, worst-case conditions are well beyond the normal, day-to-day operational conditions. The result is often a team of human operators who are rarely challenged and often underutilized.

The concept of a computer-based operator's associate has been proposed as one method to remedy this situation and to provide intelligent decision aid for operators of complex dynamic systems (Chambers & Nagel, 1985; Rouse, Geddes, & Curry, 1987; Rubin, Jones, & Mitchell, 1988). A computer-based operator's associate is a computer-based system that acts as an assistant to the human operator. Functionally, an operator's associate can offer the operator timely advice and reminders, and at the operator's request, assume responsibility for portions of the supervisory control task.

The subordinate role of the operator's associate is a fundamental assumption that characterizes this research effort. The rationale for this assumption is that in complex dynamic systems it is impossible to anticipate and plan for all the contingencies. Thus, a computer system cannot act as the principal or sole "expert" in the system control; a human decision maker will always be present and ultimately responsible for effective and safe system operation. Thus, it is essential to design the system so that the operator is an integral part of the control and decision processes.

The intelligence and utility of the operator's associate rests on its ability to understand the operator's current intentions and provide context-sensitive assistance in the form of operator aids (e.g., suggestions, advice, and reminders) or by assuming responsibility for portions of the control task. To ensure generalizability, the operator's associate requires a well-defined knowledge structure. Knowledge concerning the operator's functions, and operator intentions must be represented (Chambers & Nagel, 1985; Rouse, et al., 1987; Rubin et al., 1987; Carroll & McKendree, 1987; Geddes, 1989; Hollan & Coombs, 1983).

The understanding properties of the computer-based associate are based upon the existence of a model that prescribes the operator's interaction with the system (Rouse, et al., 1987; Rubin et al., 1988; Geddes, 1989). Based on this model of the operator's actions, the automated associate must be able to monitor the operator's actions and model the current status of the decision maker.

*Bushman is now with the Training Systems Division, Air Force Human Resources Laboratory, Brooks Air Force Base, Texas, 78235.*

*Rubin is now with ParcPlace Systems, 1550 Plymouth St, Mountain View, California, 94043.*
PRINCIPLES OF COOPERATION

The final property of a computer-based associate is that it should be based on the metaphor of human-to-human cooperation. The computer-based associate should interact with the human operator in a manner similar to the way in which humans interact in a cooperative environment (Carroll & McKendree, 1987; Hollan et al., 1986; Fischhoff, 1986; Roth, Bennett & Woods, 1987; Woods, 1986a, 1986b; Woods, Roth & Bennett, 1987). An extensive empirical study was undertaken to investigate the nature of human-to-human cooperation that could serve as the basis for the architecture of an operator's associate.

The general principles of cooperation were derived from two sources. First, an extensive review of the literature was undertaken on cooperative problem solving. Second, extensive data was collected observing a team of experienced operators in the GT-NSOCC system (a typical cooperative supervisory control system) (Mitchell, 1987). The two operators were free to develop a "natural" style of interaction and cooperation. Verbal protocols were collected of the interactions between the operators and data describing their performance were collected. The verbal records and data were then analyzed to describe the nature of their cooperative behavior.

A review of the literature indicated that a key principle of cooperation is that operators use multiple mental models to represent their knowledge of the physical system and their functions and to represent their knowledge of the other members of the cooperative team (Athans, 1982; Rasmussen, 1984, 1985; Tenney & Sandell, 1981a, 1981b). These distinct models serve to define and guide the interaction with the system and their interaction among the other operators.

The second feature of cooperation is referred to as cognitive balancing. This term is coined from the cognitive engineering approach to designing human-machine systems (Woods, 1986a, 1986b). Woods argues that the demands of the human and the system need to be considered and supported during the design of a human-machine system. With respect to a cooperative environment, the interacting operators must be aware of the cognitive demands and limitations of the other operators in order for efficient coordination and interaction to occur. One of the objectives of a cooperative team of problem solvers is to attempt to balance the joint cognitive demands of the team, as a whole. This balance is achieved through a mix of communication and delegation.

The final characteristic of cooperation is flexible levels of interaction. Empirical evidence supports the use of Rasmussen's levels of abstraction and aggregation (Rasmussen, 1984, 1985, 1986) to describe the content of the various mental models maintained by the operators and to describe the degree of interaction among the operators. The appropriate level of interaction is dynamic and is determined by the need for cooperation strategy. Interaction among the operators occurs at the levels of abstraction and aggregation common to the operators.

ALLY: A COMPUTER-BASED ASSOCIATE

These properties of a computer based associate and the principles of cooperation form the basis for the development of an architecture for a computer-based associate. The architecture is based on the OFMspert architecture (Rubin et al., 1988). The architecture incorporates multiple models that represent the system knowledge, procedural knowledge, and operator intentions. The OFMspert architecture uses the operator function modeling (OFM) methodology as the basis for the design of an operator's associate. A key component of an operator's associate is the intent inferencing capability which provides the understanding properties for an intelligent operator's associate. The intent inferencing capability uses a blackboard architecture to understand the operator's current goals. The OFMspert intent inferencing capability was validated in Jones et al. (1989).

ALLY, a computer based associate, is based on an extension of the OFMspert architecture with control capabilities. The architecture provides an interface to the operator that allows the operator to retain complete control over the computer-based associate. The operator can delegate to the associate as many or a few of the tasks as desired.

ALLY was developed to assist an operator in carrying out the supervisory control function for a simulated NASA ground control system, called the Georgia Tech Multisatellite Operations Control Center (GT-MSOCC) (Mitchell 1987; Saisi, 1986). The design was based on a model of the GT-MSOCC operator control functions and attempts to duplicate the capabilities of a human associate. A detailed description of ALLY can be found in Bushman (1989).

The operational concept behind ALLY's design is that ALLY is based on observations of the relationship that developed between a human operator and a human associate controlling the GT-MSOCC system. The human operator was in complete control of the human associate. The human associate, however, understood the cognitive complexities of the operator functions actively monitored the system for failures, and when necessary, would troubleshoot the system. ALLY functions in a manner similar to the human associate. The operator has delegate as few or as many of the tasks to ALLY as desired. ALLY also actively monitors and troubleshoots the system on its own. ALLY was developed in Smalltalk-80 on a Macintosh II. ALLY interacts with the GT-MSOCC system in a distributed fashion. ALLY acts like another operator of GT-MSOCC system in a distributed fashion. ALLY acts like another operator of GT-MSOCC (see Figure 1). A distributed architecture is consistent with the concept of an assistant that executes autonomously and in its own environment.

Figure 2 provides an example of the ALLY interface to the operator. ALLY performs both delegated and automatic control tasks. The TMSmalltalk-80 is a trademark of ParcPlace Systems, Inc.
operator delegates tasks to ALLY by selecting the corresponding control button. Each control button represents a specific operator control function as described in the GT-MSOCC operator function model (OFM) (Mitchell, 1987). Associated with each control button is a series of tasks that the human operator can delegate to ALLY.

In addition to the delegated tasks, ALLY performs two tasks automatically. ALLY continuously monitors and troubleshoots the equipment networks. ALLY also automatically monitors critical events and offers reminders when it appears the the events might have been missed. This behavior is similar to that observed in a human associate working with the operator to control the GT-MSOCC system.

AN EXPERIMENT

An experiment was conducted to evaluate the effectiveness of ALLY as an operator’s associate. The experiment compared the performance of an operator controlling GT-MSOCC working with ALLY as an associate with the performance of an operator working with a human associate.

Experimental Setup

The baseline GT-MSOCC system is a single operator system. In order to conduct the experiment, GT-MSOCC was modified to accommodate two operators. One operator serves as the primary operator and the second operator serves as an associate.

To support the associate position, two additional display screens were added to the baseline configuration. These two screens are functionally equivalent to the left and right screen in the baseline configuration. The center screen showing the GT-MSOCC configuration and status page is shared by the operator and associate. Although the physical display terminals are arranged in a different order, the functionality of the screens remain the same.

Each position is capable of issuing any of the GT-MSOCC operator control and information request command. Each position also has a dedicated audible alarm for system alarms. Common alarms indicating system events are sent to both positions, while operator error messages are only sent to the position which originated the error.

Subjects

Ten paid volunteer undergraduate Air Force ROTC cadets from Georgia Institute of Technology participated as subjects for the experiment. The subjects consisted of one female and nine males. The subjects included one junior, one sophomore, and eight freshman cadets. The subjects were paid six dollars per session.

Experimental Materials

Four sets of written instructions were used in the experiment. The first set consisted of an introduction to the baseline GT-MSOCC system and the operator supervisory control functions. These baseline instructions are found in Salis (1986).

The second set of instructions briefly described the operator-associate operations concept. The third set described the human associate concept and the modified GT-MSOCC workstation for a team of operators. Finally, the last set of instructions described the capabilities of ALLY and the user interface.

Several questionnaires were used during the experiment to collect subjective data. At the end of each data collection session, the subjects were asked to complete a Cooperation Evaluation...
carrying out the GT-MSOCC supervisory control functions. In addition, the subjects were asked to complete an ALLY Exit Questionnaire and a Human Exit Questionnaire at the end of their last data session with respective associate. The purpose of this these questionnaires was to elicit their opinions about various aspects of the associates. Finally, at the end of the experiment, the subjects were asked to complete a Subjective Comparison Rating questionnaire to compare their opinions about the two associates subjectively.

**Overview of Experimental Sessions**

The subjects were divided into two groups of five subjects each to control the order in which the subjects received the different associates. One group worked with the human associate first and the other group worked with ALLY first. In addition, to control for the variability of a human associate, a confederate was used in the experiment. The confederate was an expert GT-MSOCC operator and served as the human associate for each subject. The expert was instructed to use the same strategy for carrying out the operator control functions consistently to control the bias that might enter into the experiment from repeatedly seeing the same experimental sessions.

The subjects participated in twenty-four sessions: eight sessions of baseline GT-MSOCC training, three sessions of human associate training, four sessions of ALLY training, and five sessions of ALLY data collection. A total of 240 hours of data was collected. The sessions were approximately 65 minutes in length. The sessions were run on consecutive days with typically one session per day. Occasionally, the subjects missed a day and made up the session by running multiple sessions in a single day.

Within each session, three hardware failures and six software failures were scheduled to occur. The failures were scheduled to occur at set times (as determined by the seed of a random number generator) on identical equipment across subjects for a given session. However, since the subjects did not operate the system identically, occasionally failures occurred on different pieces of equipment. In addition, three requests for support of unscheduled spacecraft contacts were also scheduled every session. Again, the sessions were structured such that the requests were identical across subjects for a given session.

**Dependent Measures**

Eleven baseline dependent measures were developed for GT-MSOCC (Mitchell & Saisi, 1987; Mitchell & Forren, 1987; Saisi, 1986). These measures plus five additional measures to determine how many of the different types of equipment failures were corrected by the subjects were used in the experiment. The performance measures are grouped into four categories: fault compensation, equipment configuration and deconfiguration, operator errors, and percentage of failures corrected.

The fault compensation measures reflect the time to compensate for each of the four types of failures. If the subject failed to compensate for the failure, the measure reflects the total time the failure was present in the system. The next group of performance measures reflect the time to respond to various equipment configuration and deconfiguration requests.

The operator error measures reflect the number of errors committed by the operator. Two types of errors can occur. The first type is when the operator causes a conflict with the automated scheduler. The second type occurs when the operator replaces a component that has not failed. The last group of performance measures reflect the accuracy of the operator’s fault detection strategy. The measure reflects the percentage of errors of a given type that the subject corrected during the session. A separate measure is used for each type of failure. In addition, a separate measure was used to reflect the percentage of total errors corrected.

**Analysis**

A mixed effect, nested factorial design was used to analyze the data. Because some of the dependent measures did not have a fixed number of repetitions per cell, the design was unbalanced in some cases.

The primary factor of interest is Condition which reflects the type of associate, i.e., human associate or ALLY. The experimental design was a repeated measures design in that each subject was exposed to both of the experimental conditions.

To control for the variability across the subjects, Subject was included as a factor in the experimental design. The Subject effect was included in both of 10 levels to reflect the 10 experimental subjects. In order to account for any variability in the order in which the subjects worked with the two associates, Group was added as a factor in the experimental design. The Group factor includes two levels. The subjects in Group 1 worked with the human associate first, and the subjects in Group 2 worked with ALLY first. Subject, therefore, is a nested factor within Group.

Finally, Session was included as a factor to account for any variability between the sessions. The Session effect included four levels to reflect the four data collection sessions.

Analyses of variance were performed to determine the effect of each of the four independent variables (Condition, Group, Session, and Subject) on each of the sixteen dependent measures. An alpha level of .10 was used to detect significant effects.

Since the experimental design was a mixed design with random and fixed effects, approximate F statistics were constructed using Satterthwaite's method (Montgomery, 1984). Statistical analyses were performed using the General Linear Model (GLM) procedure of the SAS statistical software package (Spector, Goodnight, Sall, and Sarle, 1985). The GLM procedure computes the expected mean squares which were used to compute Satterthwaite's approximate F-statistic and the adjusted degrees of freedom. These values were then used to compute the significance level of the effects. In addition to the statistical analysis, the results of the surveys and analysis of audit logs of the subjects' activities were examined to gain additional insight into the individual interaction.
A questionnaire to capture subjectively the strategy they used to interact with the associate in strategies used by the subjects. These analyses, in conjunction with the statistical analyses, were used to evaluate the effectiveness of ALLY as an operator's associate and to evaluate the proposed theory of cooperation as it was implemented in ALLY.

**Discussion**

The experimental results are summarized in Figure 3 and 4. Figure 3 summarizes the means and standard deviations for the two associate conditions across the 16 performance measures. Figure 4 provides a graphical comparison of ALLY's performance compared with the human associate. While these figures indicate that, on the average, ALLY tended to perform better than the human associate, only two of the performance measures yielded significant differences. These were the time to compensate for software type 1 failures (i.e., software failure characterized by termination of data flow) and the number of correct responses to unscheduled support requests. On all other performance measures, ALLY performed as well as the human associate. A more exhaustive discussion of the results is found in Bushman (1989).

<table>
<thead>
<tr>
<th>Dependent Measure</th>
<th>Human Associate Mean</th>
<th>Std. Dev</th>
<th>ALLY Mean</th>
<th>Std. Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hardware failures</td>
<td>33.4</td>
<td>22.9</td>
<td>26.5</td>
<td>19.3</td>
</tr>
<tr>
<td>software failures 1</td>
<td>115.9</td>
<td>55.9</td>
<td>89.4</td>
<td>49.3</td>
</tr>
<tr>
<td>software failures 2</td>
<td>219.9</td>
<td>164.0</td>
<td>139.1</td>
<td>106.0</td>
</tr>
<tr>
<td>software failures 3</td>
<td>109.4</td>
<td>82.5</td>
<td>102.7</td>
<td>91.4</td>
</tr>
<tr>
<td>scheduled contacts</td>
<td>32.5</td>
<td>36.0</td>
<td>35.6</td>
<td>36.5</td>
</tr>
<tr>
<td>correct responses</td>
<td>7.5</td>
<td>6.7</td>
<td>2.6</td>
<td>0.5</td>
</tr>
<tr>
<td>support requests</td>
<td>172.1</td>
<td>156.0</td>
<td>108.5</td>
<td>117.1</td>
</tr>
<tr>
<td>unscheduled contacts</td>
<td>185.3</td>
<td>161.6</td>
<td>120.5</td>
<td>174.6</td>
</tr>
<tr>
<td>unscheduled failures</td>
<td>7.7</td>
<td>7.7</td>
<td>8.7</td>
<td>11.6</td>
</tr>
<tr>
<td>correct responses</td>
<td>12.2</td>
<td>9.8</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>error operation</td>
<td>1.0</td>
<td>1.0</td>
<td>1.3</td>
<td>1.8</td>
</tr>
<tr>
<td>% hardware fixed</td>
<td>89.2</td>
<td>6.5</td>
<td>100.0</td>
<td>0.0</td>
</tr>
<tr>
<td>% software 1 fixed</td>
<td>83.7</td>
<td>23.7</td>
<td>93.1</td>
<td>48.1</td>
</tr>
<tr>
<td>% software 2 fixed</td>
<td>85.0</td>
<td>23.8</td>
<td>93.7</td>
<td>20.2</td>
</tr>
<tr>
<td>% software 3 fixed</td>
<td>91.2</td>
<td>18.2</td>
<td>98.7</td>
<td>7.2</td>
</tr>
<tr>
<td>% total fixed</td>
<td>90.8</td>
<td>7.5</td>
<td>96.7</td>
<td>6.3</td>
</tr>
</tbody>
</table>

**Figure 3. Summary Performance Measures**

**Figure 4a. Mean Performance Measures by Condition**

**Figure 4b. Mean Performance Measures by Condition**

**Figure 4c. Mean Performance Measures by Condition**

While in only two cases a significant difference was detected between ALLY and the human associate, in most of the performance measures a significant condition by subject interaction was detected. This section presents the results of an in-depth analysis of the differences among the subjects. The following sections present a discussion of the results in the four major categories of performance measures: fault compensation, equipment configuration, operator errors, and percentage of errors detected. Finally, the section concludes with a discussion of the subjective evaluations of the experiment derived from questionnaires.

**Fault Compensation**

The first category of performance measures reflects the time to detect and compensate for failures in the system. The analysis indicates that the effect ALLY had on performance depended
primarily on the cooperation strategy the subjects used. Subjects that used a more active strategy that takes advantage of ALLY's monitoring and troubleshooting control tasks were able to perform generally better with ALLY than with the human associate. Subjects that used a more passive strategy by relying on ALLY's automatic monitoring and troubleshooting capability, however, performed as well as with the human associate. Overall, the use of ALLY as an associate resulted in performance that was at least as effective as the human associate.

Equipment Configuration

The effectiveness of using ALLY as an associate in response to the various configuration and deconfiguration functions primarily is a factor of the subject's style of interaction. In responding to conflicts with the automated schedule, those subjects that chose to perform these tasks manually performed better than subjects that used ALLY. Lack of planning (ALLY cannot foresee these events) and the need to check ALLY's answers were the contributing factors to ALLY's slower performance.

ALLY performed as well as the human associate in responding to unscheduled support requests. ALLY, however, resulted in fewer incorrect responses than the human associate. No differences were detected with deconfiguration requests because the subjects performed most of these tasks manually, even when they had ALLY as an associate.

Operator Errors

The next category of performance measures relate to operator errors. Two types of errors were recorded. The first type of error relates to operator actions that cause a conflict with the automated schedule. The other type relates to replacing a component that had not failed.

With respect to the first type of error (schedule conflicts), the analysis indicated that the subjects that used a more cautious strategy tended to generate fewer schedule conflicts. They would regularly check ALLY's replacements and the equipment it identified for support requests. The subjects that gave more responsibility to ALLY to replace components and schedule missions tended to generate more schedule conflicts.

No significant differences were detected with respect to the number of times the operator replaced a component that had not failed. This indicates that ALLY was just as effective as the human associate in correctly identifying equipment failures.

Percentage of Failures Detected

The analysis indicated that the subjects that used a more active fault compensation and detection strategy were able to detect more of the failures than the subjects that used a more passive strategy. The more successful subject consistently used ALLY to identify software failures before ALLY's automatic processing would detect them.

Subjective Evaluations

In addition to the above quantitative analysis, the subjects were asked to provide subjective evaluations of the two associates. Several different types of questionnaires were used to collect this information. This section summarizes the significant findings from these questionnaires.

In summary, the subjects felt that ALLY brought definite strengths to the task. ALLY's speed and accuracy at performing the monitoring tasks were cited as its major strengths. In addition, ALLY could quickly search schedules for free equipment. On the other hand, they indicated several limitations to the use of ALLY. They had to build their trust in the system. Some of the subjects were able to build the confidence in ALLY and gave it more responsibility. Others, however, needed more experience with the associate before the trust could be established.

At times, ALLY was "resistant" in that it would not change its mind once it found an answer, but the subjects tended to feel like they were out of control because they had the capability to override ALLY's choices manually.

A common "fault" found with ALLY was that it made the job too easy. Those subjects that actively worked with ALLY to get it to do things, however, felt like they had more control over the situation because they were relieved from the mundane tasks.

Summary

Overall, the performance of the subjects using ALLY as an associate was as effective as performance with the human associate. Individual strategies enabled some of the subjects to perform better with ALLY than with the human associate. The primary area that was affected by personal strategies was in detecting and compensating for software failures. Several subjects were able to develop a style of interacting with ALLY that enabled them to detect software failures before either one of them would on their own. This enabled them to detect the failures faster and to correct a larger percentage of the total failures.

Since ALLY does not have the capability to anticipate schedule conflicts, it is not able to plan for these events in advance. The subjects that relied on ALLY's capability to respond to these schedule conflicts could not take advantage of their planning ability. The subjects that performed the best with ALLY did not rely heavily on ALLY, but relied on their own capabilities to anticipate and plan for these events.

An unexpected result was a side-effect associated with the difficulty ALLY has with planning. ALLY performed as well as the human associate in responding to unscheduled support requests. However, because the subjects knew that this was one area in which ALLY can make mistakes, they regularly checked ALLY's answers. As a result, this additional checking resulted in more correct responses to support requests with ALLY.

Conclusions

This experiment demonstrated that a computer-based associate based on a model of the operator's function can perform as well as a human associate. As with any cognitive system (either human or artificial), ALLY brought with it strengths and limitations. The subjects that performed the best...
with ALLY were able to capitalize on its strengths and compensate for its weaknesses. The result was an overall increase in the system performance.

This research has demonstrated that a computer-based associate founded on the identified principles of human-machine cooperation can achieve performance compatible with a human associate. In addition, this research has provided a "starting point" from which a finer theory of cooperation can be developed. The significance of this research is that it has provided empirical research concerning the nature of human-machine cooperation.

Quantitative experimental data demonstrated the feasibility of the architecture for a computer-based associate that can perform at least as well as a human associate. Qualitative data, in the form of subjective evaluations, identified some of the varied strategies used by operators to interact with a computer-based associate. These quantitative and qualitative analyses may form the basis of a more refined theory of human-machine cooperation. Since no theory exists, this exploratory research is essential to develop a more definitive theory of cooperation.

ACKNOWLEDGMENTS

This research was supported in part by NASA Goddard Space Flight Center under contract NASS-28575 (Walt Truszkowski, technical monitor), and NASA Ames Research Center NAG 2-413 (Everett Palmer, technical monitor).

REFERENCES


*Bushman is now with the Training Systems Division, Air Force Human Resources Laboratory, Brooks Air Force Base, Texas, 78235.

*Rubin is now with ParcPlace Systems, 1550 Plymouth St, Mountain View, California, 94043.