Chapter 0

Adjustable Autonomy and Human-Agent Teamwork in Practice: An Interim Report on Space Applications

JEFFREY M. BRADSHAW*, MAARTEN SIERHUIS**, ALESSANDRO ACQUISTI***, PAUL FELTOVICH*, ROBERT HOFFMAN*, RENIA JEFFERS*, DEBBIE PRESCOTT****, NIRANJAN SURI*, ANDRZEJ USZOK*, and RON VAN HOOF**

** Institute for Human and Machine Cognition (IHMC), University of West Florida, 40 S. Alcaniz, Pensacola, FL 32501 {jbradshaw, pfeltovich, rhoffman, rjeffers, nsuri, auszok}@ai.uwf.edu*

*** RIACS and QSS, NASA Ames, MS T35B-1, Moffett Field, CA 94035, {msierhuis, rvanhoof}@mail.arc.nasa.gov*

**** SIMS, UC Berkeley, 102 South Hall, Berkeley, CA 94720, acquisti@mail.arc.nasa.gov*

***** The Boeing Company, 499 Boeing Blvd. MS JW-75, P.O. Box 240002, Huntsville, AL 35824, debbie.prescott@boeing.com*

Key words: Teamwork, human-agent interaction, human-robotic teamwork, human-centered computing, cognitive prostheses, augmented cognition, mixed-initiative, adjustable autonomy, multi-agent systems, Brahms, KAoS, NOMADS, policy, agent-based social simulations, International Space Station, space applications, Personal Satellite Assistant, Robonaut, DAML*OIL

Abstract: *We give a preliminary perspective on the basic principles and pitfalls of adjustable autonomy and human-centered teamwork. We then summarize the interim results of our study on the problem of work practice modeling and human-agent collaboration in space applications, the development of a broad model of human-agent teamwork grounded in practice, and the integration of the Brahms, KAoS, and NOMADS agent frameworks We hope our work will benefit those who plan and participate in work activities in a wide variety of space applications, as well as those who are interested in design and execution tools for teams of robots that can function as effective assistants to humans.*

1. INTRODUCTION

Tomorrow's world will be filled with agents embedded everywhere in the places and things around us [43; 44]. Providing a pervasive web of sensors and effectors, teams of such agents will function as *cognitive prostheses*—computational systems that leverage and extend human cognitive, perceptual, and collaborative capacities, just as the steam shovel was a sort of muscular prosthesis or eyeglasses a sort of visual prosthesis [40]. Thus the focus of AI research is destined to shift from Artificial Intelligence to Augmented Intelligence [12; 48].

While simple software and robotic assistants of various kinds today capture our attention, the future surely holds much more novel and sophisticated agent-powered devices than we can currently imagine. A key requirement for such devices is for real-time cooperation with people and with other autonomous systems. While these heterogeneous cooperating entities may operate at different levels of sophistication and with dynamically varying degrees of autonomy, they will require some common means of representing and appropriately participating in joint tasks. Just as important, developers of such systems will need tools and methodologies to assure that such systems will work together reliably, even when they are designed independently.

Teamwork has become the most widely-accepted metaphor for describing the nature of multi-agent cooperation. The key concept usually involves some notion of shared knowledge, goals, and intentions that function as the glue that binds team members together [27]. By virtue of a largely-reusable explicit formal model of shared intentions, team members attempt to manage general responsibilities and commitments to each other in a coherent fashion that facilitates recovery when unanticipated problems arise. For example, a common occurrence in joint action is when one team member fails and can no longer perform in its role. A general teamwork model might entail that each team member be notified under appropriate conditions of the failure, thus reducing the requirement for special-purpose exception handling mechanisms for each possible failure mode.

Whereas early research on agent teamwork focused mainly on agent-agent interaction, there is a growing interest in various dimensions of human-agent interaction [13]. Unlike autonomous systems designed primarily to take humans out of the loop, many new efforts are specifically motivated by the need to support close human-agent interaction [42; 53; 59; 81]. Under NASA sponsorship, we are investigating issues in human-robotic teamwork and adjustable autonomy. Future human missions to the Moon and to Mars will undoubtedly need the increased capabilities for human-robot collaborations we envision. Astronauts will live, work, and perform laboratory experiments in collaboration with robots inside and outside their spacecraft and habitats on planetary surfaces.

An adequate approach to design of cooperative autonomous systems requires first and foremost a thorough understanding of the kinds of interactive contexts in which humans and autonomous systems will cooperate. With our colleagues at the Research Institute for Advanced Computer Science (RIACS) at NASA Ames Research Center, we have begun to investigate the use of Brahms [75] as an agent-based design toolkit to model and simulate realistic work situations in space. The ultimate objective is to produce agent-based simulations in Brahms that could form the basis for the design of robotic and software agent functions for actual operations. On its part, IHMC is enhancing the KAoS [16; 18; 19] and NOMADS [80] agent frameworks to incorporate explicit general models of teamwork, mobility, and resource control appropriate for space operations scenarios. These models are represented largely in the form of policies [19; 32; 66].

In the following pages, we outline a preliminary perspective on the basic principles and pitfalls of adjustable autonomy and human-centered teamwork (section 2). We emphasize the importance of careful observation and modeling of human work practice as a foundation for teamwork theory, and the central role of adjustable autonomy in the design of agent systems. We then describe how we are beginning to apply this approach to human-agent collaboration in space applications (section 3). We introduce Brahms, a work practice modeling and social simulation environment that is being adapted to operate in conjunction with teamwork policies implemented within KAoS agent services [16; 19], and strong mobility and resource management services implemented in NOMADS [80]. We summarize our initial results in section 4.

2. HUMAN-CENTERED PERSPECTIVES ON TEAMWORK[1]

"Agents occupy a strange place in the realm of technology," observes Donald Norman, "leading to much fear, fiction, and extravagant claims" [69, p. 49]. Because ever more powerful intelligent agents will increasingly differ in important ways from conventional software of the past, we need to take into account the social issues no less than the technical ones if the agents we design and build are to be acceptable to people:

> "The technical aspect is to devise a computational structure that guarantees that from the technical standpoint, all is under control. This is not an easy task. The social part of acceptability is to provide reassurance that all is working according to plan… This is [also] a non-trivial task" [69, p. 51].

Elsewhere we have written about our efforts to address some of the technical aspects of control for increased human acceptability: safety, privacy, and other guards against error and maliciousness in long-lived agent communities [19]. In this chapter, we will focus mainly on issues of social acceptability, in particular with making sure that interactions between agents and people are as natural and effective as possible.[2]

Specific approaches to human-agent teamwork have been explored, including interface agents [31; 53; 59; 60]), mixed-initiative systems [22; 37], and collaboration theory [47; 71]. In addition, researchers such as Tambe [74; 81] have successfully adapted principles of agent-agent teamwork to human-agent interaction in various settings.

Unlike other approaches to agent teamwork, a human-centered perspective requires that the design of agents be problem-driven, activity-centered, and context-bound [85]. Thus we must begin with a detailed understanding of how people actually work. To enable such study, we have developed and used Brahms, a language coupled with an agent modeling and simulation environment that can capture complexities of observation, communication, and collaboration in the context of group work [75]. Our approach seeks to incorporate the best of previous research on human-centered collaboration and teamwork, while simultaneously grounding new findings in our own work practice study experience.[3]

2.1 Desiderata for Human-Centered Systems

Following a careful study of the introduction of automation into complex domains, Billings [8] described a set of desiderata for human-centered systems:

[1] Most of section 2 appears in expanded form in [13].

[2] For an entertaining and informative general characterization of various approaches to human-centered computing, see [51].

[3] Researchers in human-agent teamwork have used the term in two broad ways: 1) as a conceptual analogy for heuristically directing research (e.g., to build systems that facilitate fluent, coordinated interaction between the human and agent elements of the system as "team players") and 2) as the subject matter for research (e.g., to understand the nature of teamwork in people). The first activity focuses on practical engineering of useful systems through application of human-centered design principles, empirical studies of the use of these systems, and often a limited commitment to studying teamwork among people. The second activity is explicitly framed as a scientific study, and may have two angles: 1) providing information relevant to the design of successful human-agent systems, and 2) independent of application, understanding the nature of cognition, communication, and cooperation in people and animals. The latter activity is seen by these researchers as essential for achieving the ultimate goals of artificial intelligence. Since members of our research team are drawn from each tradition, our approach attempts to reflect sensitivity to both: neither undervaluing the independent study of social and cognitive aspects of human teamwork, nor slavishly imitating superfluous aspects of natural systems in the development of artificial ones, like an engineer who insists that successful airplane designs must necessarily feature flapping wings because all birds have them [41].

"*Premise*: Humans are responsible for outcomes in human-machine systems.
Axiom: Humans must be in command of human-machine systems.
Corollary: Humans must be actively involved in the processes undertaken by these systems.
Corollary: Humans must be adequately informed of human-machine system processes.
Corollary: Humans must be able to monitor the machine components of the system.
Corollary: The activities of the machines must therefore be predictable.
Corollary: The machines must also be able to monitor the performance of the humans.
Corollary: Each intelligent agent in a human-machine system must have knowledge of the intent of the other agents."[4]

Note that Billings' main premise ("Humans are responsible for outcomes in human-machine systems") implicitly assumes a fundamental asymmetry between humans and today's agents. Notwithstanding this assumption, we expect the broad balance between human and agent initiative and responsibility to co-evolve commensurate with the degree of trust humans are willing (or required) to exercise in particular kinds of technology for specific contexts of use. Already people rely routinely on technology to do things automatically for them that were unthinkable not too long ago. Section 2.2 shows how within the perspective of human-centered agent teamwork these asymmetries between people and machines are best viewed as broad complementarities. Section 2.3 briefly discusses research in adjustable autonomy that aims to enable principled yet dynamic flexibility in the roles of people and agents.

The corollaries to Billings' main premise serve to underscore the importance of maintaining appropriate mutual awareness among team members. Each actor, both human and agent, must not only be able to realistically assess the overall situation and current the state and intentions of the other team members, but also to accurately ascertain trends and reliably predict future states. Section 2.4 discusses the challenges that researchers face in trying to overcome these mutual "gulfs of evaluation and execution."

Finally, section 2.5 aims to correct the common misperception that automation is a simple substitution of machine for human effort. Instead, results of many years of research makes it clear that, more fundamentally, automated assistance changes the nature of the task itself.

2.2 Complementary Asymmetries in Human-Agent Interaction

Humans and artificial agents are two disparate kinds of entities that exist in very different sorts of worlds. For the foreseeable future there will be a fundamental asymmetry in their capabilities: the brightest agents will be limited in the generality if not the depth of their inferential, adaptive, social, and sensory capabilities; humans, though fallible, are functionally rich in reasoning strategies and their powers of observation, learning, and sensitivity to context [2]. Moreover, agents interact directly and efficiently in cyberspace but indirectly and awkwardly in the material sphere; humans shine in the world of atoms but cannot juggle bits on their own. Adapting to appropriate mutual roles that take advantage of the respective strengths of humans and agents, and crafting natural and effective modes of interaction are key challenges.

All this being said, we do not wish to extend here the long tradition of MABA-MABA (men-are-better-at/machines-are-better-at) lists that began with the classic report of Paul Fitts *et al.* [38]. The point is not to think so much about which tasks are best performed by humans and which by agents

[4] Note that this applies to people and to software agents.

but rather how tasks can best be shared to be done by both humans and agents working in concert [49]. Licklider [58] called this concept *man-computer symbiosis*.[5]

To counter the limitations of the Fitts' list, which is clearly intended to summarize what humans and machines each do well on their own, Hoffman has summarized the findings of Woods in an "un-Fitts list" [51] (see Table 1), which emphasizes how the competencies of humans and machines can be enhanced through appropriate forms of mutual interaction. These guidelines provide useful general heuristics for the development of our model of human-agent teamwork.

Machines	
Are constrained in that:	**Need people to:**
Sensitivity to context is low and is ontology-limited	Keep them aligned to context
Sensitivity to change is low and recognition of anomaly is ontology-limited	Keep them stable given the variability and change inherent in the world
Adaptability to change is low and is ontology-limited	Repair their ontologies
They are not "aware" of the fact that the model of the world is itself in the world	Keep the model aligned with the world
People	
Are not limited in that:	**Yet they create machines to:**
Sensitivity to context is high and is knowledge- and attention-driven	Help them stay informed of ongoing events
Sensitivity to change is high and is driven by the recognition of anomaly	Help them align and repair their perceptions because they rely on mediated stimuli
Adaptability to change is high and is goal-driven	Effect positive change following situation change
They are aware of the fact that the model of the world is itself in the world	Computationally instantiate their models of the world

Table 1. An "un-Fitts" list [51].

2.3 Adjustable Autonomy

Before discussing adjustable autonomy, it is important to describe the concept of autonomy itself in more detail.[6] Some important dimensions of autonomy can be straightforwardly characterized by reference to Figure 1, with degrees of autonomy corresponding to varying nested ranges of action

[5] The ultimate in such symbiosis is where the boundary between agents and people disappears altogether, with the agents being subsumed into the human's eudaemonic space (i.e., the agents seem to be part of the person).

[6] Note that there is a subtle difference in common usage of the word *autonomous* between Americans and Europeans. In the latter, the original sense of the term as describing something that is capable of self-government (Greek: *auto-* (self) + *nomos* (law)) strongly predominates, whereas in American English usage the senses of independence from outside control and self-directedness have developed stronger associations with the word (American Heritage Dictionary). This difference, while not affecting the ability of researchers from different backgrounds to understand one another, may sometimes affect the slant or emphasis put on various aspects of their respective conceptualizations of autonomy. Note, for example, Brainov and Hexmoor's emphasis on degree of autonomy as a relative measure of independence between an agent and the physical environment, and within and among social groups [20]. Luck *et al.* [61], unsatisfied with defining autonomy as a wholly relative concept, argue that the self-generation of goals should be the defining characteristic of autonomy, thus allowing it to be regarded in absolute terms that more clearly reflect the precedence of the sense of self-government.

available to an agent in a given context.[7] The nesting of the regions in the diagram is consistent with logical constraints that would ideally govern the relationship among these sets of actions—for example, the fact that an agent should not be obligated to perform an action that it is not permitted to do. The outermost scope defines a range of theoretically *possible actions* that could be taken by a maximally-autonomous agent with a "complete" set of capabilities, unqualified permissions, and no obligations. In practice, only a subset of these actions would be ones that an agent could be expected to *achieve independently*, however a larger set of actions could typically be *achievable* in concert with others. Finally, the extent of an agent's autonomy is affected by policy-based deontic constraints: the larger the range of its *permitted actions* and the smaller the set of its *obligations*, the more freely it can act.[8]

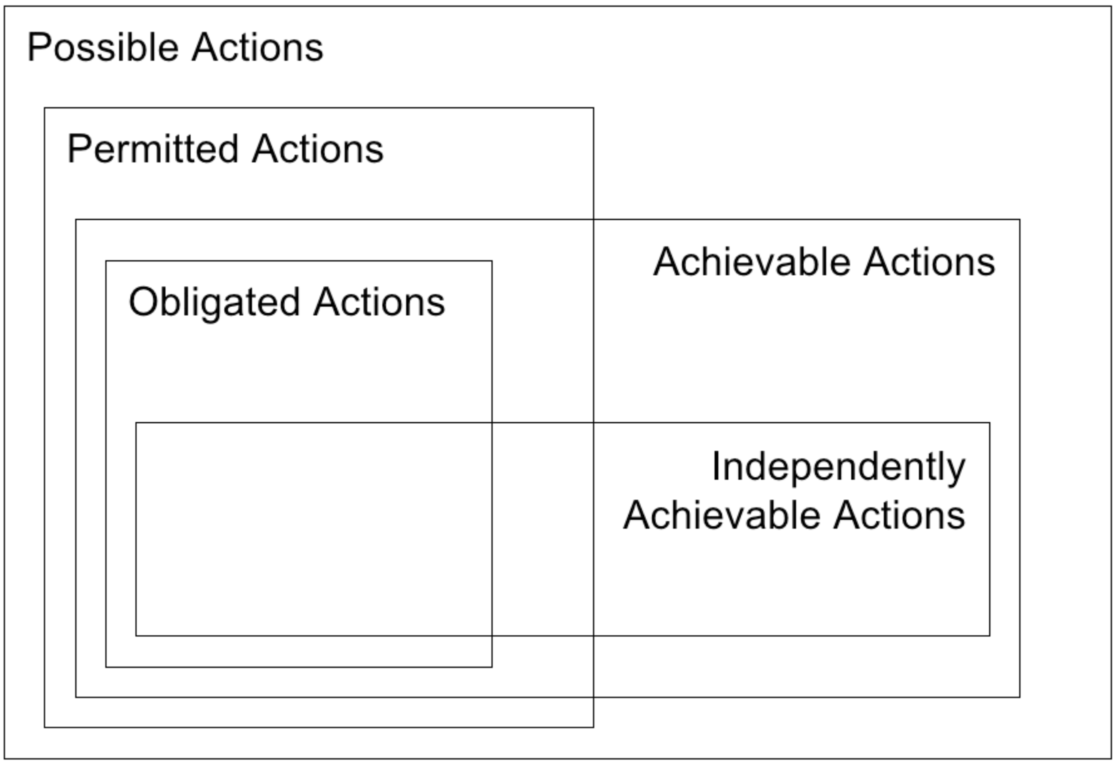

Figure 1. Degrees of autonomy corresponding to varying nested ranges of action available to an agent [13].

Humans and agents may play mutual roles that vary according to the relative degree of initiative appropriate to a given situation (Figure 2).[9] At the one extreme, traditional systems are designed to

[7] We refer to actions instead of goals because in our approach we think of goals as (roughly) a form of abstract action.

[8] We can make a rough comparison between these dimensions and many of the aspects of autonomy described by Falcone and Castelfranchi [35]. Environmental autonomy can be expressed in terms of the possible actions available to the agent—the more the behavior is wholly deterministic in the presence of a fixed set of environmental inputs, the smaller the range of possible actions available to the agent. The aspect of self-sufficiency in social autonomy relates to the ranges of what can be achieved independently vs. in concert with others; deontic autonomy corresponds to the range of permissions and obligations that govern the agent's choice among actions.

[9] For a more fine-grained presentation of a continuum of control between humans and machines, see Hancock and Scallen's [49] summary of Sheridan's ten-level formulation. Barber *et al.* differentiate three kinds of relationships among agents:

carry out the explicit commands of humans with no ability to ignore orders (i.e., *executive* autonomy [7; 35]), generate their own goals (i.e., *goal* autonomy [35; 61]), or otherwise act independently of environmental stimuli (i.e., *environmental* autonomy [20; 35]). Such systems cannot, in any significant sense, *act*; they can only be *acted upon.* At the other end of the spectrum is an imagined extreme in which agents would control the actions of humans.[10] Between these two extremes is the domain of today's agent systems, with most agents typically playing fixed roles as servants, assistants, associates, or guides. Such autonomous systems are designed with fixed assumptions about what degree of initiative is appropriate to their tasks. They execute their instructions without considering that the optimal level of autonomy may vary by task and over time, or that unforeseen events may prompt a need for either the human or the agent to take more control. At the limit of this extreme are strong, silent systems [73] with only two modes: fully automatic and fully manual. In practice this can lead to situations of human "underload," with the human having very little to do when things are going along as planned, followed by situations of human "overload," when extreme demands may be placed on the human in the case of agent failure.

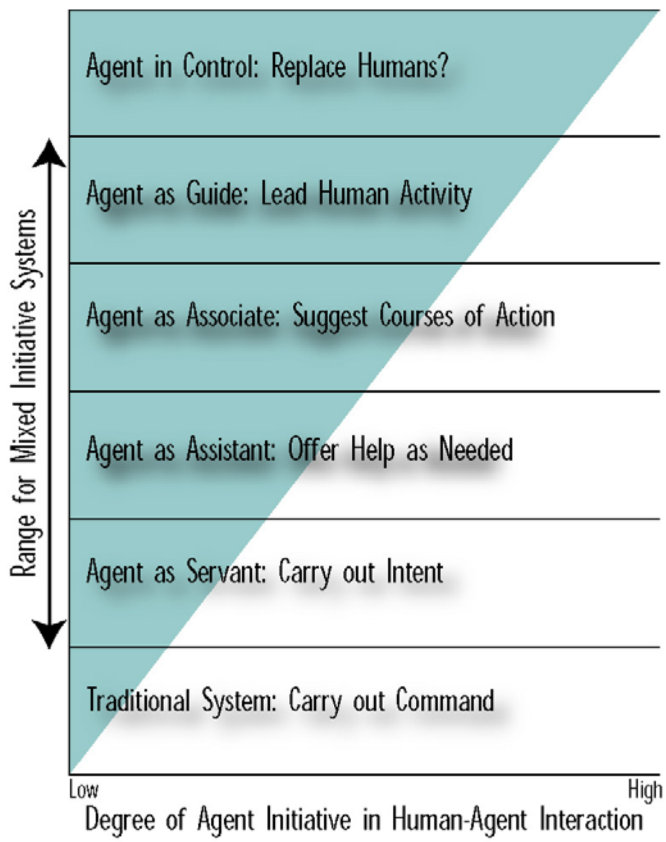

Figure 2. Spectrum of agent roles in human-agent interaction [13].

command-driven (i.e., the agent is fully subordinated to some other agent), true consensus (i.e., decision-making control is shared equally with other agents), and locally autonomous/master (i.e., the agent makes decisions without consulting other agents and may be allowed to command subordinates) [6].

[10] Of course, in real systems, the relative degree of initiative that could be reasonably taken by an agent or human would not be a global property, but rather relative to particular functions that one or the other was currently assuming in some context of joint work (see [7; 11; 46; 49].

Although in practice many do not live up to their billing, the design goal of mixed-initiative systems is to allow agents to dynamically and flexibly assume a range of roles depending on the task to be performed and the current situation [22; 29; 37]. Research in adjustable autonomy supports this goal through the development of an understanding of how to ensure that, in a given context, the agents are operating at an optimal boundary between the initiative of the human and that of the agents. People want to maintain that boundary at a sweet spot in the tradeoff curve that minimizes their need to attend to interaction with the agent while providing them with a sufficiently comfortable level of assurance that nothing will go wrong.

In principle, the actual adjustment of an agent's level of autonomy could be initiated either by a human, an agent, or some third party.[11] There are several dimensions of this level that can be varied such as: 1) type or complexity of tasks or functions it is permitted to execute, 2) which of its functions or tasks may be autonomously controlled, 3) circumstances under which the agent will override manual control, 4) duration of autonomous operation, 5) the circumstances under which a human may be interrupted (or must be interrupted) in order to provide guidance [33].

Question	Adaptive Response
Who	IF: A human performs within predetermined criteria THEN: The human shall keep task control, otherwise the task is allocated to a capable agent, if one exists.
What	IF: Only parts of tasks are being performed poorly THEN: Only these parts shall become available for dynamic allocation.
When	IF: Certain time periods are associated with increased demand, error, or loss of situation awareness THEN: These periods will be appropriate for dynamic allocation.
Where	IF: Particular environments or combinations of environmental variables are associated with increased task demand or error THEN: Encountering these environments triggers dynamic allocation.
Why	IF: Extended periods of allocation have detrimental effects (objective or subjective) THEN: Allocation shall periodically return control to the human.
How	IF: Human performance, environmental attributes, and psycho-physiological indexes are paramount for human-agent interaction THEN: All of these are inputs for allocation shift.

Table 2. Examples of the types of questions and adaptive responses to be addressed in adaptive allocation (adapted from [49, p. 526])

To the extent we can adjust agent autonomy with reasonable dynamism (ideally allowing handoffs of control among team members to occur anytime) and with a sufficiently fine-grained

[11] Cohen [30] draws a line between those approaches in which the agent itself wholly determines the mode of interaction with humans (mixed-initiative) and those where this determination is imposed externally (adjustable autonomy). Additionally, mixed-initiative systems are considered by Cohen to generally consist of a single user and a single agent. However, it is clear that these two approaches are not mutually exclusive and that, in an ideal world, agents would be capable of both reasoning about when and how to initiate interaction with the human and also of subjecting themselves to the external direction of whatever set of explicit authorization and obligation policies were currently in force to govern that interaction. Additionally, there is no reason to limit the notion of "mixed initiative" systems to the single agent-single human case. Hence we prefer to think of mixed-initiative systems as being those systems that are capable of making context-appropriate adjustments to their level of social autonomy (i.e., their level or mode of engagement with the human), whether a given adjustment is made as a result of reasoning internal to the agent or due to externally-imposed policy-based constraints.

range of levels, teamwork mechanisms can flexibly renegotiate roles and tasks among humans and agents as needed when new opportunities arise or when breakdowns occur. Such adjustments can also be anticipatory when agents are capable of predicting the relevant events [9; 35]. Research in adaptive function allocation—the dynamic assignment of tasks among humans and machines—provides some useful lessons for implementations of adjustable autonomy in intelligent systems. Examples of the types of questions and adaptive responses that might be addressed in adaptive allocation are shown in Table 2.

When evaluating options for adaptively reallocating tasks among team members, it must be remembered that dynamic role adjustment comes at a cost. Measures of expected utility can be used to evaluate the tradeoffs involved in potentially interrupting the ongoing activities of agents and humans in such situations to communicate, coordinate, and reallocate responsibilities [30; 53; 54]. It is also important to note that the need for adjustments may cascade in complex fashion: interaction may be spread across many potentially-distributed agents and humans who act in multiply-connected interaction loops. For this reason, adjustable autonomy may involve not merely a simple shift in roles among a human-agent pair, but rather the distribution of dynamic demands across many coordinated actors.[12] Defining explicit policies for the transfer of control among team members and for the resultant required modifications to coordination constraints can prove useful in managing such complexity [74]. Whereas goal adoption and the commitment to join and interact in a prescribed manner with a team sometimes occurred as part of a single act in early teamwork formulations, researchers are increasingly realizing the advantages of allowing the respective acts of goal adoption, commitment to work jointly with a team, and the choice of specific task execution strategies to be handled with some degree of independence [6; 66].

A major challenge is to ensure that the degree of autonomy is continuously and transparently adjusted to be consistent with explicitly declared policies that can ideally be imposed and removed at any time as appropriate [19; 66]. In simple terms, the goal of the agent or external entity performing such adjustments should be to make sure that the range of permissible actions do not exceed the range of those that are likely to be achievable by the agent (see figure 1).[13] While the agent is constrained to operate within whatever deontic bounds on autonomy are currently enforced as authorization and obligation policies, it is otherwise free to act. Thus, the coupling of autonomy with policy gives the agent maximum opportunity for local adaptation to unforeseen problems and opportunities while assuring humans that agent behavior will be kept within desired bounds.

2.4 The Gulfs of Evaluation and Execution

A further challenge to effective human-agent interaction is that the agents necessarily interpose a level of indirectness between ourselves and actions in the world. This indirectness can often lead to

[12] As Hancock and Scallen [49] rightfully observe, the problem of adaptive function allocation is not merely one of efficiency or technical elegance. Economic factors (e.g., can the task be more inexpensively performed by humans, agents, or some combination?), political and cultural factors (e.g., is it acceptable for agents to perform tasks traditionally assigned to humans?), or personal and moral considerations (e.g., is a given task enjoyable and challenging vs. boring and mind-numbing for the human?) are also essential considerations.

[13] If the range of achievable actions for an agent is found to be too restricted, it can, in principle, be increased in any combination of four ways: 1. removal of some portion of the environmental constraints, thus increasing the range of *possible actions*; 2. adding internal capabilities to the agent, thus increasing the range of *independently achievable actions*; 3. making additional external help available to the agent, thus increasing the overall range of *achievable actions*; or 4. reducing an agent's current set of *obligations*, thus freeing it to make other choices and to perform other tasks (see section 3.4.1 below). Of course, there is a cost in computational complexity to increasing the range of actions that must be considered by an agent—hence the judicious use of policy where certain actions can either be precluded from consideration or obligated with confidence in advance by a third party.

situations where we are misled in our expectations about the state of the world and the effects of our actions:

> "The gulfs of execution and evaluation refer to the mismatch between our internal goals and expectations and the availability and representation of information about the state of the world and how it might be changed [figure 3]. The gulf of execution refers to the difficulty of acting upon the environment (and how well the [agent] supports those actions). The gulf of evaluation refers to the difficulty of assessing the state of the environment (and how well the [agent] supports the detection and interpretation of that state)… We can conceptualize the [agent] and its interface in this way. A person is a system with an active, internal representation. For an [agent] to be usable, the surface representation must correspond to something that is interpretable by the person, and the operations required to modify the information within the [agent] must be performable by the user. The interface serves to transform the properties of the [agent's] representational system to those that match the properties of the person." [68][14]

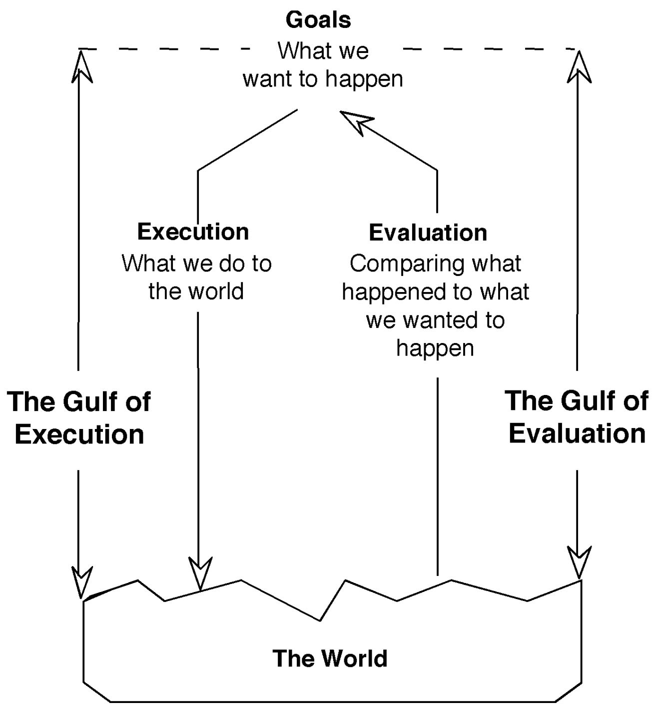

Figure 3. The gulfs of execution and evaluation [68].

In brief, people need to understand what is happening and why when a teammate tends to respond in a certain way; they need to be able to control the actions of an agent even when it does not always wait for the human's input before it makes a move; and they need to be able to reliably predict what will happen, even though the agent may alter its responses over time [34]. This problem is more

[14] We do not wish to imply that we are here taking a stance that the world is presented to us directly. Rather, as George Kelly elaborated in his principle of constructive alternativism, "'reality' does not reveal itself to us directly, but rather is subject to as many different constructions as we are able to invent" [17, p. 288]. Such considerations of the fluidity of meaning and interpretations are only recently being addressed within human-agent interaction research.

pronounced in sophisticated agents that have a model of their teammates and the environment than with relatively more passive conventional programs because in agent applications these gulfs pose a problem not only for the human acting on the agent and its world, but also for the agent trying to act on the humans and their world. Humans and agents must be aware of what team members are doing, why they are doing it, and where it is on the team member's current "agenda." Table 3 below gives examples of the kinds of things that humans and agents need to be able to know in teamwork settings.

Questions About the Shared Representation of the Problem State	**Questions About the Representation of the Activities of Humans and Agents**
What type of problem is it? Is the problem routine or difficult? Is the problem high or low priority? What types of solution strategies are appropriate? What dependencies must be considered? How is the problem state evolving?	How did we get into this state? What are they doing now? Why are they doing it? Are they having difficulties? Why? What are they doing to cope with difficulties? Are they likely to fail? How long will they be busy? What will they do next?

Table 3. Examples of the kinds of things that humans and agents need to be able to know in teamwork settings (adapted and extended from [24; 83; 85]).

Good designers help manage the problem of the gulf of evaluation by providing an understandable and controllable level of feedback about the agent's intentions and actions. They must also think about how to accurately convey the agent's capabilities and limitations so that people are not misled in their expectations. Part of the problem is the natural enthusiasm of agent researchers; part of the problem is people's tendency to falsely anthropomorphize.[15] Though we can carefully describe agent capabilities and limitations within accompanying instructional manuals, it is even more important to find clever ways to subtly weave this information into the agent interface itself.

Overcoming the gulf of execution requires providing people with straightforward means to accurately convey their intentions and requests to agents. And if the agent is not currently capable of performing a desired action, how do ordinary people program the agent to do what they want? While programming-by-demonstration or simplified graphical or scripting languages have been suggested, none of them yet seem adequate to specify the kinds of complex tasks envisioned for future intelligent agents [31; 59; 76].[16]

15 Fortunately, people have a lot of experience in judging the limitations of those with whom they communicate: "Sometimes people overstate what the computer can do, but what people are extremely good at is figuring out what they can get away with. Children can size up a substitute teacher in about five minutes" [56]. For evidence that developers of intelligent software are no less prone than other people to overestimate the capabilities of their programs, see McDermott [63].

16 *Automatic programming* is an enterprise with a long history of insatiable requirements and moving expectations. For example, Rich and Waters [72] remind us that "compared to programming in machine code, assemblers represented a spectacular level of automation. Moreover, FORTRAN was arguably a greater step forward than anything that has happened since. In particular, it dramatically increased the number of scientific end users who could operate computers without having to hire a programmer." Today, no one would call FORTRAN a form of automatic programming, though in 1958 the term was quite appropriate. The intractability of fully-automated, completely-general programming is analogous to the problem of automated knowledge acquisition [17; 39]. As Sowa observes: "Fully automated knowledge acquisition is as difficult as unrestricted natural language understanding. The two problems, in fact, are different aspects of exactly the same problem: the task of building a formal model for some real world system on the basis of informal descriptions in ordinary language. Alan Perlis once made a remark that characterizes that difficulty: *You can't translate informal specifications into formal specifications by any formal algorithm.*" [77].

The discovery of means for humans to more effectively overcome these gulfs with agents—and vice versa—is a prime focus of research on *mediating representations* [39]. The choice of representation can have an enormous effect on human problem solving performance (e.g., [45; 57]. As a simple example, consider the impact that representing numbers in binary, Arabic, or Roman numeral form would have on the ability of humans or agents to efficiently multiply. As another example, concrete or abstract diagrams can be a particularly powerful form of knowledge representation for humans because they allow the explicit depiction of relevant information and effective hiding of inessential features which otherwise would have to be done at the expense of large amounts of cognitive work.[17] Moreover, the appropriate mode of interaction needs to be considered in the context of the task, physical environment, and the individual preferences and capabilities of the human under varying conditions of cognitive load. When required, interruptions of the human by an agent must be done judiciously—with no more than the just the necessary degree of obtrusiveness. In addition, as the level of sophistication of communication between agents and humans increases, future forms of human-agent collaboration must enable people and agents to negotiate and work with appropriate tools for shaping shared understanding.

Erickson [34] raises a related concern when he argues that designers ought to take advantage of the ontological expectations that users bring with them when they interact with various portrayals of functionality in graphical user interfaces. For example, specific computing functionality can be portrayed as an object or an agent, depending on what is most natural. The desktop metaphor takes advantage of users' previous knowledge that office artifacts are visible, are passive, have locations, and may contain things. "Objects stay where they are: nice, safe predictable things that just sit there and hold things" [34, p. 94]. Ontological knowledge of a different sort comes into play when the agent metaphor is employed. Our common sense knowledge of what agents can do tells us that, unlike typical desktop objects, they can notice things, carry out actions, know and learn things, and go places.[18] "Agents become the repositories for adaptive functionality" [34, p. 94]. Erickson concludes that research "which focuses on the portrayal of adaptive functionality, rather than [solely] on the functionality itself, is a crucial need if we wish to design agents that interact gracefully with their users" [34, p. 95].

2.5 The Substitution Myth

A persistent misperception about all forms of automation is the notion that such assistance is a simple multiplier of human capability. Such a view is natural because, from the point of view of an outsider observing the assisted human, it seems that—in the successful cases at least—the person is able to perform the task better or faster than he or she could without help. In reality, however, help of whatever kind does not simply enhance our ability to perform the task: it changes the nature of the task itself [68]. Those who have had a five-year-old child offer to help them with the dishes know this to be true—from the point of view of an adult, such "help" does not necessarily diminish the effort involved, it merely effects a transformation of the work from the physical action of washing the dishes to the cognitive task of monitoring the progress (and regress) of the child.

[17] "Chernoff faces," which leverage the efficiency of human facial recognition into the domain of statistical analysis, are an ideal exemplar of the kind of integrative pattern-based feedback that works well for people [23]. Another outstanding example is the OZ cockpit display [78].

[18] It is also easy for people to assume less tangible qualities about agents, e.g., that they are internally consistent, are rational, act in good faith, can introspect, can cooperate to achieve common goals, and have a persistent mental state. Obviously, agent designers need to work hard to make sure that human expectation accords with the stark facts of reality in each of these dimensions. Violated trust inevitably breeds user hostility.

Ignorance of such considerations leads to what Wiener [83] called "clumsy automation" and what Christoffersen and Woods [24] term the "substitution myth": the erroneous notion that "automation activities simply can be substituted for human activities without otherwise affecting the operation of the system."[19] In refutation of the substitution myth, Table 4 contrasts the putative benefits of automated assistance with the results of empirical study. Ironically, even when technology succeeds in making tasks more efficient, the human workload is not reduced accordingly. As noted by many researchers and summarized by Woods as the *law of stretched systems:* "every system is stretched to operate at its capacity; as soon as there is some improvement, for example in the form of new technology, it will be exploited to achieve a new intensity and tempo of activity" [84].

Putative Benefit	Real Complexity
Better results are obtained from "substitution" of machine activity for human activity.	Transforms practice; the roles of people change; old and sometimes beloved habits and familiar features are altered—the *envisioned world problem.*
Frees up human by offloading work to the machine.	Creates new kinds of cognitive work for the human, often at the wrong times; every automation advance will be exploited to require people to do more, do it faster, or in more complex ways—the *law of stretched systems.*
Frees up limited attention by focusing human on the correct answer.	Creates more threads to track; makes it harder for people to remain aware of and integrate all of the activities and changes around them.
Less human knowledge is required.	New knowledge and skill demands are imposed on the human.
Agent will function autonomously.	Team play with people is critical to success.
Same feedback to human will be required.	New levels and types of feedback are needed to support peoples' new roles.
Agent enables more flexibility to the system in a generic way.	Resulting explosion of features, options, and modes creates new demands, types of errors, and paths toward failure—*automation surprises.*
Human errors are reduced.	Both agents and people are fallible; new problems are associated with human-agent coordination breakdowns.

Table 4. Putative benefits of automation vs. actual experience (adapted and expanded from [73; 84]).

Notwithstanding these challenges, adult humans and radically less-abled entities (e.g., children, dogs, video game characters) are capable of working together effectively in a variety of situations where a subjective experience of collaborative teaming is often maintained despite the magnitude of their differences. Generally this is due to the ability of humans to rapidly size up and adapt to the limitations of their teammates in relatively short order. More study is needed to understand how to take advantage of these human abilities in a general way to make human-agent interaction more natural and effective in complex agent applications. As with all automation, the introduction of agents into human work practices, particularly agents who do not yet generally exhibit the intelligence of a five-year old child, must be done carefully to ensure that the cost of the coordination and monitoring demands on the human do not exceed the value of the agent assistance offered [30].

[19] The substitution myth is one example of what Feltovich and colleagues [36] have termed the "reductive bias": the tendency for designers and operators to oversimplify complex phenomena, often with detrimental effects.

3. WORK-PRACTICE MODELING AND HUMAN-AGENT COLLABORATION ONBOARD THE INTERNATIONAL SPACE STATION

In this section we describe our efforts to apply the principles of human-centered teamwork to the design of robotic agents appropriate for use aboard the International Space Station (ISS). Following a discussion of the background of the project (section 3.1), we describe our progress and lessons learned to date in using Brahms work practice modeling and simulation approaches (section 3.2). We briefly discuss the implications of our preliminary studies of teamwork in space applications (section 3.3). We then show how Brahms capabilities are combined with KAoS services for analysis and enforcement of policies for teamwork and adjustable autonomy, and NOMADS strong mobility and fine-grained resource management capabilities (section 3.4).

3.1 Background and Motivation

Enhancing the crew’s ability to perform their duties is critical for successful, productive, and safe space operations. Crew time on such missions is a precious resource. The limited number of crew members are required to maintain complex systems, assist with life-critical environmental health monitoring and regulation, perform dozens of major simultaneous payload experiments, and perform general housekeeping. As one example, consider the challenges of Shuttle Mission 89’s flight in January, 1998:

> “*One* astronaut, Andy Thomas, will undertake *several hundred* research runs involving *26* different science projects in *five* disciplines. The projects are provided by *33* principal investigators from the U.S., Canada, Germany and the U.K” (Yuri Gawdiak, personal communication).

Other considerations also are important motivators for efficient human teaming and some degree of robotic assistance. Considerable time and tedium, as well as long- and short-term risks, are involved in human EVA activity. Safety considerations and size constraints are also important issues for many manned mission activities, especially in emergency situations. Even if it were physically possible for an astronaut to enter congested spacecraft areas, protruding debris or other environmental hazards of one kind or another could pose serious safety risks [21].

QuickTime™ and a TIFF (LZW) decompressor are needed to see this picture.

Figure 4. The Personal Satellite Assistant (image courtesy Greg Dorais).

One example of a system being designed to support such activities is the Personal Satellite Assistant (PSA), a softball-sized flying robot designed to operate onboard spacecraft in pressurized micro-gravity environments [42] (figure 4). The PSA will incorporate environmental sensors for gas, temperature, and fire detection, providing the ability for the PSA to monitor spacecraft, payload and crew conditions. Video and audio interfaces will support navigation, remote monitoring, and video-conferencing. Ducted fans will provide propulsion and batteries will provide portable power.

Other examples of such systems involving very different kinds of interaction with humans are Robonaut, a dexterous robot with torso, head, arms, and five-fingered hands designed for fine movement [5] (figure 5), and Mini-AERCam, a flying spherical "eye" for use outside the ISS in multi-robot multi-person EVA scenarios. Robots of various forms will also support humans performing various forms of surface exploration on interplanetary expeditions.

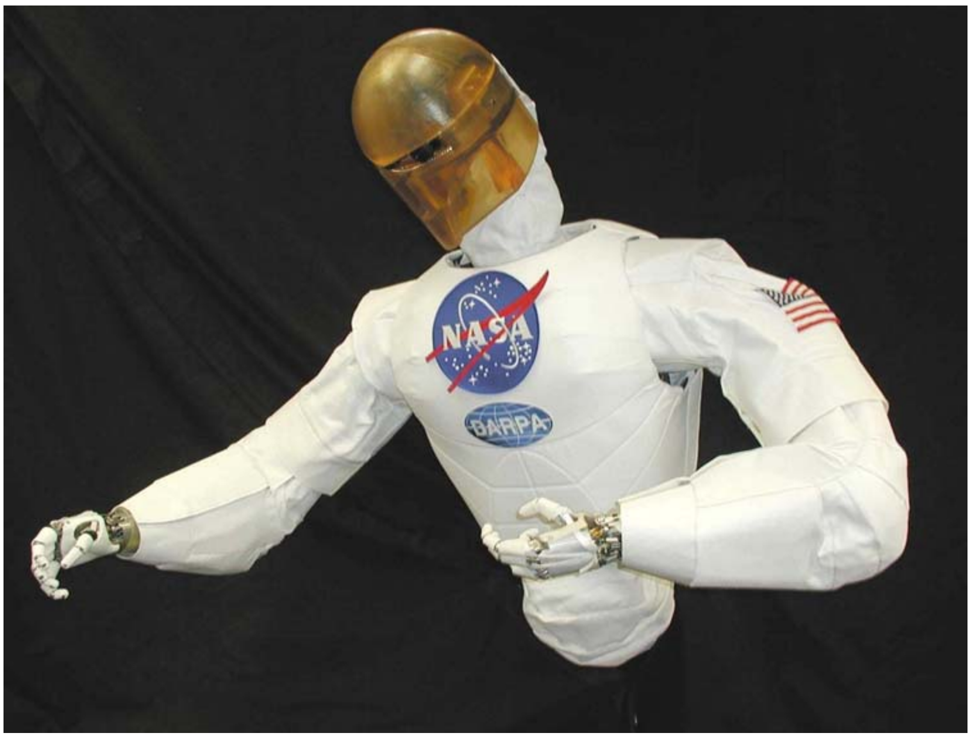

Figure 5. Robonaut is a dexterous robot with torso, head, arms, and five-fingered hands designed for fine movement [5].

To support the objective of efficient and effective space operations, we are midway through a three-year study to show whether a model of human-robot collaboration implemented in Brahms, KAoS, and NOMADS can be used not only as a design tool to understand human-robotic interaction, but also in conjunction with agents in the execution environment. The overall plan of the study is shown in figure 6. In the next sections, we discuss our efforts to date to use Brahms in performing an astronaut work practice study and work systems design for use of collaborative robots onboard the ISS (section 3.2). We then share some of our current efforts to extend teamwork theory consistent with the results of our modeling, with behavioral science studies, and with previous research on human-centered teamwork (section 3.3). Finally, we describe our progress to date on integrating Brahms, KAoS, and NOMADS in preparation for the experimental phase of the study (section 3.4).

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

Figure 6. Plan for study of human-robot collaboration on the ISS.

3.2 Modeling Life Onboard The International Space Station

3.2.1 Brahms

Brahms is a language and modeling and simulation environment based on the idea of situated action [25; 79] and offers to the researcher a tool to represent and study the richness of activity theory and work practice [26]. A traditional task or functional analysis of work leaves out informal logistics, especially how environmental conditions come to be detected and how problems are resolved. Without consideration of these factors, analysts cannot accurately model how work and information actually flow, nor can they properly design software agents that help automate human tasks or interact with people as their collaborators. For these goals, what is needed is a model that includes aspects of reasoning found in an information-processing model, plus aspects of geography, agent movement, and physical changes to the environment found in a multi-agent simulation – interruptions, coordination, impasses, and so on. A model of work practice focuses on informal behavior in specific locations and circumstances. It emphasizes the behavior by which synchronization occurs (such that the task contributions of humans and machines flow together to accomplish goals) and allows the researcher to capture much of the richness of activity theory.

Brahms is based on an Agent-Oriented Language (AOL) with a well-defined syntax and semantics. A Brahms model can be used to simulate human-machine systems, for what-if experiments, for training, user modeling, or driving intelligent assistants and robots. For a full description of Brahms, the reader is referred to [75]. The run-time component—the Brahms virtual machine—can execute a Brahms model, also referred to as a simulation run (figure 7).

The Brahms architecture is organized around the following representational constructs:

Groups of groups containing
  Agents who are located and have
    Beliefs that lead them to engage in
      Activities specified by
        Workframes
Workframes consist of
  Preconditions of beliefs that lead to
    Actions, consisting of
      Communication Actions
      Movement actions
      Primitive Actions
      Other composite activities
          Consequences of new beliefs and facts
    Thoughtframes that consist of
      Preconditions and
      Consequences

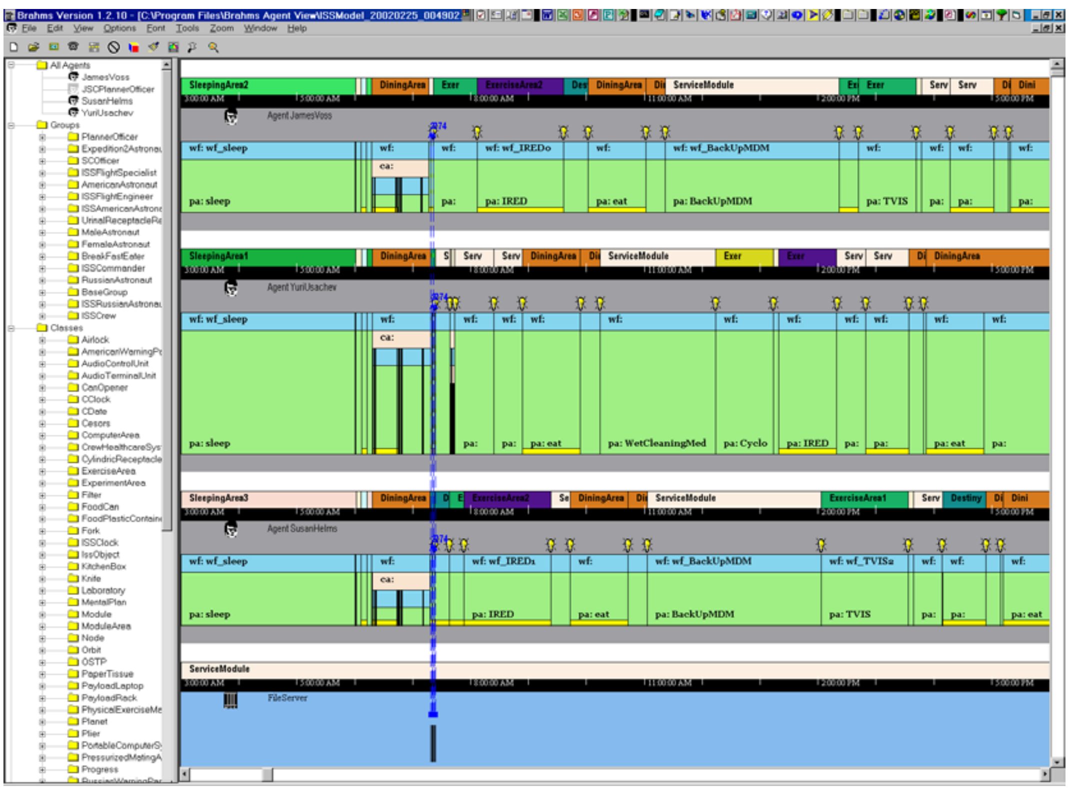

Figure 7. A day in the life of the ISS crew: 2D output of a Brahms simulation

Brahms relates knowledge-based models of cognition (e.g., task models) with discrete simulation and the behavior-based subsumption architecture. In Brahms, agents' behaviors are organized into activities, inherited from groups to which agents belong. Most importantly, activities associate behaviors of people and their tools with particular times and locations, so that resource availability and informal human participation can be taken into account. A model of activities doesn't necessarily describe the intricate details of reasoning or calculation, but instead brings out aspects of the social-physical context in which reasoning occurs. Thus Brahms differs from other simulation systems by incorporating:

- Activities of multiple agents located in time and space;
- Conversations;
- Descriptions of how information is represented, transformed, reinterpreted in various physical modalities.

3.2.2 A Human-Centered Perspective on Autonomous Systems Design

Consistent with our emphasis on understanding teamwork in practice, our project necessarily began with a detailed study of how the astronauts actually work onboard the ISS. Our first step has been to develop a Brahms model of the daily work practice of the crew. Secondly, we would extend that model to include various PSA and Robonaut use scenarios. Finally, to support real-time operational teamwork capabilities for these robots, we would enhance the Brahms model and software to run in conjunction with KAoS policies and agent services, and NOMADS capabilities (see section 3.5 below).

The objective of studying crew work practice is threefold: 1) to study how actual crew work practices emerge from planned activities and written procedures; 2) to discover opportunities for collaborative robots such as the PSA to assist with the crew's work; and 3) to learn more about how teamwork actually occurs in practice. A more complete account of the Brahms modeling effort to date may be found in [1].

3.2.3 A Day in the Life

In a typical day, each ISS crewmember divides his or her time between physical exercise, maintenance, experiments, communication with ground personnel, unstructured time, and essential personal activities (e.g., rest, eating). These activities are critical for the well-being of the crew. Hence, the planned maintenance and research activities must be scheduled around them. At the same time, several interdependent structural constraints ensure crew safety and productivity: thermal control, power management, communication bandwidth management, and regulation of other systems. These form a network of components that must be accurately timed and orchestrated around crew activities and needs.

An elaborate planning process is necessary to meet the human and system constraints. The planning complexity is such that the major planning rule for the ISS is actually: "Thou shalt not replan" (see [67], p. 2.1-21). This means that—with the exception of "unacceptable failures" and "job jar" items left to the self-organization of the crew—any activity that can not be performed at its allotted time will *not* be replanned in real-time. It will simply not be performed, "with the expectation that [it] will be rescheduled into the operational flow at some later date." Such discrepancies are actually frequent, as the comparison between daily plans and actual ship logs shows.

3.2.4 Developing the Brahms Model

06:00 - 06:10 ISS morning inspection
06:10 - 06:40 Post Sleep
06:40 - 07:30 Breakfast
07:30 - 08:00 Prep for work
08:02 - 08:17 DPC via S-Band
08:30 - 09:15 (FE-1) TVIS Video Survey
08:15 - 08:30 (FE –2) SSC Daily Maintenance
08:35 - 08:50 (FE –2) MEC card swap
09:00 - 10:00 (FE –2) Physical Exercise, Active Rest
09:15 - 10:45 (FE –1) Physical Exercise, Active Rest
09:00 - 09:15 (CDR) URAGAN, visual observations
09:50 - 10:20 (CDR) Replacement of urine-receptacle in Toilet
10:00 - 10:30 (FE –2) MEC Exercise Data Downlink
10:20 - 11:00 (CDR) ECLSS maintenance by MCC GO
10:55 - 11:20 (FE –1) LAB PL Status/Monitor
11:00 - 11:30 (CDR) SOYUZ Window Inspection
11:30 - 11:50 7A TAGUP via S-Band
12:00 - 13:00 LUNCH
13:05 - 13:25 WPC via S-Band
13:30 - 15:30 (CDR) Wet Cleaning/ ODF Medical Support
13:30 -16:30 (FE -1, FE –2) Back Up MDM S/U
16:15 - 17:15 (CDR) Physical Exercises, VELO-1
16:55 -17:15 (FE –1) Prep of Delta File (IMS)
16:45 - 18:15 (FE –2) Physical Exercise, Active Rest
17:15 - 18:15 (FE –1) Physical Exercise, Active Rest
17:15 - 18:15 (CDR) Physical Exercise. RED-1
18:15 - 18:45 Familiarization with next day's plan
18:45 - 19:30 Prep of Report
19:05 - 19:20 DPC via S-Band
19:30 - 19:55 Dinner

Table 5. A Day in the Life of the ISS Crew (Derived from the onboard plan for May 7th, 2001 uploaded to the ISS)

We sought data that could lead us to understand and represent a generic “day in the life” of the ISS crew. However, we also dedicated more attention to specific activities and scenarios (such as emergency scenarios) that appeared of great relevance to our research objectives. We consulted ISS documentation and manuals, onboard procedures and flight rules, crew daily plans and ship logs, crew debriefings, and, particularly, ISS crew videos. This information was interpreted, analyzed, and validated through interviews with astronauts, astronaut trainers, and flight controllers at Mission Control. While the day chosen for modeling was May 7th, 2001 (there was a relatively large amount of data available for that day), we generalized the model so that we could later simulate any typical day.

In our analysis of the data gathered during the first phase of our research we looked for patterns in the crew activities and the emergence of work practices that are specific to onboard life. We generalized and represented the individual astronaut’s daily behavioral patterns as learned and shared activities at the informal group level. For example, the activity of eating breakfast is represented in the Brahms ISS model at the Crew group-level. This way, all agents that are a member of the ISS Crew "know" how to perform this activity. The group structure also allows us to represent differences between social, cultural and other type of communities (for example, the behavioral differences between American and Russian crewmembers, and between male and female crewmembers).

In order to make our model reusable and applicable to any typical day and scenario on the ISS, we represented procedures, daily plans, and flight rules as physical and conceptual objects in the model, that agents can access, have beliefs about, manipulate, and act upon. We categorized activities according to a 2-by-2 matrix, with the degree in which the activity was scheduled (scheduled vs. unscheduled activities) represented on one axis, and the uniqueness or repeatability (day-specific vs. recurrent activities) of the activity represented on the other axis (see Table 6). This allowed us to model elements of the crew's situated action by letting the crew agents perform a just-in-time replanning activity[20] through which they change their mental plan—that was first constrained by the OSTP and coordinated with Mission Control—during the day, based on the context of the day's activities.

	Scheduled activity	*Unscheduled activity*
Day-specific activity	Maintenance activities (e.g., Replacement of urine-receptacle in Toilet). Experiments (e.g., LAB PL Status/Monitor). …	Emergencies. Job-Jar activities Unexpected maintenance or repair activities. …
Recurrent activity	Physical exercise. Daily Planning Conference. Eating (lunch, dinner, breakfast) …	Going to the toilet. Sending personal email. …

Table 6. A 2 by 2 matrix of sample ISS activities

Discrepancies between plan and reality come out of the simulation of the model, as they do during the actual day of the astronauts. While procedures and scheduled activities suggest a certain idealized scenario, several issues can emerge: procedures might not be clear, American and Russian versions of the same document might differ slightly and thereby generate confusion, the time needed to complete an activity might be longer than expected, tools might get lost, and, significantly, new work practices might emerge.

The discrepancies we refer to are not only those caused by imprecise timing of new activities, or triggered by unforeseeable error and mismatches with systems or procedures.[21] As Table 7 shows, we also focus on more substantial discrepancies that involve a deliberate (though possibly not planned-in-advance) behavior of the crew. For example, it is often the case that "highly motivated crew members sacrifice personal time to 'get the job done'" [70]. Traditional approaches do not take into consideration any of these kinds of unplanned events. More importantly, they rarely deal with the concatenated effects caused by the highlighted discrepancies. For example, the activity of printing out a procedure rather then reading it on a laptop implies that the astronauts must move to the printer location; it also implies that paper must be available, otherwise new paper must be fed into the

[20] We do not suggest that astronauts perform this activity by executing a computational algorithm similar to artificial intelligence planning systems. We rather represent the astronaut's ability to change the order they decide to perform their activities, based on situational awareness and context.

[21] In this regard, Expedition 2 reported a substantial improvement with respect to Expedition 1 in the accuracy of the predicted duration of scheduled activities and in the feasibility of the planned daily workload.

printer. In contrast to typical planning approaches, the Brahms simulation is capable of showing how the practice of onboard activities often diverges, both in timing and execution, from the originally-scheduled activities and procedures. Distances and movements, noises, tools location, work practice, and so forth are considered. Hence, delays caused by crew movement constraints, the search for tools and other items, and the inability to share resources or access to electronic procedures can be revealed by the simulation. For example, we model the fact that the work practice of the astronauts is to move from one module to the other to communicate face-to-face, rather than using the internal audio system.

Cause of Discrepancy	**Example**	
	Plan	*Practice*
Procedures are not easily accessible or are not clear	During emergency, refer to procedure	During emergency, rely on training and memory
Noise level on internal audio system	Use internal audio system to communicate between modules	Move from module to module to communicate with crew members
Personal preferences	Do medical tests as scheduled	Do medical tests in the morning
Shared resources are not always available	Upload on laptop computer medical/physical data after experiment/exercise	Upload data rarely
Personal habits	Read procedure	Read electronic procedure (from laptop), or read printed procedure
Inventory system is not always reliable	Use the tools indicated in procedure	Tools must be found and time can be lost in this operation
Inventory system is not always reliable	Use the bar-code reader for inventory	Rarely use bar-code reader

Table 7. Discrepancies between plan and practice.[22]

3.3 Extending Teamwork Theory

In extending traditional teamwork theory [28; 82], we seek to incorporate the best of previous research on human-centered collaboration and teamwork, while simultaneously grounding new findings in our own work practice study experience. In addition to surveying studies of multi-agent teamwork and human-centered perspectives cited above, we are assessing the contributions of allied fields ranging from cognitive function analysis [10], to studies of animal signaling and cooperation, the roles of values and affect, and the enablers of effective delegation in humans [64].

[22] Sources: ISS Ship logs; ISS debriefs; interviews with ISS training specialists.

From our preliminary studies of human teamwork to date we have realized that the longstanding emphasis in teamwork research on "joint goals" and fixed task-specific "roles" has previously overshadowed other important aspects of teamwork such as shared context, awareness, identity, and history. Unlike the relatively rigid joint intentions of typical agent teamwork models, experience in work practice underscores the importance of conceptualizing agreements among team members as things that are forever tentative and subject to ongoing negotiation. An interesting finding has been that some of the most important settings for the coordination of teamwork are not in formal planning meetings but around the breakfast table (figure 8). Such settings can also be a time for sharing knowledge gleaned from practice. To take a simple example, we observed Commander Usachev engaged vigorously spinning one of his food tins at the breakfast table—what a naïve observer might initially think was just "playing with his food" could instead prove to be a useful strategy to heat the can or to keep it from drifting off while he attends to other activities.

Figure 8. Breakfast onboard the ISS during Expedition 2 (Commander Yuri Usachev on the left).

Additionally, we expect that as we study divergences between plan and practice we will better understand how people plan and coordinate their activities in the real world. Such detail in the Brahms model has already helped us notice unexpected opportunities for robotic assistance. For example, a simple thing like one crew member having to stand and hold a flashlight for another or having to stop and look for missing tools can tie up valuable human resources. Building features to support mundane activities that waste valuable crew time into the PSA has turned out to be of great importance.

3.4 Integrating Brahms, KAoS, and NOMADS

In preparation for the experimental phase of our work, we have integrated a version of Brahms with KAoS, and are developing an approach to take advantage of NOMADS as well. These capabilities and their expected roles during the experimental phase are briefly described below.

3.4.1 KAoS

KAoS is a collection of componentized agent services compatible with several popular agent frameworks, including the DARPA CoABS Grid, DARPA the ALP/Ultra*Log Cougaar framework, and Objectspace Voyager. The adaptability of KAoS is due in large part to its pluggable infrastructure based on Sun's Java Agent Services (JAS) (http://www.java-agent.org). For a full description of KAoS, the reader is referred to [16; 18; 19].

There are two key KAoS services relevant to the effort described here: *policy services* and *domain management services.*

Policy services are used to define, manage, and enforce constraints assuring coherent, safe, effective, and natural interaction among teams of human and agents. Knowledge is represented declaratively in DAML+OIL ontologies. The current KAoS Policy Ontology (KPO) specification defines basic ontologies for actors, actions, entities that are the targets of actions (e.g., other actors, computing resources), places, policies, and policy conditions. We have extended these ontologies to represent simple atomic Java permissions, as well as more complex NOMADS, and KAoS policy constructs. The ontologies are currently being extended with additional classes, individuals, and rules to represent our model of human-agent teamwork, reflecting both the traditional concerns of team formation and maintenance as well as newer concerns about effectiveness of human-agent interaction in specific contexts of practice.

The policy ontology distinguishes between positive and negative *authorizations* (i.e., constraints that permit or forbid some action) and positive and negative *obligations* (i.e., constraints that require some action to be performed, or else serve to waive such a requirement) [9; 32]. Dynamic changes or additions to policy require logical inference to determine first of all which if any policies are in conflict and second how to resolve these conflicts [62]. We have implemented a general-purpose algorithm for policy conflict detection and harmonization whose initial results promise a relatively high degree of efficiency and scalability.[23] Figure 9 shows the three types of conflict that can currently be handled: positive vs. negative authorization (i.e., being simultaneously permitted and forbidden from performing some action), positive vs. negative obligation (i.e., being both required and not required to perform some action), and positive obligation vs. negative authorization (i.e., being required to perform a forbidden action). We use subsumption-based reasoning to allow policy conflicts to be detected and resolved even when the actors, actions, or targets of the policies are specified at vastly different levels of abstraction.

Beyond the mechanisms for assuring consistency among permissions and obligations, we are developing an approach to determine how and when to make policy changes based on adjustable autonomy considerations. To accommodate reasoning and decision-making under uncertainty in these situations, we are incorporating a reasoning component based on knowledge-based construction and evaluation of Bayesian networks and influence diagrams [4; 14; 15; 52; 55]. The results of this analysis are used to compute the expected utility of potential adjustments to the agent's autonomy. The goal is to delegate as much as appropriately can be done to the agent while continually adjusting

[23] A detailed description of the KAoS policy ontologies and our policy conflict detection and resolution process is currently being prepared for publication.

the policies in force so that the range of permissible actions does not exceed the range of those deemed as likely to be achievable. The need to maintain an appropriate level of mutual awareness of the intentions of team members and of the state of the world while minimizing unnecessary intrusiveness is also being taken into consideration in the model. While initial results are encouraging, these are challenging problems that will require many additional years of long-term research.

We are also exploring approaches to increase the agent's autonomy in a more absolute sense by automatically expanding its external capabilities so that the agent is able to take on new kinds of tasks relying on outside assistance that it could not have taken on by itself alone (or through the use of external resources of which it was previously aware). For example, one approach might be to provide support for transparent resource discovery and coordination processes. A related approach to increasing the agent's autonomy in an absolute sense is to provide support for it to efficiently reduce its obligations through various forms of facilitation, and renegotiating (or, in exceptional circumstances, reneging on) current obligations so that the agent can focus on higher priority tasks.

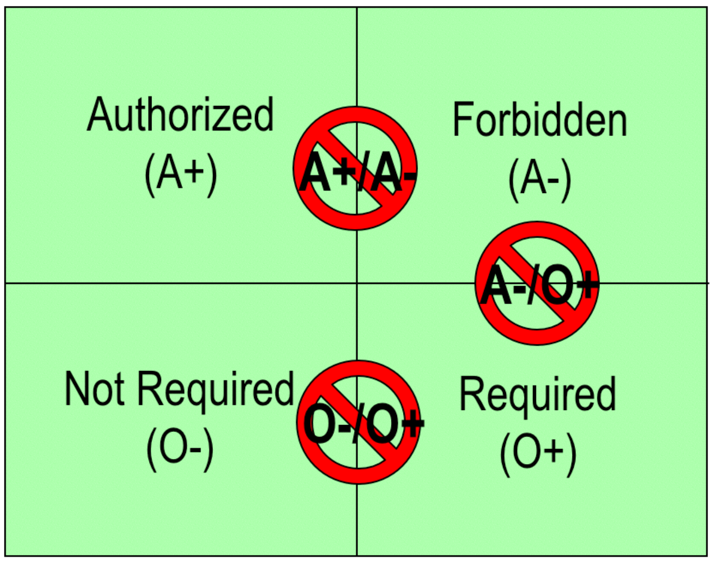

Figure 9. Three types of policy conflict are detected and resolved in KAoS.

Policy enforcement mechanisms built either into the execution environment or the agent platform aim to assure policy compliance for any agent or program running in that environment or platform, regardless of how that agent or program was written.[24] Viewing the knowledge governing the interaction of agent communities from a policy-based perspective has proven natural and effective in

[24] Our need to support heterogeneous agents and platforms precludes us from using an approach such as Myers *et al.* [66] that provides for policy enforcement by requiring all agents to adopt a common BDI framework. Scerri *et al.* [74] and Cohen *et al.* [30] differ from both Myers *et al.* and our approach in that the determination of the preferable mode of interaction with humans and other agents is wholly performed through the reasoning processes of the agents themselves rather than allowing for policy constraints to be explicitly imposed by external authority. In the context of our space applications, we see the requirement to embrace both approaches: providing a model and mechanisms for self-initiated reasoning about mixed-initiative interaction while assuring that the actions of agents are consistent with externally-specified policy constraints.

a variety of our recent agent applications (e.g., [3; 12]), and we are finding that many of the lessons learned can be straightforwardly generalized to the domain of human-agent teamwork.

Domain services are used to facilitate the structuring of agents into complex organizational structures, administrative groups, and dynamic task-oriented teams, and to provide a common point of administration and policy enforcement. Through various property restrictions, a given policy can be variously scoped, for example, either to individual agents, to agents of a given class, to agents belonging to an intensionally- or extensionally-defined domains or teams, or to agents running in a given physical place or computational environment (e.g., host, VM).

The basic approach for the integration of Brahms and KAoS can be described as follows. A Brahms world runs in parallel with the robotic execution environment, providing a work practice context when KAoS services are invoked. KAoS is in turn responsible for message transport, management and execution of conversation, adjustable autonomy, and teamwork policies among humans and agents. KAoS is also responsible for providing input from the teamwork model to general-purpose robotic planning, scheduling, and execution capabilities that will be used in our testbed experimentation. For example, we are working with Muscettola and his colleagues on an interface between KAoS and EUROPA/IDEA for this purpose [65].

3.4.2 NOMADS

NOMADS is the name we have given to the combination of Aroma, an enhanced Java-compatible Virtual Machine (VM), with its Oasis agent execution environment [80]. It is designed to provide environmental protection of two kinds:

– assurance of availability of system resources, even in the face of changing resource priorities, buggy agents or denial-of- service attacks;
– protection of agent execution state, even in the face of unanticipated system failure.

These basic capabilities of NOMADS provide essential features of reliability and safety required for interaction with humans in dynamic and demanding space environments. We are currently working with Sun Microsystems on incorporating resource management features similar to NOMADS into a future version of the commercial Java platform.

Thus, in summary, the model of work practice provided by Brahms of activities, humans, agents, and objects that are part of those activities provides needed contextual parameters that will be used to tune the constraints of the KAoS teamwork model and its operational mechanisms. The policy specification, representation, conflict resolution, and enforcement mechanisms of KAoS will assure that a coherent set of teamwork policies can be continuously in effect. The strong mobility and safe execution features of NOMADS will enable protection and optimal use of scarce onboard computing resources.

4. SUMMARY

We have outlined a preliminary perspective on the basic principles and pitfalls of adjustable autonomy and human-centered teamwork, and some of our preliminary observations on a teamwork theory incorporating insights from the cognitive and behavioral sciences, and our own studies of teamwork in practice. We then described the first phases of its application to the development of an agent-based model of the work practice of the ISS crew. We use Brahms—an agent-oriented, activity-based language—to model the ISS crew's situated action, communication, and collaboration during the course of their daily activities. In our modeling of a day in the life onboard the ISS we include resource availability, both scheduled and unscheduled activities, and the emergence of work practices. In addition, we model human-machine interaction (such as the collaboration between the crew and robotic systems such as the PSA and the Robonaut). In the next phase of our research we will experimentally explore the use of an enhanced Brahms ISS model as part of an execution environment for teamwork between ISS crews and onboard collaborative software- and robotic-agents relying on integration with the KAoS and NOMADS frameworks. As the study proceeds over the long term, we hope our work will benefit those who plan and participate in work activities in a wide variety of space applications, as well as those who are interested in design and execution tools for teams of robots that can function as effective assistants to humans.

5. ACKNOWLEDGMENTS

The authors gratefully acknowledge the sponsorship of this research by the NASA Cross-Enterprise and Intelligent Systems Programs, and a joint NASA-DARPA ITAC grant. Additional support was provided by DARPA's CoABS, Ultra*Log, and Augmented Cognition programs, and by the Army Research Lab's Advanced Decision Architectures program (ARLADA). We are also grateful for the contributions of Guy Boy, Maggie Breedy, Bill Clancey, Rob Cranfill, Greg Dorais, Ken Ford, Yuri Gawdiak, Jack Hansen, Pat Hayes, Henry Hexmoor, Matt Johnson, Shri Kulkarni, James Lott, Nicola Muscettola, Mike Shafto, Milind Tambe, and David Woods.

6. 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] Agnew, N. M., & Brown, J. L. (1989). Foundations for a theory of knowing: II. Fallible but functional knowledge. *Canadian Psychology*, 30, 168-183.

[3] Allsopp, D., Beautement, P., Bradshaw, J. M., Durfee, E., Kirton, M., Knoblock, C., Suri, N., Tate, A., & Thompson, C. (2002). Coalition Agents eXperiement (CoAX): Multi-agent cooperation in an international coalition setting. *A. Tate, J. Bradshaw, and M. Pechoucek (Eds.), Special issue of IEEE Intelligent Systems*.

[4] Almond, R. G., Bradshaw, J. M., & Madigan, D. (1994). Reuse and sharing of graphical belief network components. In P. Cheeseman & R. W. Oldford (Ed.), *Selecting Models from Data: AI and Statistics IV.* (pp. 113-122). Berlin: Springer-Verlag.

[5] Ambrose, R., Culbert, C., & Rehnmark, F. (2001). An experimental investigation of dexterous robots using Eva tools and interfaces. *AIAA, 4593*.

[6] Barber, K. S., Gamba, M., & Martin, C. E. (2002). Representing and analyzing adaptive decision-making frameworks. In H. Hexmoor, C. Castelfranchi, & R. Falcone (Ed.), *Agent Autonomy.* (pp. 23-42). Dordrecht, The Netherlands: Kluwer.

[7] Barber, K. S., & Martin, C. E. (1999). Agent autonomy: Specification, measurement, and dynamic adjustment. *Proceedings of the Workshop on Autonomy Control Software, International Conference on Autonomous Agents*. Seattle, WA,

[8] Billings, C. E. (1997). Issues concerning human-centered intelligence systems: What's 'human-centered' and what's the problem? *National Science Foundation Workshop on Human-Centered Systems: Information, Interactivity, and Intelligence*. Arlington, VA, http://www.ifp/uiuc.edu/nsfhcs/abstracts/,

[9] Boella, G. (2002). Obligations and cooperation: Two sides of social rationality. In H. Hexmoor, C. Castelfranchi, & R. Falcone (Ed.), *Agent Autonomy.* (pp. 57-78). Dordrecht, The Netherlands: Kluwer.

[10] Boy, G. (1998). *Cognitive Function Analysis*. Stamford, CT: Ablex Publishing.

[11] Boy, G. (2002). Human-centered design of artificial agents: A cognitive function analysis approach. In J. M. Bradshaw (Ed.), *Handbook of Software Agents.* (pp. in press). Cambridge, MA: AAAI Press/The MIT Press.

[12] Bradshaw, J. M., Beautement, P., Kulkarni, S., Suri, N., & Raj, A. (2002). *Toward a deliberative and reactive agent architecture for augmented cognition. DARPA Augmented Cognition Program White Paper.*, Pensacola, FL: Institute for Human and Machine Cognition, University of West Florida.

[13] Bradshaw, J. M., Boy, G., Durfee, E., Gruninger, M., Hexmoor, H., Suri, N., Tambe, M., Uschold, M., & Vitek, J. (Ed.). (2002). *Software Agents for the Warfighter. ITAC Consortium Report.* Cambridge, MA: AAAI Press/The MIT Press.

[14] Bradshaw, J. M., Chapman, C. R., Sullivan, K. M., Almond, R. G., Madigan, D., Zarley, D., Gavrin, J., Nims, J., & Bush, N. (1993). KS-3000: An application of DDUCKS to bone-marrow transplant patient support. *Proceedings of the Sixth Annual Florida AI Research Symposium (FLAIRS '93),* (pp. 89-95). Ft. Lauderdale, FL,

[15] Bradshaw, J. M., Covington, S. P., Russo, P. J., & Boose, J. H. (1990). Knowledge acquisition for intelligent decision systems: integrating *Aquinas* and *Axotl* in DDUCKS. In M. Henrion, R. Shachter, L. N. Kanal, & J. Lemmer (Ed.), *Uncertainty in Artificial Intelligence.* (pp. 255-270). Amsterdam: Elsevier.

[16] Bradshaw, J. M., Dutfield, S., Benoit, P., & Woolley, J. D. (1997). KAoS: Toward an industrial-strength generic agent architecture. In J. M. Bradshaw (Ed.), *Software Agents.* (pp. 375-418). Cambridge, MA: AAAI Press/The MIT Press.

[17] Bradshaw, J. M., Ford, K. M., Adams-Webber, J. R., & Boose, J. H. (1993). Beyond the repertory grid: New approaches to constructivist knowledge acquisition tool development. In K. M. Ford & J. M. Bradshaw (Ed.), *Knowledge Acquisition as Modeling.* (pp. 287-333). New York: John Wiley.

[18] Bradshaw, J. M., Greaves, M., Holmback, H., Jansen, W., Karygiannis, T., Silverman, B., Suri, N., & Wong, A. (1999). Agents for the masses: Is it possible to make development of sophisticated agents simple enough to be practical? *IEEE Intelligent Systems*(March-April), 53-63.

[19] Bradshaw, J. M., Suri, N., Cañas, A., Davis, R., Ford, K. M., Hoffman, R., Jeffers, R., & Reichherzer, T. (2001). Terraforming cyberspace. *IEEE Computer*(July), 49-56. An expanded and updated version of this article is available in D. Marinescu and C. Lee (Eds.), Process Coordination and Ubiquitous Computing. Boca Raton, FL: CRC Press, 2002, 165-185.

[20] Brainov, S., & Hexmoor, H. (2002). Quantifying autonomy. In H. Hexmoor, C. Castelfranchi, & R. Falcone (Ed.), *Agent Autonomy.* (pp. 43-56). Dordrecht, The Netherlands: Kluwer.

[21] Burrough, B. (1998). *Dragonfly: NASA and the Crisis Aboard Mir*. New York, NY:: HarperCollins.

[22] Burstein, M. H., & McDermott, D. V. (1996). Issues in the development of human-computer mixed-initiative planning. In B. Gorayska & J. L. Mey (Ed.), *Cognitive Technology: In Search of a Humane Interface.* Elsevier Science.

[23] Chernoff, H. (1973). The use of faces to represent points in k-dimensional space graphically. *Journal of the American Statistical Association*, 68(June), 361-368.

[24] Christofferson, K., & Woods, D. D. (2002). How to make automated systems team players. In E. Salas (Ed.), *Advances in Human Performance and Cognitive Engineering Research, Vol. 2.* JAI Press, Elsevier.

[25] Clancey, W. J. (1997). *Situated Cognition: On Human Knowledge and Computer Representations*. Cambridge, England: Cambridget University Press.

[26] Clancey, W. J. (2002). Simulating activities: Relating motives, deliberation, and attentive coordination. *Cognitive Systems Review, special issue on Situated and Embodied Cognition*.

[27] Cohen, P. R., & Levesque, H. J. (1991). *Teamwork*. Technote 504. Menlo Park, CA: SRI International, March.

[28] Cohen, P. R., Levesque, H. R., & Smith, I. (1997). On team formation. In J. Hintikka & R. Tuomela (Ed.), *Contemporary Action Theory.* Synthese.
[29] Cohen, R., Allaby, C., Cumbaa, C., Fitzgerald, M., Ho, K., Hui, B., Latulipe, C., Lu, F., Moussa, N., Pooley, D., Qian, A., & Siddiqi, S. (1998). What is initiative? *User Modeling and User-Adapted Interaction*, 8(3-4), 171-214.
[30] Cohen, R., & Fleming, M. (2002). Adjusting the autonomy in mixed-initiative systems by reasoning about interaction. In H. Hexmoor, C. Castelfranchi, & R. Falcone (Ed.), *Agent Autonomy.* (pp. 105-122). Dordrecht, The Netherlands: Kluwer.
[31] Cypher, A. (Ed.). (1993). *Watch What I Do: Programming by Demonstration.* Cambridge, MA: MIT Press.
[32] Damianou, N., Dulay, N., Lupu, E. C., & Sloman, M. S. (2000). *Ponder: A Language for Specifying Security and Management Policies for Distributed Systems, Version 2.3*. Imperial College of Science, Technology and Medicine, Department of Computing, 20 October 2000.
[33] Dorais, G., Bonasso, R. P., Kortenkamp, D., Pell, B., & Schrekenghost, D. (1999). Adjustable autonomy for human-centered autonomous systems on Mars. *Proceedings of the AAAI Spring Symposium on Agents with Adjustable Autonomy. AAAI Technical Report SS-99-06*. Menlo Park, CA, Menlo Park, CA: AAAI Press,
[34] Erickson, T. (1997). Designing agents as if people mattered. In J. M. Bradshaw (Ed.), *Software Agents.* (pp. 79-96). Cambridge, MA: The AAAI Press/The MIT Press.
[35] Falcone, R., & Castelfranchi, C. (2002). From automaticity to autonomy: The frontier of artificial agents. In H. Hexmoor, C. Castelfranchi, & R. Falcone (Ed.), *Agent Autonomy.* (pp. 79-103). Dordrecht, The Netherlands: Kluwer.
[36] Feltovich, P., Spiro, R. R., & Coulson, R. L. (1997). Issues of expert flexibility in contexts characterized by complexity and change. In P. Feltovich, K. M. Ford, & R. Hoffman (Ed.), *Expertise in Context: Human and Machine.* (pp. 125-146). Menlo Park, CA: AAAI/MIT Press.
[37] Ferguson, G., Allen, J., & Miller, B. (1996). TRAINS-95: Towards a mixed-initiative planning assistant. *Proceedings of the Third Conference on Artificial Intelligence Planning Systems (AIPS-96),* (pp. 70-77). Edinburgh, Scotland,
[38] Fitts, P. M. (Ed.). (1951). *Human Engineering for an Effective Air Navigation and Traffic Control System.* Washington, D.C.: National Research Council.
[39] Ford, K. M., Bradshaw, J. M., Adams-Webber, J. R., & Agnew, N. M. (1993). Knowledge acquisition as a constructive modeling activity. In K. M. Ford & J. M. Bradshaw (Ed.), *Knowledge Acquisition as Modeling.* (pp. 9-32). New York: John Wiley.
[40] Ford, K. M., Glymour, C., & Hayes, P. (1997). Cognitive prostheses. *AI Magazine*, 18(3), 104.
[41] Ford, K. M., & Hayes, P. (1998). On computational wings: Rethinking the goals of Artificial Intelligence. *Scientific American. Special issue on "Exploring Intelligence"*, 9(4), 78-83.
[42] Gawdiak, Y., Bradshaw, J. M., Williams, B., & Thomas, H. (2000). R2D2 in a softball: The Personal Satellite Assistant. H. Lieberman (Ed.), *Proceedings of the ACM Conference on Intelligent User Interfaces (IUI 2000),* (pp. 125-128). New Orleans, LA, New York: ACM Press,
[43] Gelernter, D. (1992). *Mirror Worlds*. New York: Oxford University Press.
[44] Gershenfeld, N. A. (1999). *When Things Start to Think*. New York: Henry Holt and Company.
[45] Glasgow, J., Narayanan, N. H., & Chandrasekaran, B. (Ed.). (1995). *Diagrammatic Reasoning: Computational and Cognitive Perspectives.* Menlo Park, CA: AAAI Press.
[46] 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,
[47] Grosz, B. J. (1996). Collaborative systems. *AI Magazine*, 17(2), 67-85.
[48] Hamilton, S. (2001). Thinking outside the box at IHMC. *IEEE Computer*, 61-71.
[49] Hancock, P. A., & Scallen, S. F. (1998). Allocating functions in human-machine systems. In R. Hoffman, M. F. Sherrick, & J. S. Warm (Ed.), *Viewing Psychology as a Whole.* (pp. 509-540). Washington, D.C.: American Psychological Association.

[50] Hoffman, R. (1997). How to doom yourself to repeat the past: Some reflections on the history of cognitive technology. *Cognitive Technology*, 2(2), 4-15.
[51] Hoffman, R., Feltovich, P., Ford, K. M., Woods, D. D., Klein, G., & Feltovich, A. (2002). A rose by any other name… would probably be given an acronym. *IEEE Intelligent Systems*, July-August, 72-80.
[52] Holtzman, S. H. (1989). *Intelligent Decision Systems*. Reading, Massachusetts: Addison-Wesley.
[53] Horvitz, E. (1999). Principles of mixed-initiative user interfaces. *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '99)*. Pittsburgh, PA, New York: ACM Press,
[54] Horvitz, E., Jacobs, A., & Hovel, D. (1999). Attention-sensitive alerting. *Proceedings of the Conference on Uncertainty and Artificial Intelligence (UAI '99),* (pp. 305-313). Stockholm, Sweden,
[55] Howard, R. A., & Matheson, J. E. (1984). Influence diagrams. In R. A. Howard & J. E. Matheson (Ed.), *Readings on the Principles and Applications of Decision Analysis.* (pp. 719-762). Menlo Park, California: Strategic Decisions Group.
[56] Kahle, B. (1993). Interview of Brewster Kahle. *Intertek*, 4, 15-17.
[57] Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11, 65-99.
[58] Licklider, J. C. R. (1960). Man-computer symbiosis. *IRE Transactions in Electronics. New York: Institute of Radio Engineers.*, 4-11.
[59] Lieberman, H. (Ed.). (2001). *Your Wish is My Command: Programming By Example.* San Francisco, CA: Morgan Kaufmann.
[60] Lieberman, H., & Selker, T. (2002). Agents for the user interface. In J. M. Bradshaw (Ed.), *Handbook of Software Agents.* Cambridge, MA: AAAI Press/The MIT Press.
[61] Luck, M., D'Inverno, M., & Munroe, S. (2002). Autonomy: Variable and generative. In H. Hexmoor, C. Castelfranchi, & R. Falcone (Ed.), *Agent Autonomy.* (pp. 9-22). Dordrecht, The Netherlands: Kluwer.
[62] Lupu, E. C., & Sloman, M. S. (1999). Conflicts in policy-based distributed systems management. *IEEE Transactions on Software Engineering—Special Issue on Inconsistency Management*.
[63] McDermott, D. (1976). Artificial intelligence meets natural stupidity. *SIGART Newsletter*(57), 4-9.
[64] Milewski, A. E., & Lewis, S. H. (1994). *Design of intelligent agent user interfaces: Delegation issues*. AT&T Corporate Information Technologies Services Advanced Technology Planning, October 20.
[65] Muscettola, N., Dorais, G., Fry, C., Levinson, R., & Plaunt, C. (2002). IDEA: Planning at the core of autonomous reactive agents. *Proceedings of the Workshop on Online Planning and Scheduling, Sixth International Conference on AI Planning and Scheduling*. Toulouse, France,
[66] Myers, K., & Morley, D. (2002). Directing agents. In H. Hexmoor, C. Castelfranchi, & R. Falcone (Ed.), *Agent Autonomy.* (pp. 143-162). Dordrecht, The Netherlands: Kluwer.
[67] National_Aeronautics_and_Space_Administration (1999). *International Space Station Operations and Planned Training Manual. ISS OPS and PL TM 21109, TD 9711, Rev. A*. Houston, TX: Johnson Space Center.
[68] Norman, D. A. (1992). Cognitive artifacts. In J. M. Carroll (Ed.), *Designing Interaction: Psychology at the Human-Computer Interface.* (pp. 17-38). Cambridge: Cambridge University Press.
[69] Norman, D. A. (1997). How might people interact with agents? In J. M. Bradshaw (Ed.), *Software Agents.* (pp. 49-55). Cambridge, MA: The AAAI Press/The MIT Press.
[70] Peacock, B., Rajulu, S., & Novak, J. (2001). Human factors and the International Space Station. *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting*.

[71] Rich, C., Sidner, C., & Lesh, N. (2001). COLLAGEN: Applying collaborative discourse theory. *AI Magazine*.
[72] Rich, C., & Waters, R. C. (1988). Automatic programming: Myths and prospects. *IEEE Computer*, 21(8), 40-51.
[73] Sarter, N., Woods, D. D., & Billings, C. E. (1997). Automation surprises. In G. Salvendy (Ed.), *Handbook of Human factors/Ergonomics, 2nd Edition.* New York, NY: John Wiley.
[74] Scerri, P., Pynadath, D., & Tambe, M. (2002). Adjustable autonomy for the real world. In R. Falcone (Ed.), *Agent Autonomy.* (pp. 163-190). Dordrecht, The Netherlands: Kluwer.
[75] Sierhuis, M. (2001) *Brahms: A Multi-Agent Modeling and Simulation Language for Work System Analysis and Design*. Doctoral, University of Amsterdam.
[76] Smith, D. C., Cypher, A., & Spohrer, J. (1997). KidSim: Programming Agents without a Programming Language. In J. M. Bradshaw (Ed.), *Software Agents.* (pp. 165-190). Menlo Park, CA: The AAAI Press/The MIT Press.
[77] Sowa, J. F. (1989). Knowledge acquisition by teachable systems. In J. P. Martins & E. M. Morgado (Ed.), *EPIA 89, Lecture Notes in Artificial Intelligence.* (pp. 381-396). Berlin: Springer-Verlag.
[78] Still, D. L., & Temme, L. A. (2001). OZ: A human-centered cockpit display. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference*. Orlando, FL,
[79] Suchman, L. A. (1987). *Plans and Situated Action: The Problem of Human-Machine Communication*. Cambridge, MA: Cambridge University Press.
[80] Suri, N., Bradshaw, J. M., Breedy, M. R., Groth, P. T., Hill, G. A., & Jeffers, R. (2000). Strong Mobility and Fine-Grained Resource Control in NOMADS. *Proceedings of the 2nd International Symposium on Agents Systems and Applications and the 4th International Symposium on Mobile Agents (ASA/MA 2000).* Zurich, Switzerland, Berlin: Springer-Verlag,
[81] Tambe, M., Pynadath, D., Chauvat, C., Das, A., & Kaminka, G. (2000). Adaptive agent architectures for heterogeneous team members. *Proceedings of the International Conference on Multi-Agent Systems (ICMAS 2000)*.
[82] Tambe, M., Shen, W., Mataric, M., Pynadath, D. V., Goldberg, D., Modi, P. J., Qiu, Z., & Salemi, B. (1999). Teamwork in cyberspace: Using TEAMCORE to make agents team-ready. *Proceedings of the AAAI Spring Symposium on Agents in Cyberspace*. Menlo Park, CA, Menlo Park, CA: The AAAI Press,
[83] Wiener, E. L. (1989). *Human Factors of Advanced Technology, "Glass Cockpit" Transport Aircraft. NASA Contractor Report No. 177528*. NASA Ames Research Center,
[84] Woods, D. D. (2002). Steering the reverbereations of technology change on fields of practice: Laws that govern cognitive work. *Proceedings of the Annual Meeting of the Cognitive Science Society*.
[85] Woods, D. D., & Sarter, N. (2000). Learning from automation surprises and going sour accidents. In N. Sarter & R. Amalberti (Ed.), *Cognitive Engineering in the Aviation Domain.* Hillsdale, NJ: Lawrence Erlbaum.