Teamwork-Centered Autonomy for Extended Human-Agent Interaction in Space Applications

J. M. Bradshaw*, A. Acquisti*, J. Allen*, M. Breedy*, L. Bunch*, N. Chambers*, L. Galescu*, M. Goodrich****, R. Jeffers*, M. Johnson*, H. Jung*, S. Kulkarni*, J. Lott*, D. Olsen****, M. Sierhuis**, N. Suri*, W. Taysom,* G. Tonti*, A. Uszok*, R. van Hoof****

* Institute for Human and Machine Cognition (IHMC), 40 S. Alcaniz, Pensacola, FL 32502
{jbradshaw, jallen, mbreedy, lbunch, rjeffers, mjohnson, hjung, skulkarni, jlott, nsuri, gtonti, auszok}@ihmc.us
** RIACS and QSS, NASA Ames, MS T35B-1, Moffett Field, CA 94035, {msierhuis, rvanhoof}@mail.arc.nasa.gov
*** School of Public Policy and Mgmt., Carnegie Mellon University, Hamburg Hall 2105C, Pittsburgh, PA, acquisti@andrew.cmu.edu
**** Computer Science Department, Brigham Young University, Provo, UT 84602, {mike, olsen}@cs.byu.edu

Abstract

This paper summarizes our efforts to bring together and extend the best in current theory and technologies for *teamwork-centered autonomy* for space applications. Traditional planning technologies at the foundation of intelligent robotic systems typically take an *autonomy-centered* approach, with representations, mechanisms, and algorithms that have been designed to ingest a set of goals and output a complete plan in the most efficient and sound fashion possible. A *teamwork-centered autonomy* approach, on the other hand, takes as a beginning premise that people are working in parallel alongside autonomous systems, and hence adopts the stance that the processes of understanding, problem solving, and task execution are necessarily incremental, subject to negotiation, and forever tentative. Thus, a successful approach to teamwork-centered autonomy will require that every element of the autonomous system be designed to facilitate the kind of give-and-take that quintessentially characterizes natural and effective teamwork among groups of people. We briefly describe the major components of this approach and current efforts to apply and evaluate its utility from both human-centered and cost-benefit perspectives.

Introduction

Because ever more powerful intelligent agents will interact with people in increasingly sophisticated and essential ways, greater attention must be given to the technical and social aspects of human-agent teamwork. Although recent descriptive and theoretical research have begun to elucidate the principles of human-agent teamwork in realistic settings [7], the implementation of these principles in operational scenarios has been impeded by the lack of suitable *teamwork-centered autonomy* technologies.

Traditional planning technologies at the foundation of intelligent robotic systems typically take an *autonomy-centered* approach, with representations, mechanisms, and algorithms that have been designed to ingest a set of goals and output a complete plan in the most efficient and sound fashion possible. A *teamwork-centered autonomy* approach, on the other hand, takes as a beginning premise that people are working in parallel alongside autonomous systems, and hence adopts the stance that the processes of understanding, problem solving, and task execution are necessarily incremental, subject to negotiation, and forever tentative. Thus, a successful approach to teamwork-centered autonomy will require that every element of the autonomous system be designed to facilitate the kind of give-and-take that quintessentially characterizes natural and effective teamwork among groups of people. This represents an important departure from past work in autonomy, which has historically ignored such issues in core design of supporting technologies—token gestures to tack on a thin veneer of human interface as an afterthought notwithstanding!

Moreover, past research has paid scant attention to the development and use of targeted methodologies and metrics to understand, measure, and predict the impact of autonomous agent systems that work closely and continuously with people in realistic work settings. Evaluating and predicting this impact can be done from two perspectives: *human-centered* and *cost-benefit*. From a *human-centered evaluation* perspective, we want to draw from and extend previous research on human-computer interaction to answer questions about operator impact on the use of autonomous agent technologies, and elucidate aspects of system design that will lead to the development of technologies that work well with space scientists and other operators. A well-designed system minimizes *interaction time*, defined as the time required for a human to interact with the system, and maximizes *neglect time*, defined as the time that transpires between human-machine interactions. From a *cost-benefit* perspective, because autonomous systems typically involve significant development costs over extended periods of time, we need to be able to justify the investment in such systems in joint human-agent work settings, and compare them to scenarios involving either exclusively human or exclusively robotic alternatives.

From this perspective, four aspects of a teamwork-centered autonomy approach merit particular attention in our research approach:

- careful observation, modeling, and simulation of *work practice;*
- *teamwork policies* that embody a body of principles of natural and effective human-robotic interaction derived from this observation, modeling, and experimental interventions;
- *collaborative planning technologies* that are designed from the ground up to support partial sharing and incremental revision of plan state;
- appropriate *evaluation methodologies* extended and applied, first, to determine how well the joint human-agent systems being studied achieve their intended effects in various work settings—and to discover and assess their unintended effects—and, second, to provide continuous feedback into the loop of ongoing modeling and design refinement.

In this paper we outline new aspects of an effort we have recently undertaken to bring together and extend the best in current theory and technologies for teamwork-centered autonomy. This approach involves the following components:

- Brahms, a simulation environment with a rich library of models of human and agent activity in a variety of space environments [10; 15];
- KAoS, a set of services oriented to the management of policies specifying natural and effective teamwork behavior [5; 6; 7; 9];
- A Collaboration Management Agent (CMA), based on a framework for collaborative planning uniquely suited to the incremental addition and removal of constraints and the development of partial plan solutions [2; 3; 4].
- Extensions and applications to formal methodologies to analyze and compare the performance and effectiveness of autonomous systems from both the *human-centered* [12; 13] and *cost-benefit* [17] perspectives.

This new research draws on our experience in previous research to develop theory and tools supporting human-agent teamwork in space applications [1; 7; 8; 16]. Extending these previous components to support a CMA, which is built on a planner designed from the ground up to support human collaboration, will allow us to accelerate the transition of teamwork-centered autonomous systems from design and simulation to effective implementation and operation.

The relationship between these components is significant and synergistic. The rich model provided by Brahms of activities, humans, agents, and objects that are part of those activities represents the dynamic decision context that tunes KAoS policies and informs CMA problem-solving functionality. CMA manages problem-solving interaction among humans and agents within the bounds of KAoS policy constraints and consistent with the situational context provided by the Brahms model. The policy specification, representation, conflict resolution, and enforcement mechanisms of KAoS will assure that a coherent set of teamwork policies for safe and effective human-agent interaction can be continuously in effect throughout the ongoing problem-solving and human-agent interaction processes.

Our current research applications require both a continuation of full-scale field tests in joint human-robotic simulated surface exploration experiments [11; 16] as well as limited-objective experiments that will be performed at the IHMC human-robotic testbed site. Consistent with our work practice orientation, we rely heavily on simulation and field-testing as a guide to the design of work systems that use autonomous agents to support human activity.

The following sections briefly describe the major elements of our approach, including:

- Observing, modeling, and simulating work practice;
- Collaborative problem solving in human-agent teams;
- Teamwork policies and mechanisms;
- Evaluation methodologies.

Observing, Modeling, and Simulating Work Practice

Results from previous field tests that show just how many practical challenges face humans and robots just to jointly perform basic logistical activities as part of simulated surface exploration, let alone their science missions [11; 16]. Although previous field tests uncovered these challenges, time and resources did not always allow for full analysis of their causes or for exploration of solutions. Making extensive use of the Brahms modeling and simulation environment, we are using post hoc analyses of past field test data as well as data from new field studies and limited objective tests to develop aspects of Brahms models for teamwork-centered autonomy for extended human-robotic surface exploration missions, under different assumptions of crew size, mission duration, current and future robotic capabilities, and so forth.

As one example, field test data show that communication breakdowns are frequent and more attention needs to be paid to the development of requirements and capabilities in support of human team members under conditions of frequent disconnected operation. We are currently performing a careful analysis of examples of human planning to understand how they can inform our understanding of how to build effective collaborative planning systems. The role of high-level policies to ensure astronaut safety and control of agent behavior under all circumstances is being examined, as well as policies for adjustable autonomy. In later phases of our study, dealing effectively with resource limitations (e.g., CPU, power, bandwidth, time) and real-time recovery from buggy software or malfunctioning hardware will be an additional important focus based on our observations in the field.

One of the most significant limitations of the current Mobile Agents Architecture (MAA) being deployed in planetary surface exploration studies is that agents do not use a model of teamwork to help coordinate their interactions with humans and other agents. As one consequence, the agents are not able to process more than one request at a time, nor to associate a series of tasks or requests that support a common team goal. Agents need to be able to draw connections between requests from different agents, prioritize which request they should handle first, and keep other team members appropriately informed of the status of the discharge of team goals. Within the annual full-scale field tests each spring, we plan to apply models and policies for teamwork to MAA.

Moreover, while the MAA currently supports the EVA astronauts during the EVA in capturing science and biosensor data, it offers no help for astronauts needing to store this data in the habitat, nor is there support for the crew to access this data either during or after the EVA for science analysis and discussions with other crewmembers. We are extending the multiagent teamwork model to handle for data storage and access environment for planetary EVAs.

Collaborative Problem Solving in Human-Agent Teams

The collaboration management agent (CMA) is designed to support human-agent, human-human, and agent-agent interaction and collaboration within mixed human-robotic teams. While it will always be the case that agents may operate autonomously without the support of the CMA, our goal is to show that when the CMA is "in the loop" the overall activities of the human and autonomous agents are more coordinated and efficient. By implementing the CMA as an optional "assist" to the agents rather than a coordinator that must be consulted for every aspect of agent interaction we avoid concerns of having a team of agents depend on a single central processor that might become inoperable or simply just be out of communication range. The CMA improves overall effectiveness during those times that it is available.

The CMA interacts with individual agents in order to 1) maintain an overall picture of the current situation and status of the overall plan, as complete as possible based on available reports, 2) detect possible failures that become more likely as the plan execution evolves and invoke replanning; 3) evaluate the viability of proposed changes to plans by agents, 4) manage re-planning when situations exceed the capabilities of individual agents, including recruiting more capable agents to perform the re-planning, 5) manage the re-tasking of agents when changes are made, 6) adjust its communications to the capabilities of the agents (e.g., graphical interfaces work well for a human but wouldn't help most other agents). For example, the CMA allows agents with limited planning capabilities to benefit from planning assistance from other more capable agents when problems arise. Consider a non-planning agent that cannot continue executing its plan because of some obstacle. It reports the problem and waits. Although the CMA has the ability to resolve some problems autonomously, in this case, let's say it decides that it requires human assistance. The CMA then reports the problem to a human who interacts with it to explore the problem and decide on a solution. The CMA collaborates in this process, evaluating proposed solutions to find potential problems. Once a solution is found, the CMA then manages the tasking to ensure all the individual agents change their plans as needed. And when changes are made to the overall plan—say, by a human in response to an unexpected situation, all the agents involved in the changes have to be re-tasked in a manner that is consistent with organizational policies, approved procedures, and socially-acceptable practices. Since the agents will be in different states based on how much of their original plan they have executed, the CMA must support further negotiation and re-planning among team members before the overall plan is deemed viable and put into execution.

Core research issues we are currently addressing include 1) the development of a temporal representation of the situation and of collaborative goals and plans that supports re-planning; 2) the development of an ontology of collaborative problem-solving and the types of interactions agent's perform when collaborating; 3) the development of intention recognition algorithms that can identify the intended collaborative act from user input; 4) the development of an agent communication language that will work across Brahms, CMA, and KAoS; 5) representing and reasoning about the capabilities of agents to assist in automatic tasking; 6) modeling communicative capabilities to facilitate agent communication across agents of very different capabilities; and 7) the development of effective human interfaces for interaction with the CMA. While we support a multi-modal spoken language interface for the human team members, we are using an existing technology base [3] and so this is not be a focus of the current research.

Teamwork Policies and Mechanisms

Whereas early research on agent teamwork focused mainly on agent-agent interaction, teamwork principles have now been proposed in the context of human-agent interaction [5; 7]. The vision of future human-robotic operations is that of loosely coordinated groups of humans and agents [14]. As capabilities and opportunities for autonomous operation grow in the future, autonomous agents will perform their tasks for increasingly long periods of time with only intermittent supervision. Most of the time routine operation is managed by the agent software that controls these vehicles while the human crews perform other tasks. Occasionally however, when unexpected problems or novel opportunities arise, operators must assist the agents. Because of the loose nature of these groups, such communication and collaboration must proceed asynchronously and in a mixed-initiative manner. Humans must quickly come up to speed on situations with which they may have had little involvement for hours or days. Then they must cooperative effectively and naturally with the agents as true team members.

This vision points to two major opportunities for KAoS policy and domain services in the context of the current research:

- the use of policy to assure that unsupervised autonomous agent behavior is kept within safe and secure limits of prescribed bounds, even in the face of buggy, poorly designed, unsophisticated, or malicious agent code;
- the use of policy to assure effective and natural human-agent team interaction, without individual agents having to be specifically programmed with the knowledge to do so.

Current rigid and simplistic approaches supporting maintenance of joint team goals among autonomous agents are not up to the task of supporting mixed groups of humans and agents. Over the past few years, we have been engaged in the process of abstracting the results of work practice studies and field tests into declaratively-specified teamwork policies in KAoS to support close and continuous interaction among agents and people [5].

Policy-based mechanisms embedded within the robots aim to enable a high-level of trust and efficiency. For example, policies governing sensitive or risky behavior attempt to ensure that the robots operate safely, robustly, and strictly within the range of permitted action even in the face of system failures, loss of communication capabilities, potentially buggy software, or human error. Policies regulating adjustable autonomy for surface exploration scenarios are being designed and tested to appropriately offload human tasks as needed while assuring that final decision authority rests with people. Coordination policies working in conjunction with collaborative planning capabilities and work practice models are designed to make sure that interaction is as effective and natural as possible, with the robots assuming the role of true team players alongside the humans.

One limitation of previous versions of KAoS is that the user interface for creating and deploying policies (i.e., KPAT) was oriented toward technical specialists. We would like to see whether we can make the policy management capabilities invisible to the astronauts working with the astronauts, i.e., to make it seem that they are just giving advice to the robot, when in reality behind the scenes policies are being created, deleted, and maintained. In addition to KPAT, we have begun developing dynamic visualization components coupled with spoken dialogue that will integrate with information provided by CMA and Brahms.

Building on the work of previous NASA IS-sponsored research, our new work will put us in a position to verify the effectiveness of KAoS policies and services through a series of tests assessing *survivability* (ability to maintain effectiveness in the face of unforeseen software or hardware failures), *safety* (ability to prevent certain classes of dangerous actions or situations), *predictability* (assessed correlation between human judgment of predicted vs. actual behavior), *controllability*(immediacy with which an authorized human can prevent, stop, enable, or initiate agent actions), *effectiveness* (assessed correlation between human judgment of desired vs. actual behavior), and *adaptability* (ability to respond to changes in context).

Evaluation Methodologies

Besides these issues in developing a viable approach to teamwork-centered autonomy, we note that appropriate methodologies and metrics for comparing and evaluating the performance and effectiveness of mixed human-agent teams are lacking. Evaluating and predicting this impact can be done from two perspectives: *human-centered* and *cost-benefit.*

From a *human-centered evaluation* perspective, we want to draw from and extend previous research on human-computer interaction to answer questions about the development of technologies that work well with space scientists and other operators [12]. A well-designed system minimizes *interaction time*, defined as the time required for a human to interact with the system, and maximizes *neglect time*, defined as the time that transpires between human-machine interactions [13].

Once the first version of the base system is complete, we propose to begin two iterative parallel approaches for the human-centered evaluation of the system:

- The first approach uses secondary task experiments in a laboratory setting to evaluate the cognitive impact of a new autonomous system. This approach assesses neglect tolerance and interaction efficiency by creating artificial loads on the operator. Under these artificial load conditions, task-relevant performance is measured. By associating the relationship between load and performance, predictions can be made about how new autonomy, new activities, or new mission conditions will affect performance.
- The second approach models limiting cases of human-machine interaction. This approach describes the maximum number of human activities that can be performed using various interaction technologies. By associating various technologies with a maximum activity number, predictions can be made about how performance will decline or improve as new autonomy, new activities, or new mission conditions are introduced.

Our primary contribution will be to extend our current evaluation methodology so that it can be used in the proposed project. The element of this project that most need to be addressed in the methodology is the distribution of abilities across human and heterogeneous artificial agents. As results from laboratory experiments are obtained, they will be incorporated as part of the limited-objective and full-scale field experiments.

Several methodologies have been proposed in the planning and business literature to measure performance, effectiveness, or cost efficiency of tasks and processes from a *cost-benefit* perspective. While these existing methodologies have been useful in the study of profit-based enterprises, they have to be modified to satisfy the particular requirements of research-oriented organizations such as NASA. We are currently assimilating the literature on cost analysis, benefit analysis, research and development evaluation studies in the context of metrics for evaluation of human-computer interaction, so that in later phases of the research we will be able to analyze issues specific to NASA needs, focusing on the unique properties of agent-based systems and the output productivity of research activities. A well-grounded methodology will also require the analysis of existing surveys and new interviews of humans developing and interacting with autonomous systems, analysis of their various costs (e.g., development, customization, implementation, learning curve, failures, breakdowns, etc.), analysis of their various benefits (cost efficiencies time efficiencies, learning by doing), and analysis of comparable methodologies in the literature. By year two of the research, we will have brought the first version of the methodology to a point that it can be used in conjunction with the field tests.

References

[1] Acquisti, A., Sierhuis, M., Clancey, W. J., & Bradshaw, J. M. (2002). Agent-based modeling of collaboration and work practices onboard the International Space Station. *Proceedings of the Eleventh Conference on Computer-Generated Forces and Behavior Representation*. Orlando, FL,

[2] Allen, J., Byron, D. K., Dzikovska, M., Ferguson, G., Galescu, L., & Stent, A. (2000). An architecture for a generic dialogue shell. *Journal of Natural Language Engineering*, 6(3), 1-16.

[3] Allen, J. F., Byron, D. K., Dzikovska, M., Ferguson, G., Galescu, L., & Stent, A. (2001). Towards conversational human-computer interaction. *AI Magazine*, 22(4), 27-35.

[4] Allen, J. F., & Ferguson, G. (2002). Human-machine collaborative planning. *Proceedings of the NASA Planning and Scheduling Workshop*. Houston, TX,

[5] Bradshaw, J. M., Beautement, P., Raj, A., Johnson, M., Kulkarni, S., & Suri, N. (2003). Making agents acceptable to people. In N. Zhong & J. Liu (Ed.), *Intelligent Technologies for Information Analysis: Advances in Agents, Data Mining, and Statistical Learning*. (pp. in press). Berlin: Springer Verlag.

[6] Bradshaw, J. M., Jung, H., Kulkarni, S., & Taysom, W. (2004). Dimensions of adjustable autonomy and mixed-initiative interaction. In M. Klusch, G. Weiss, & M. Rovatsos (Ed.), *Computational Autonomy*. (pp. in press). Berlin, Germany: Springer-Verlag.

[7] Bradshaw, J. M., Sierhuis, M., Acquisti, A., Feltovich, P., Hoffman, R., Jeffers, R., Prescott, D., Suri, N., Uszok, A., & Van Hoof, R. (2003). Adjustable autonomy and human-agent teamwork in practice: An interim report on space applications. In H. Hexmoor, R. Falcone, & C. Castelfranchi (Ed.), *Agent Autonomy*. (pp. 243-280). Kluwer.

[8] Bradshaw, J. M., Sierhuis, M., Acquisti, A., Feltovich, P., Hoffman, R. R., Jeffers, R., Suri, N., Uszok, A., & Van Hoof, R. (2003). Living with agents and liking it: Addressing the technical and social acceptability of agent technology. *AAAI Stanford Spring Symposium on Human Interaction with Autonomous Systems in Complex Environments*. Menlo Park, CA, AAAI Press,

[9] Bradshaw, J. M., Uszok, A., Jeffers, R., Suri, N., Hayes, P., Burstein, M. H., Acquisti, A., Benyo, B., Breedy, M. R., Carvalho, M., Diller, D., Johnson, M., Kulkarni, S., Lott, J., Sierhuis, M., & Van Hoof, R. (2003). Representation and reasoning for DAML-based policy and domain services in KAoS and Nomads. *Proceedings of the Autonomous Agents and Multi-Agent Systems Conference (AAMAS 2003)*. Melbourne, Australia, New York, NY: ACM Press,

[10] Clancey, W. J., Sachs, P., Sierhuis, M., & van Hoof, R. (1998). Brahms: Simulating practice for work systems design. *International Journal of Human-Computer Studies*, 49, 831-865.

[11] Clancey, W. J., Sierhuis, M., & et_al. (2003). Advantages of Brahms for specifying and implementing a multiagent human-robotic exploration system. *Proceedings of the Sixteenth Annual Conference of the Florida AI Society (FLAIRS 2003)*. St. Augustine, FL,

[12] Goodrich, M. A., Olsen Jr., D. R., Crandall, J. W., & Palmer, T. J. (2001). Experiments in adjustable autonomy. *Proceedings of the IJCAI_01 Workshop on Autonomy, Delegation, and Control: Interacting with Autonomous Agents*. Seattle, WA, Menlo Park, CA: AAAI Press,

[13] Olsen Jr., D. R., & Goodrich, M. A. (2003). Metrics for evaluating human-robot interactions. *Proceedings of PERMIS 2003*.

[14] Schreckenghost, D., Martin, C., & Thronesbery, C. (2003). Specifying organizational policies and individual preferences for human-software interaction. *Submitted for publication*.

[15] Sierhuis, M. (2001) *Brahms: A Multi-Agent Modeling and Simulation Language for Work System Analysis and Design*. Doctoral, University of Amsterdam.

[16] Sierhuis, M., Bradshaw, J. M., Acquisti, A., Van Hoof, R., Jeffers, R., & Uszok, A. (2003). Human-agent teamwork and adjustable autonomy in practice. *Proceedings of the Seventh International Symposium on Artificial Intelligence, Robotics and Automation in Space (i-SAIRAS)*. Nara, Japan,

[17] Staubus, G. J. (1971). *Activity Costing and Input-Output Accounting*. Homewood, IL: R. D. Irwin.