Policy-Based Coordination in Joint Human-Agent Activity*

Jeffrey M. Bradshaw, Paul Feltovich, Hyuckchul Jung, Shri Kulkarni, James Allen,
Larry Bunch, Nathanael Chambers, Lucian Galescu, Renia Jeffers, Matthew Johnson,
Maarten Sierhuis, William Taysom, Andrzej Uszok, and Ron Van Hoof

Institute for Human and Machine Cognition (IHMC), 40 S. Alcaniz, Pensacola, FL 32502 - USA
{jbradshaw, hjung, skulkarni, jallen, lbunch, nchambers, pfeltovich, lgalescu, rjeffers, mjohnson, wtaysom, auszok}@ihmc.us

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

Abstract – *In this paper, we outline an approach to policy-based coordination in joint human-agent activity. The approach is grounded in a theory of joint activity originally developed in the context of discourse, and now applied to the broader realm of human-agent interaction. We have been gradually implementing selected aspects of policy-based coordination within the KAoS services framework and have been developing a body of examples that will guide additional testing of these ideas through detailed studies of work practice.*

Keywords: Policy, KAoS, coordination, joint activity, human-agent interaction.

1 Introduction

The concept of automation—which began with the straightforward objective of replacing whenever feasible any task currently performed by a human with a machine that could do the same task better, faster, or cheaper—became one of the first issues to attract the notice of early human factors researchers. These researchers attempted to characterize the general strengths and weaknesses of humans and machines [29]. The resulting discipline of *function allocation* aimed to provide a rational means of determining which system-level functions should be carried out by humans and which by machines.

Over time it became plain to researchers that things were not as simple as they first appeared. For example, many functions in complex systems are shared by humans and machines; hence the need to consider synergies and conflicts among the various performers of joint actions. Also, the suitability of a particular human or machine to take on a particular task may vary by time and over different situations; hence the need for methods of function allocation that are dynamic and adaptive [34]. Moreover, it has become clear that function allocation is not a simple process of transferring responsibilities from one component to another [5]. Automated assistance of whatever kind does not simply enhance our ability to perform the task: it changes the nature of the task itself [16; 27; 47]. Those who have had a five-year-old child help them by doing 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.

Some of the most important contributions to a more sophisticated understanding of what makes automation effective have been made in the field of software and robotic agents. The ultimate desire of agent researchers is to make automation a team player [16; 45]. In contrast to significant early research that focused almost exclusively on how to make individual agents more autonomous, much of current agent research seeks to understand and satisfy requirements for the basic aspects of joint activity, either within multi-agent systems or as part of human-agent teamwork. Specific approaches to human-agent teamwork have been explored by researchers, albeit in many forms and with somewhat divergent perspectives. For example, research communities have formed around the topics of interface agents and assistants [17; 21; 35; 40; 41], adjustable autonomy [9; 11; 24; 25; 32; 43; 44], mixed-initiative systems [2; 3; 9; 14; 28], and collaboration theory [33; 50].

An adequate approach to joint human-agent activity requires three things:

- grounding in an appropriate theory,
- a comprehensive, flexible, and dynamic implementation of the theory, and
- experience in the ongoing study of work practice.

With respect to theory, we have made a first attempt to generalize the concept of joint activity, developed by Clark [18] in the context of discourse, in order to apply it to the broader realm of human-agent interaction [38] (Section 2). We advocate a policy-based coordination approach (Section 3). We have been gradually implementing selected aspects of policy-based coordination within the KAoS services framework [6; 58] (Section 4) and have begun to develop a body of examples (Section 5) that will guide additional testing of these ideas through detailed studies of work practice [52; 53].

* **0-7803-8566-7/04/$20.00 © 2004 IEEE.**

2 Joint Activity and Coordination[1]

A joint activity is an extended set of behaviors that are carried out by an ensemble of people who are coordinating with each other [18, p. 3]. In a joint activity, the parties involved must intend to produce something that is a genuine joint product—as Woods writes, "It's not cooperation if either you do it all or I do it all" [61]. In order to carry out the joint activity, the parties effectively enter into what we call a "Basic Compact"—an agreement (usually tacit) that all parties will support the process of coordination. If there is no need for substantive coordination among the various parties as they carry out their actions, then this is parallel—not joint—activity.

Joint activity is a *process*, extended in space and time. There is a time when the parties enter into joint activity and a time when it has ended. These are not "objective" points of time that would necessarily be agreed on by any "observer-in-the-world," but most importantly are interpretations arrived at by the parties involved [18, p. 84]. In some circumstances the entry and exit points may be very clear such as when two people play a classical duet; the same would probably not be said of musicians involved in a jam session or of participants in a mass demonstration.

The overall *structure* of joint activity is one of embedded sets of actions, some of which may also be joint and some of which may be accomplished more or less individually. All these actions likewise have entry and exit points, although as we have mentioned earlier, these points are not epistemologically "objective." Synchronizing entry and exit points of the many embedded phases involved in complex joint activity is a major challenge to coordination.

2.1 Requirements for effective coordination

There are three requirements for effective coordination: interpredictability, common ground, and directability:

- *Interpredictability:* In highly interdependent activities, it becomes possible to plan one's own actions (including coordination actions) only when what others will do can be accurately predicted. Skilled teams become interpredictable through shared knowledge and idiosyncratic coordination devices developed through extended experience in working together; bureaucracies with high turnover compensate for experience by substituting explicit predesigned structured procedures and expectations.

- *Common ground:* Common ground refers to the pertinent mutual knowledge, beliefs, and assumptions that support interdependent actions in the context of a given joint activity [19]. This includes initial common ground prior to engaging in the joint activity as well as mutual knowledge of shared history and current state that is obtained while the activity is underway. Unless I can make good assumptions about what you know and what you can do, we cannot effectively coordinate.

- *Directability:* Directability refers to the capacity for deliberately assessing and modifying the actions of the other parties in a joint activity as conditions and priorities change [16]. Effective coordination requires responsiveness of each participant to the influence of the others as the activity unfolds

2.2 Different aspects of coordination

In attempting to define coordination, various researchers have emphasized different aspects. For example, Olson, Malone, and Smith [49] state that "Coordination is managing dependencies between activities." This definition is fine as far as it goes, but it omits mention of the important process of conflict resolution among interacting goals. Klein [37] has stated that "Coordination is the attempt by multiple entities to act in concert in order to achieve a common goal by carrying out a script they all understand." However, many joint activities requiring coordination (e.g., driving on a highway, engaging in casual conversation) neither involve an overriding script nor a common goal other than the goal of working cooperatively in order to achieve individual goals.[2]

From our perspective—and consistent with the overall thrust of Clark's work—effective coordination in joint activity may require a complex choreography of events, comprising various phases of action, that are guided by signaling and coordination devices, with the end of maintaining the quality of the interaction, while expending a minimum of coordination cost. Given a structure of embedded actions—some of which may be joint actions—as well as overall joint activity, there are two questions that need answering:

1. How does coordination take place in the more local joint acts that make up an overall joint activity? (section 2.1)
2. How does coordination take place at the more macro level of the overall joint activity itself? (section 2.2)

[1] In this section, we draw extensively on the analysis of Klein *et al.* [38]. See this reference for more detail and examples, as well as a discussion of issues in the design of agents for joint activity.

[2] Cartwright and Zander [15] define three levels of goals that are relevant in understanding the behavior of a team: individual goals; team goals, and individual goals for the team. We have extended this goal ontology with the concept of team goals for the individual [53].

2.3 Coordination devices use in local actions

With regard to the first question, the "coordination devices" [18, pp. 64-65] play a major role in local joint actions. These include:

- *Agreement:* Coordinating parties are sometimes simply able to communicate their intentions and work out elements of coordination. This category includes diverse forms of signaling that have shared meaning for the participants, including language, signs, gestures, and the like.
- *Convention:* Often, prescriptions of various types apply to how parties interact. These can range from rules and regulations, to less formal codes of appropriate conduct such as norms of practice in a particular professional community, or established practices in a workplace. Coordination by convention depends on structures outside of a particular episode of joint activity.
- *Precedent:* Coordination by precedent is like coordination by convention, except that it applies to norms and expectations developed within an episode of the ongoing process of a joint activity (or across repeated episodes of such activity if the participants are a long-standing team that repeats conduct of some procedure): "That's the way we did it last time."
- *Salience:* Salience is perhaps the coordination device that is most difficult to understand and describe.[3] It has to do with how the ongoing work of the joint activity arranges the workspace so that next move becomes highlighted or otherwise apparent among the many moves that could conceivably be chosen. For example, in a surgery, exposure of a certain element of anatomy, in the course of pursuing a particular surgical goal, can make it clear to all parties involved what to do next. Coordination by salience is a sophisticated kind of coordination produced by the very conduct of the joint activity itself. It requires little or no overt communication and is likely the predominant mode of coordination among long-standing, highly practiced teams.

2.4 Overall activity coordination and phases

Coordination across the entire course of an extended joint activity is in some ways similar and in some ways different from the more local coordination involved in individual joint actions and subactions. For instance, there may be "scripted" conventions for conducting an entire procedure-just as there are for conducting more local components of it. That is, joint activities can be more or less open in execution, depending on the presence of applicable norms, policies, procedures, and the like. In addition to regulatory coordination mechanisms, there are other kinds of macro guides that serve to coordinate across the course of an entire joint activity. Examples are agreements about plans and policies for some activity worked out in advance by the participants, or a prior extensive outline worked out by authors involved in writing a joint academic manuscript. It has been argued that some of the reasons for "standardizing" procedures are to aid coordination and to prevent untoward interactions so that some earlier move does not clobber some necessary later move (e.g., [51]). Of course, any of these overarching coordination devices usually needs to be revisited, and very likely adjusted, as the actual work unfolds.

Previously, we have discussed five general phases in joint activity [53], which can only be noted briefly here :[4]

1. Recognition of the need for joint activity;
2. Team formation;
3. Ongoing coordination, support for maintenance of common ground and team integrity throughout the activity;
4. Recognition of either resolution of the need for joint activity or impasse with respect to such a resolution; and
5. Team disbanding.

2.5 Common ground and integrity support

Phase 3 above refers to the ongoing need for support of common ground and team integrity. Participants in joint activities work to establish and maintain common ground in several ways:

- Structuring the *preparations* ("practicing, rehearsing") in order to establish an initial calibration of beliefs, assumptions, harmonization of any individual and joint goals, and to establish routines for use during execution;
- *Sustaining* common ground through clarifications and reminders that either serve to verify beliefs or to provide team members a chance to challenge assumptions;
- *Updating* others about changes that they may have missed;
- *Monitoring* the other team members for anomalies or signs of erosion of common ground;
- *Repairing* losses in common ground.

Team integrity is preserved in part through effective maintenance of common ground. Additional actions to reinforce team identity, establish rewards and sanctions, and protect the team from encroachment of competing allegiances may also be important.

[3] Part of the complication is the relationships among these mechanisms. For example, conventions and precedents may be essential in salience "assignment."

[4] A classic reference for team development in people is Tuckman [57], who coined the following memorable list of stages: forming (the orientation stage), storming (the conflict stage), norming (the cohesion stage), performing (the task-performance stage), and adjourning (the dissolution stage).

3 Policy-Based Coordination

In this section, we describe our approach to policy-based coordination. We briefly summarize the role of social constraints and other regulatory mechanisms: in joint activity and more broadly in human culture (section 3.1). We then give our motivations for a policy-based approach to regulation, and, in turn, the predictability necessary for coordination (section 3.2).

3.1 Human culture regulatory mechanisms

In his description of joint activity, Clark [18] borrows a definition from Levinson [39, p. 69] that emphasizes the central role of social constraints governing the actions of participants:

> "I take the notion of an activity type to refer to a fuzzy category whose focal members are goal-defined, socially constituted, bounded, events with constraints on participants, setting, and so on, but above all on the kinds of allowable contributions."

Taking inspiration from biological and anthropological perspectives, Feltovich *et al*. [26] have pursued a similar line of thinking. For example, Feltovich, *et al*, quoting Geertz [30], argues that because of our vast behavioral repertoire, and because we are so underdetermined in our biology, a very large portion of what humans do and create is constituted to "control ourselves"! In this view, the role of human culture is that of a vast, fabricated self-regulatory mechanism:[5]

> "I want to propose two ideas: The first of these is that culture is best seen not as complexes of concrete behavior patterns—customs, usages, traditions, habit clusters—as has, by and large, been the case up to now, but as a set of control mechanisms—plans, recipes, rules, instructions (what computer engineers call 'programs')—for the governing of behavior. The second idea is that man is precisely the animal most desperately dependent upon such extragenetic, outside-the-skin mechanisms, such cultural programs, for ordering his behavior..." [30, p. 44].

Order and predictability in human culture may have a basis in simple cooperative acts among people, in which the parties "contract" to engage together in a set of interlinked, mutually beneficial joint activities [54]. From this simple base, there are constructed elaborate and intricate systems of regulatory tools, from formal legal systems, to standards of professional practice, to norms of proper everyday behavior (along with associated methods of punishment or even simple forms of shaming for violations of these).

3.2 Motivation for a policy-based approach

We believe that many of these kinds of constraints and regulatory mechanisms can be effectively represented using policies [6; 9]. In common use of the term in a computer science context, policies are a means to dynamically regulate the behavior of a system without changing code or requiring the cooperation of the components being governed:

- Through policy, people can precisely express bounds on autonomous behavior (permissions) and expectations of performance (obligations) in a way that is consistent with their appraisal of an agent's *competence* in a given context.
- Deliberate *maliciousness* and inadvertent bugs can be minimized. Because policy enforcement is handled externally to the agent, malicious and buggy agents can no more exempt themselves from the constraints of policy than benevolent and well-written ones can. Moreover, a separation of concerns can be maintained between core task-related agent competencies designed in advance by developers, and policy-governed behavior that is specified, enforced, and adapted at runtime.
- The ability to change policies dynamically means that poorly performing agents can be immediately brought into *compliance* with corrective measures.

Elsewhere we have pointed out other benefits of policy-based approaches, including reusability, efficiency, extensibility, context-sensitivity, verifiability, support for both simple and sophisticated components, and external reasoning about component behavior [6].

From the perspective of joint activity, policies can be used to explicitly express agreements, conventions, precedents, and salience conditions that help make automated components more effective players in mixed human-agent teams. They can be used to enforce bounds and expectations that increase *interpredictability,* they can be used to establish and maintain *common ground*, and their ability to be imposed and adjusted at runtime enables dynamic *directability*.

[5] We recognize that Geertz represents only one of many views of culture, but a discussion of competing views is beyond the scope of this paper. Though his basic claim that the role of culture is primarily for self-regulation may at first seem far-fetched, if we consider not only explicit permissions and obligations as forms of regulation but also regulation by design of affordances that make appropriate actions easy and inappropriate ones difficult or impossible (e.g., positioning of doorways and sidewalks, connectors on Lego pieces), then we find Geertz's argument more appealing. Such considerations are topics of our ongoing research.

4 KAoS Policy and Domain Services

In the mid-1990's, we began to define the initial version of KAoS, a set of platform-independent services that enable people to define policies ensuring adequate predictability and controllability of both agents and traditional distributed systems [10; 13; 42; 55; 59; 60]. Since that time, we have also become involved in a series of projects requiring close and continuous interaction between humans and agents in military and space settings. In collaboration with a variety of research partners, we have been developing a generic model of policy-based joint activity that includes policies to assure natural and effective interaction in mixed teams of people and software/robotic agents [1; 6; 9; 12; 53].

KAoS supports a wide range of policy and domain services. KAoS Domain Services provide the capability for groups of software components, people, resources, and other entities to be semantically described and structured into organizations of domains and subdomains to facilitate collaboration and external policy administration. KAoS Policy Services allow for the specification, management, conflict resolution, and enforcement of policies within domains. While initially oriented to the dynamic and complex requirements of software and robotic agent platforms (e.g., Nomads, DARPA CoABS Grid, Cougaar, Voyager, Brahms, TRIPS, SFX), KAoS services have been extended to work equally well with traditional clients in CORBA, Grid Computing, and Web Services environments [58; 60]. A comparison between KAoS, Rei, and Ponder for policy specification, representation, reasoning, and enforcement is given in [56]. More complete descriptions of KAoS and Nomads can be found in [6; 58].

The KAoS policy ontology distinguishes between *authorizations* (i.e., constraints that permit or forbid some action) and *obligations* (i.e., constraints that require some action to be performed when a state- or event-based trigger occurs, or else serve to waive such a requirement) [22]. Other policy constructs (e.g., delegation, role-based authorization) are built out of the basic primitives of domains plus these four policy types.

The use of OWL (Web Ontology Language, http://www.w3.org/TR/owl-features/) to represent policies enables reasoning about the controlled environment, about policy relations and disclosure, policy conflict resolution, as well as about domain structure and concepts. KAoS reasoning methods are built on top of Stanford's Java Theorem Prover (http://www.ksl.stanford.edu/software/JTP/) and exploit description-logic-based subsumption and instance classification algorithms and, if necessary, controlled extensions to description logic (e.g., role-value maps). No rules are used in policy representation—rather conditions are expressed as property restrictions on actions associated with the policy ontologies.

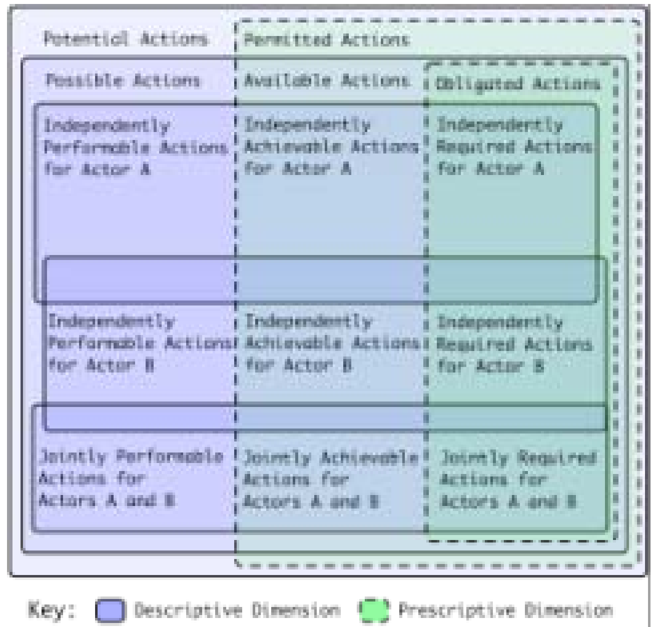

Figure 1. Dimensions of adjustable autonomy.

4.1 KAoS and manual adjustment of policy

KAoS allows people to manually specify, analyze, and modify authorization and obligation policies at runtime through the use of the KPAT (KAoS Policy Administration Tool) graphical user interface . KPAT hides the complexity of the OWL representation from users. The reasoning and representation capabilities of OWL are used to full advantage to make the process as simple as possible.

4.2 Automatically adjustable autonomy

Realizing that policies themselves may sometimes need to be modified automatically in response to changing conditions and priorities, we have developed an initial implementation of Kaa (KAoS adjustable autonomy), a component that can perform adjustments of policy as well as other aspects of agent autonomy [9; 11].

Figure 1 illustrates some important dimensions relating to the adjustment of autonomy. Note that the figure does not show every possible configuration of the dimensions, but rather exemplifies a particular set of relations holding for the actions of a particular set of actors in a given situation. There are two basic dimensions:

- a *descriptive* dimension corresponding to a first sense of autonomy (self-sufficiency) that stretches horizontally to describe the actions an actor in a given context is *capable* of performing; and
- a *prescriptive* dimension corresponding to a second sense of autonomy (self-directedness) running vertically to describe the actions an actor in a given context is allowed to perform or which it must perform by virtue of *policy* constraints in force.

The outermost rectangle, labeled *potential actions*, represents the set of all actions across all situations defined in some ontology under current consideration. The rectangle labeled *possible actions* represents the set of potential actions whose performance by one or more actors is deemed plausible in a given situation [4; 23]. Note that the definition of possibilities is strongly related to the concept of affordances [31; 46], in that it relates the features of the situation to classes of actors capable of exploiting these features in the performance of actions.[6]

Of these possible actions, only certain ones will be deemed *performable* for a given actor (e.g., Actor A) in a given situation. *Capability*, i.e., the power that makes an action performable, is a function of the *abilities* (e.g., knowledge, capacities, skills) and *conditions* (e.g., ready-to-hand resources) necessary for an actor to successfully undertake some action in a given context. Certain actions may be *independently performable* by either Actor A or B; other actions can be independently performed by either one or the other uniquely. Yet other actions are *jointly performable* by a set of actors.

Along the prescriptive dimension, declarative policies may specify various *permissions* and *obligations* [22]. An actor is *free* to the extent that its actions are not limited by permissions or obligations. *Authorities* may impose or remove involuntary policy constraints on the actions of actors. Alternatively, actors may voluntarily enter into *agreements* that mutually bind them to some set of policies for the duration of the agreement. The *effectivity* of an individual policy specifies when it is in or out of force.

The set of *permitted actions* is determined by *authorization policies* that specify which actions an actor or set of actors is allowed (*positive authorizations* or *A+* policies) or not allowed (*negative authorizations* or *A-* policies) to perform in a given context. The intersection of what is possible and what is permitted delimits the set of *available actions*. Of those actions that are available to a given actor or set of actors, some subset may be judged to be *independently achievable* in the current context. Some actions, on the other hand, would be judged to be only *jointly achievable.*

Finally, the set of *obligated actions* is determined by *obligation policies* that specify actions that an actor or set of actors is required to perform (*positive obligations* or *O+* policies) or for which such a requirement is waived (*negative obligations* or *O-* policies).[7] *Jointly obligated actions* are those that two or more actors are explicitly required to perform.

Assistance from Kaa in making limited adjustable autonomy decisions might typically be required when it is anticipated that the current configuration of policies has led to or is likely to lead to poor performance or failure, and when there is no set of competent and authorized humans available to make the decisions themselves. Ultimately, it is a matter of expected utility: the utility of making the change vs. the utility of the status quo.

The current implementation of Kaa uses influence-diagram-based decision-theoretic algorithms to determine what if any changes should be made in agent autonomy [7; 8; 36]. However, Kaa is designed to allow other kinds of decision-making components to be plugged-in if an alternative approach is preferable. When invoked, Kaa first compares the utility of various adjustment options (e.g., increases or decreases in permissions and obligations, acquisition of capabilities, proactive changes to the situation to allow new possibilities), and then—if a change in the status quo is warranted—takes action to implement the recommended alternative.

4.3 KAoS and mixed-initiative behavior

Effective joint activity obviously requires dynamic give and take among all the participants. The concept of *mixed-initiative interaction*—involving some combination of humans and agents—has been succinctly described by Allen as follows:

> “Mixed-initiative refers to a flexible interaction strategy, where each agent can contribute to the task what it does best. Furthermore, in the most general cases, the agents’ roles are not determined in advance, but opportunistically negotiated between them as the problem is being solved. At any one time, one agent might have the initiative—controlling the interaction—while the other works to assist it, contributing to the interaction as required. At other times, the roles are reversed, and at other times again the agents might be working independently, assisting each other only when specifically asked. The agents dynamically adapt their interaction style to best address the problem at hand” [2, p. 14].

The taking of initiative is obviously a decision that derives from the results of an agent consulting its own reasoning processes, and not one directly instigated by external policy-related components. The set of policies in force and adjustments made to this set, however, can affect the process of initiative-taking in a number of ways [9].

The use of policies in various aspects of joint activity is illustrated briefly in the examples in the following section.

---

[6] As expressed by Norman: “Affordances reflect the possible relationships among actors and objects: they are properties of the world” [48].

[7] A negative obligation corresponds to the idea of “you are not obliged to” rather than “you are obliged not to”—this second sense corresponds to a negative authorization with the subject doing the enforcing (similar to Ponder’s *refrain* policies).

5 Examples

Coordination and phases of the overall activity. Among other things, the classic teamwork research of Cohen and Levesque [20] addressed the kinds of communication required to form, maintain, and abandon joint goals. At least some of these important heuristics could be represented in KAoS as particular kinds of obligations, for example: If it becomes known that a joint goal has been achieved, or has become unachievable or irrelevant, then all team members must be notified in an appropriate manner.

Coordination of local actions. Policies can be used to represent agreements, conventions, and precedents that establish accepted practice in the performance of joint activities. For example, authorization policies may permit or restrict the delegation of certain kinds of tasks, or the access to certain kinds of information. Other policies might mandate certain kinds of logging or signaling to take place before the performance of a given action. A form of coordination by salience can be implemented by a policy that highlights certain features of an operator's display in a given context. Policies allowing greater or requiring lessened initiative to be taken by particular team members in a given context can increase the resilience of a team in the face of unforeseen circumstances.

Support for the establishment and maintenance of common ground and team integrity. Policies constitute explicit representations of important aspects of common ground that become the subject of negotiation in structuring the preparations for joint activity. To the degree that they can mandate notifications of relevant information and signals pertaining to important events, they become a mechanism for sustaining common ground; to the degree that policy enforcement mechanisms can prevent or warn of erosion of common ground, they become a mechanisms for monitoring and repair.

6 Future Directions

Based on preliminary results of research and development of applications requiring joint human-agent activity funded by DARPA, NASA, ONR, and the Army, we have now started to construct a more systematic body of policy-based coordination examples that will guide additional testing of these ideas in detailed studies of work practice. We are optimistic that a policy-based coordination approach can be used to enforce bounds and expectations that increase *interpredictability*, to establish and maintain *common ground*, and to better enable dynamic *directability*.

References

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

[2] Allen, J. F. (1999). Mixed-initiative interaction. *IEEE Intelligent Systems*, September-October, 14-16.

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

[4] Barwise, J., & Perry, J. (1983). *Situations and Attitudes*. Cambridge, MA: MIT Press.

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

[6] Bradshaw, J. M., Beautement, P., Breedy, M. R., Bunch, L., Drakunov, S. V., Feltovich, P. J., Hoffman, R. R., Jeffers, R., Johnson, M., Kulkarni, S., Lott, J., Raj, A., Suri, N., & Uszok, A. (2004). Making agents acceptable to people. In N. Zhong & J. Liu (Ed.), *Intelligent Technologies for Information Analysis: Advances in Agents, Data Mining, and Statistical Learning*. (pp. 355-400). Berlin: Springer Verlag.

[7] Bradshaw, J. M., & Boose, J. H. (1990). Decision analysis techniques for knowledge acquisition: Combining information and preferences using *Aquinas* and *Axotl*. *International Journal of Man-Machine Studies*, 32(2), 121-186.

[8] 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.

[9] Bradshaw, J. M., Feltovich, P., Jung, H., Kulkarni, S., Taysom, W., & Uszok, A. (2004). Dimensions of adjustable autonomy and mixed-initiative interaction. In M. Klusch, G. Weiss, & M. Rovatsos (Ed.), *Computational Autonomy*. (pp. 17-39). Berlin, Germany: Springer-Verlag.

[10] 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.

[11] Bradshaw, J. M., Jung, H., Kulkarni, S., Allen, J., Bunch, L., Chambers, N., Feltovich, P. J., Galescu, L., Jeffers, R., Johnson, M., Taysom, W., & Uszok, A. (2004). Toward trustworthy adjustable autonomy and mixed-initiative in KAoS. *Proceedings of the Autonomous Agents and Multi-Agent Systems (AAMAS) 2004 Trust Workshop*. New York City, NY, pp. 9-20,

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

[13] Bradshaw, J. M., Suri, N., Breedy, M. R., Canas, A., Davis, R., Ford, K. M., Hoffman, R., Jeffers, R., Kulkarni, S., Lott, J., Reichherzer, T., & Uszok, A. (2002). Terraforming cyberspace. In D. C. Marinescu & C. Lee (Ed.), *Process Coordination and Ubiquitous Computing*. (pp. 165-185). Boca Raton, FL: CRC Press. Updated and expanded version of an article that originally appeared in IEEE Intelligent Systems, July 2001, pp. 49-56.

[14] 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.

[15] Cartwright, D., & Zander, A. (1968). *Group Dynamics: Research ant Theory*. (Third ed.), New York City, NY: Harper and Row.

[16] 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.

[17] Clancey, W. J. (2004). Roles for agent assistants in field science: Understanding personal projects and collaboration. *IEEE Transactions*

on Systems, Man, and Cybernetics—Part C: Applications and Reviews, 32(2).

[18] Clark, H. H. (1996). *Using Language*. Cambridge, UK: Cambridge University Press.

[19] Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Ed.), *Perspectives on Socially Shared Cognition.* Washington, D.C.: American Psychological Association.

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

[21] Cypher, A. (Ed.). (1993). *Watch What I Do: Programming by Demonstration.* Cambridge, MA: MIT Press.

[22] 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.

[23] Devlin, K. (1991). *Logic and Information*. Cambridge, England: Cambridge University Press.

[24] 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,

[25] Falcone, R., & Castelfranchi, C. (2002). Adjustable social autonomy.

[26] Feltovich, P., Bradshaw, J. M., Jeffers, R., Suri, N., & Uszok, A. (2004). Social order and adaptability in animal and human cultures as an analogue for agent communities: Toward a policy-based approach. In *Engineering Societies in the Agents World IV*. Berlin, Germany: Springer-Verlag.

[27] Feltovich, P. J., Hoffman, R. R., Woods, D. D., & Roesler, A. (2004). Keeping it too simple: How the reductive tendency affects cognitive engineering. *IEEE Intelligent Systems*, 19(3), 90-94.

[28] 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,

[29] Fitts, P. M. (Ed.). (1951). *Human Engineering for an Effective Air Navigation and Traffic Control System.* Washington, D.C.: National Research Council.

[30] Geertz, C. (1973). *The Interpretation of Cultures*. New York, NY: Basic Books.

[31] Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston, MA: Houghton Mifflin.

[32] 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,

[33] Grosz, B. J. (1996). Collaborative systems. *AI Magazine*, 17(2), 67-85.

[34] 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.

[35] 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,

[36] 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.

[37] Klein, G. (2001). Features of team coordination. In M. McNeese, M. R. Endsley, & E. Salas (Ed.), *New Trends in Cooperative Activities.* (pp. 68-95). Santa Monica, CA: HFES.

[38] Klein, G., Feltovich, P. J., Bradshaw, J. M., & Woods, D. D. (2004). Common ground and coordination in joint activity. In W. B. Rouse & K. R. Boff (Ed.), *Organizational Simulation.* (pp. in press). New York City, NY: John Wiley.

[39] Levinson, S. C. (1992). Activity types and language. In P. Drew & J. Heritage (Ed.), *Talk at Work*. (pp. 66-100). Cambridge, England: Cambridge University Press.

[40] Lieberman, H. (Ed.). (2001). *Your Wish is My Command: Programming By Example.* San Francisco, CA: Morgan Kaufmann.

[41] 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.

[42] Lott, J., Bradshaw, J. M., Uszok, A., & Jeffers, R. (2004). KAoS policy management for control of security mechanisms in a large-scale distributed system. (pp. submitted for publication).

[43] 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.

[44] Maheswaran, R. T., Tambe, M., Varakantham, P., & Myers, K. (2004). Adjustable autonomy challenges in personal assistant agents: A position paper. In M. Klusch, G. Weiss, & M. Rovatsos (Ed.), *Computational Autonomy.* (pp. in press). Berlin, Germany: Springer.

[45] Malin, J. T., Schreckenghost, D. L., Woods, D. D., Potter, S. S., Johannesen, L., Holloway, M., & Forbus, K. D. (1991). *Making intelligent systems team players: Case studies and design issues (NASA Technical Memorandum 104738)*. NASA Johnson Space Center,

[46] Norman, D. A. (1988). *The Psychology of Everyday Things*. New York: Basic Books.

[47] 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.

[48] Norman, D. A. (1999). Affordance, conventions, and design. *Interactions*, May, 38-43.

[49] Olson, G. M., Malone, T. W., & Smith, J. B. (2001). *Coordination Theory and Collaboration Technology*. Mahwah, NJ: Lawrence Erlbaum Associates.

[50] Rich, C., Sidner, C., & Lesh, N. (2001). COLLAGEN: Applying collaborative discourse theory. *AI Magazine*.

[51] Shalin, V. L., Geddes, N. D., Bertram, D., Szczepkowski, M. A., & DuBois, D. (1997). Expertise in dynamic, physical task domains. In P. J. Feltovich, K. M. Ford, & R. R. Hoffman (Ed.), *Expertise in Context: Human and Machine*. (pp. 195-217). Menlo Park, CA: AAAI/MIT Press.

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

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

[54] Smith, W. J. (1995). The biological bases of social attunement. *Journal of Contemporary Legal Issues*, 6.

[55] Suri, N., Carvalho, M., Bradshaw, J. M., Breedy, M. R., Cowin, T. B., Groth, P. T., Saavendra, R., & Uszok, A. (2003). Mobile code for policy enforcement. *Policy 2003*. Como, Italy,

[56] Tonti, G., Bradshaw, J. M., Jeffers, R., Montanari, R., Suri, N., & Uszok, A. (2003). Semantic Web languages for policy representation and reasoning: A comparison of KAoS, Rei, and Ponder. In D. Fensel, K. Sycara, & J. Mylopoulos (Ed.), *The Semantic Web—ISWC 2003. Proceedings of the Second International Semantic Web Conference, Sanibel Island, Florida, USA, October 2003, LNCS 2870*. (pp. 419-437). Berlin: Springer.

[57] Tuckman, B. (1965). Developmental sequence in small groups. *Psychological Bulletin*, 63, 384-399.

[58] Uszok, A., Bradshaw, J. M., & Jeffers, R. (2004). KAoS: A policy and domain services framework for grid computing and semantic web services. *Proceedings of the Second International Conference on Trust Management*. Oxford, England,

[59] Uszok, A., Bradshaw, J. M., Jeffers, R., Johnson, M., Tate, A., Dalton, J., & Aitken, S. (2004). Policy and contract management for semantic web services. *AAAI 2004 Spring Symposium Workshop on Knowledge Representation and Ontology for Autonomous Systems*. Stanford University, CA, AAAI Press,

[60] Uszok, A., Bradshaw, J. M., Jeffers, R., Johnson, M., Tate, A., Dalton, J., & Aitken, S. (2004). Policy and contract management for semantic web services. *IEEE Intelligent Systems*, 26-35.

[61] Woods, D. D. (2002). Steering the reverberations of technology change on fields of practice: Laws that govern cognitive work. *Proceedings of the Annual Meeting of the Cognitive Science Society*.