Ten Challenges for Making Automation a "Team Player" in Joint Human-Agent Activity

Gary Klein, *Klein Associates*
David D. Woods, *Cognitive Systems Engineering Laboratory*
Jeffrey M. Bradshaw, Robert R. Hoffman, and Paul J. Feltovich,
Institute for Human and Machine Cognition

We propose 10 challenges for making automation components into effective "team players" when they interact with people in significant ways. Our analysis is based on some of the principles of human-centered computing that we have developed individually and jointly over the years, and is adapted from a more comprehensive examination of common ground and coordination.[1]

Requirements for joint activity among people

We define *joint activity* as an extended set of actions that are carried out by an ensemble of people who are coordinating with each other.[1,2]

Joint activity involves at least four basic requirements. All the participants must

- Enter into an agreement, which we call a *Basic Compact*, that the participants intend to work together
- Be mutually predictable in their actions
- Be mutually directable
- Maintain common ground

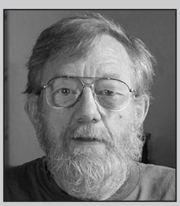
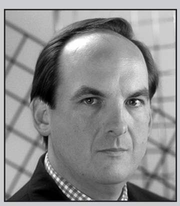
The Basic Compact

To carry out joint activity, each party effectively enters into a Basic Compact—an agreement (often tacit) to facilitate coordination, work toward shared goals, and prevent breakdowns in team coordination. This Compact involves a commitment to some degree of goal alignment. Typically this entails one or more participants relaxing their own shorter-term goals in order to permit more global and long-term team goals to be addressed. These longer-term goals might be shared (for example, a relay team) or individual (such as highway drivers wanting to ensure their own safe journeys).

The Basic Compact is not a once-and-for-all prerequisite to be satisfied but rather has to be continuously reinforced or renewed. It includes an expectation that the parties will repair faulty mutual knowledge, beliefs, and assumptions when these are detected. Part of achieving coordination is investing in those actions that enhance the Compact's integrity as well as being sensitive to and counteracting those factors that could degrade it.

For example, remaining in a Compact during a conversation manifests in the process of accepting turns, relating understandings, detecting the need for and engaging in repair, displaying a posture of interest, and the like. When these sorts of things aren't happening, we might infer that one or more of the parties isn't wholeheartedly engaged. The Compact requires that if one party intends to drop out of the joint activity, he or she must signal this to the other parties. Breakdowns occur when a party abandons the team without clearly signaling his or her intentions to others.

While in traffic, drivers might have defensible motives for rejecting a Compact about following the rules of the road, as when they're responding to an emergency by rushing someone to the nearest hospital. At such times, drivers might turn on their emergency blinkers to signal to other drivers that their actions are no longer as predictable. But in most kinds of joint activity, the agreement itself is tacit, and the partners depend on more subtle signals to convey that they are or aren't continuing in the joint activity. In a given context, sophisticated protocols might develop to acknowledge receipt of a signal, transmit some construal of a signal's meaning back to the sender, indicate preparation for consequent acts, and so forth.

Editors: Robert R. Hoffman, Patrick J. Hayes, and Kenneth M. Ford
Institute for Human and Machine Cognition, University of West Florida
rhoffman@ai.uwf.edu

1541-1672/04/$20.00 © 2004 IEEE
Published by the IEEE Computer Society

Mutual predictability

For effective coordination to take place during the course of the joint activity, team members rely on the existence of a reasonable level of mutual predictability. In highly interdependent activities, planning our own actions (including coordination actions) becomes possible only when we can accurately predict what others will do. Skilled teams become mutually predictable through shared knowledge and idiosyncratic coordination devices developed through extended experience in working together. Bureaucracies with high turnover compensate for lack of shared experience by substituting explicit, predesigned, structured procedures and expectations.

Directability

Team members must also be directable. This refers to the capacity for deliberately assessing and modifying other parties' actions in a joint activity as conditions and priorities change.[3] Effective coordination requires participants' adequate responsiveness to the others' influence as the activity unfolds.

Common ground

Finally, effective coordination requires establishing and maintaining common ground.[4] Common ground includes the pertinent knowledge, beliefs, and assumptions that the involved parties share. Common ground enables each party to comprehend the messages and signals that help coordinate joint actions. Team members must be alert for signs of possible erosion of common ground and take preemptive action to forestall a potentially disastrous breakdown of team functioning.

As an example, we had occasion to observe an Army exercise. During the exercise, a critical event occurred and was entered into the shared large-format display of the "common operating picture." The brigade commander wasn't sure that one of his staff members had seen the change, so he called that person because he felt it was important to manage his subordinate's attention and because the technology didn't let him see if the staff member had noticed the event. The commander had to act like an aide to ensure that the staff member had seen a key piece of information. Special language, often used in noisy, confusing environments (such as "acknowledge" and "roger that"), serves the same function.

Ten challenges

Many researchers and system developers have been looking for ways to make automated systems team players.[3] A great deal of the current work in the software and robotic-agent research communities involves determining how to build automated systems with sophisticated team player qualities.[5–7] In contrast to early research that focused almost exclusively on how to make agents more autonomous, much current agent research seeks to understand and satisfy requirements for the basic aspects of joint activity, either within multiagent systems or as part of human-agent teamwork.

Given the widespread demand for increasing the effectiveness of team play for complex systems that work closely and collaboratively with people, a better understanding of the major challenges is important.

> Agents must also "understand" and accept the enterprise's joint goals, their roles in the collaboration, and the need for maintaining common ground.

A Basic Compact

Challenge 1: To be a team player, an intelligent agent must fulfill the requirements of a Basic Compact to engage in common-grounding activities.

A common occurrence in joint action is when an agent fails and can no longer perform its role. General-purpose agent teamwork models typically entail that the struggling agent notify each team member of the actual or impending failure.[8]

Looking beyond current research and machine capabilities, not only do agents need to be able to enter into a Basic Compact, they must also "understand" and accept the enterprise's joint goals, understand and accept their roles in the collaboration and the need for maintaining common ground, and be capable of signaling if they're unable or unwilling to fully participate in the activity.

Adequate models

Challenge 2: To be an effective team player, intelligent agents must be able to adequately model the other participants' intentions and actions vis-à-vis the joint activity's state and evolution—for example, are they having trouble? Are they on a standard path proceeding smoothly? What impasses have arisen? How have others adapted to disruptions to the plan?

In the limited realm of what today's agents can communicate and reason about among themselves, there's been some limited success in the development of theories and implementations of multiagent cooperation not directly involving humans. The key concept here usually involves some notion of shared knowledge, goals, and intentions that function as the glue that binds agents' activities together.[8] By virtue of a largely reusable explicit formal model of shared "intentions," multiple agents try to manage general responsibilities and commitments to each other in a coherent fashion that facilitates recovery when unanticipated problems arise.

Addressing human-agent teamwork presents a new set of challenges and opportunities for agent researchers. No form of automation today or on the horizon can enter fully into the rich forms of Basic Compact that are used among people.

Predictability

Challenge 3: Human-agent team members must be mutually predictable.

To be a team player, an intelligent agent—like a human—must be reasonably predictable and reasonably able to predict others' actions. It should act neither capriciously nor unobservably, and it should be able to observe and correctly predict its teammates' future behavior. Currently, however, agents' "intelligence" and autonomy work directly against the confidence that people have in their predictability. Although people will rapidly confide tasks to simple deterministic mechanisms whose design is artfully made transparent, they are usually reluctant to trust complex agents to the same degree.[9] Ironically, by making agents more adaptable, we might also make them less predictable. The more a system takes the initiative to adapt to its operator's working style, the more reluctant operators might be to adapt their own behavior because of the confusions these adaptations might create.[10]

Directability

Challenge 4: Agents must be directable.

The nontransparent complexity and inadequate directability of agents can be a formula for disaster. In response to this concern, agent researchers have focused increasingly on developing means for controlling aspects of agent autonomy in a fashion that can be both dynamically specified and easily understood—that is directability.[3,11] Policies are a means to dynamically regulate a system's behavior without changing code or requiring the cooperation of the components being governed.[6,9] Through policy, people can precisely express bounds on autonomous behavior in a way that's consistent with their appraisal of an agent's competence in a given context. Their behavior becomes more predictable with respect to the actions controlled by policy. Moreover, the ability to change policies dynamically means that poorly performing agents can be immediately brought into compliance with corrective measures.

Revealing status and intentions

Challenge 5: Agents must be able to make pertinent aspects of their status and intentions obvious to their teammates.

Classic results have shown that the highest levels of automation on the flight deck of commercial jet aircraft (Flight Management Systems or FMSs) often leave commercial pilots baffled in some situations, wondering what the automation is currently doing, why it's doing that, and what it will do next.[12] To make their actions sufficiently predictable, agents must make their own targets, states, capacities, intentions, changes, and upcoming actions obvious to the people and other agents that supervise and coordinate with them.[13] This challenge runs counter to the advice sometimes given to automation developers to create systems that are barely noticed. We are asserting that people need a model of the machine as an agent participating in the joint activity.[14] People can often effectively use their own thought processes as a basis for inferring the way their teammates are thinking, but this self-referential heuristic is not usually effective in working with agents.

Interpreting signals

Challenge 6: Agents must be able to observe and interpret pertinent signals of status and intentions.

Sending signals isn't enough. The agents that receive signals must be able to interpret the signals and form models of their teammates. This is consistent with the Mirror-Mirror principle of HCC: *Every participant in a complex sociotechnical system will form a model of the other participant agents as well as a model of the controlled process and its environment.*[15] The ideal agent would grasp the significance of such things as pauses, rapid pacing, and public representations that help humans mark the coordination activity. Few existing agents are intended to read their operator teammates' signals with any degree of substantial understanding, let alone nuance. As a result, the devices can't recognize the operator's stance, much less appreciate the operator's knowledge, mental models, or goals, given the evolving state of the plan in progress and the world being controlled.

Charles Billings[16] and David Woods[17] argue that an inherent asymmetry in coordinative competencies between people and machines will always create difficulties for designing human-agent teams. Nevertheless, some researchers are exploring ways to stretch agents' performance to reduce this asymmetry as far as possible, such as exploiting and integrating available channels of communication from the agent to the human and, conversely, sensing and inferring the human's cognitive state through a range of physiological measures in real time. Similarly, a few research efforts are taking seriously the agent's need to interpret the physical environment. If they accomplish nothing more, efforts such as these can help us appreciate the difficulty of this problem.

> **To be a team player, an intelligent agent—like a human—must be reasonably predictable and reasonably able to predict others' actions.**

Goal negotiation

Challenge 7: Agents must be able to engage in goal negotiation.

In many common situations, participants must be able to enter into goal negotiation, particularly when the situation changes and the team has to adapt. As required, intelligent agents must convey their current and potential goals so that appropriate team members can participate in the negotiations.

If agents are unable to readily represent, reason about, or modify their goals, they will interfere with coordination and the maintenance of common ground. Traditional planning technologies for agents typically take an autonomy-centered approach, with representations, mechanisms, and algorithms that have been designed to ingest a set of goals and produce output as if they can provide a complete plan that handles all situations. This approach isn't compatible with what we know about optimal coordination in human-agent interaction.

Collaboration

Challenge 8: Support technologies for planning and autonomy must enable a collaborative approach.

A collaborative autonomy approach assumes that the processes of understanding, problem solving, and task execution are necessarily incremental, subject to negotiation, and forever tentative.[18] Thus, every element of an "autonomous" system will have to be designed to facilitate the kind of give-and-take that quintessentially characterizes natural and effective teamwork among groups of people.

James Allen and George Ferguson's research on collaboration management agents is a good example.[5] CMAs are designed to support human-agent, human-human, and agent-agent interaction and collaboration within mixed human-robotic teams. They interact with individual agents to

- Maintain an overall picture of the current situation and status of the overall plan as completely as possible based on available reports
- Detect possible failures that become more likely as the plan execution evolves, and invoke replanning
- Evaluate the viability of proposed changes to plans by agents
- Manage replanning when situations exceed individual agents' capabilities, including recruiting more capable agents to perform the replanning
- Manage the retasking of agents when changes occur

- Adjust their communications to the agents' capabilities (for example, graphical interfaces work well for a human but wouldn't help most agents)

Because the team members will be in different states depending on how much of their original plan they've executed, CMAs must support further negotiation and replanning at runtime.

Attention management

Challenge 9: Agents must be able to participate in managing attention.

As part of maintaining common ground during coordinated activity, team members direct each other's attention to the most important signals, activities, and changes. They must do this in an intelligent and context-sensitive manner, so as not to overwhelm others with low-level messages containing minimal signals mixed with a great deal of distracting noise.

Relying on their mental models of each other, responsible team members expend effort to appreciate what each other needs to notice, within the context of the task and the current situation.[19] Automation can compensate for trouble (for example, asymmetric lift due to wing icing), but currently does so invisibly. Crews can remain unaware of the developing trouble until the automation nears the limits of its authority or capability to compensate. As a result, the crew might take over too late or be unprepared to handle the disturbance once they take over, resulting in a bumpy transfer of control and significant control excursions. This general problem has been a part of several aviation incident and accident scenarios.

It will push the limits of technology to get the machines to communicate as fluently as a well-coordinated human team working in an open, visible environment. The automation will have to signal when it's having trouble and when it's taking extreme action or moving toward the extreme end of its range of authority. Such capabilities will require interesting relational judgments about agent activities: How does an agent tell when another team member is having trouble performing a function but has not yet failed? How and when does an agent effectively reveal or communicate that it's moving toward its limit of capability?

Adding threshold-crossing alarms is the usual answer to these questions in automation design. However, in practice, rigid and context-insensitive thresholds will typically be crossed too early (resulting in an agent that speaks up too often, too soon) or too late (resulting in an agent that's too silent, speaking up too little). However, focusing on the basic functions of joint activity rather than machine autonomy has already produced some promising successes.[20]

Cost control

Challenge 10: All team members must help control the costs of coordinated activity.

The Basic Compact commits people to coordinating with each other and to incurring the costs of providing signals, improving predictability, monitoring the others' status, and so forth. All these take time and energy. These coordination costs can easily get out of hand, so the partners in a coordination transaction must do what they reasonably can to keep coordination costs down. This is a tacit expectation—to try to achieve economy of effort. Coordination requires continuing investment and hence the power of the Basic Compact—a willingness to invest energy and accommodate others, rather than just performing alone in one's narrow scope and subgoals. Coordination doesn't come for free, and coordination, once achieved, doesn't allow us to stop investing. Otherwise, the coordination breaks down.

> How does an agent tell when a team member is having trouble performing a function but hasn't yet failed? When does an agent communicate that it's moving toward its limit of capability?

Keeping coordination costs down is partly, but only partly, a matter of good human-computer interface design. More than that, the agents must conform to the operators' needs rather than require operators to adapt to them. Information hand-off, which is a basic exchange during coordination involving humans and agents, depends on common ground and mutual predictability. As the notions of HCC suggest, agents must become more understandable, more predictable, and more sensitive to people's needs and knowledge.

The 10 challenges we've presented can be viewed in different lights:

- As a blueprint for designing and evaluating intelligent systems—requirements for successful operation and the avoidance or mitigation of coordination breakdowns.
- As cautionary tales about the ways that technology can disrupt rather than support coordination: Simply relying on explicit procedures, such as common operating pictures, isn't likely to be sufficient.
- As the basis for practicable human-agent systems. All the challenges have us walking a fine line between the two views of AI: the traditional one that AI's goal is to create systems that emulate human capabilities, versus the nontraditional human-centered computing goal—to create systems that extend human capabilities, enabling people to reach into contexts that matter for human purposes.

We can imagine in the future that some agents will be able to enter into some form of a Basic Compact, with diminished capability.[6] Agents might eventually be fellow team members with humans in the way a young child or a novice can be—subject to the consequences of brittle and literal-minded interpretation of language and events, limited ability to appreciate or even attend effectively to key aspects of the interaction, poor anticipation, and insensitivity to nuance. In the meantime, we hope you might use the 10 challenges we've outlined to guide research in the design of team and organizational simulations that seek to capture coordination breakdowns and other features of joint activity. Through further research, restricted types of Basic Compacts might be created that could be suitable for use in human-agent systems.

Acknowledgments

Klein Associates, Ohio State University, and the Institute for Human and Machine Cognition prepared this work through the support of the Advanced Decision Architectures Collaborative Technology Alliance, sponsored by the US Army Research Laboratory under cooperative agreement DAAD19-01-2-0009. David Woods' participation was also made possible by an IBM Faculty Award.

References

1. G. Klein et al., "Common Ground and Coordination in Joint Activity," to be published in *Organizational Simulation*, W.R. Rouse and K.B. Boff, eds., John Wiley & Sons, 2005.

2. H. Clark, *Using Language,* Cambridge Univ. Press, 1996.

3. K. Christoffersen and D.D. Woods, "How to Make Automated Systems Team Players," *Advances in Human Performance and Cognitive Eng. Research*, vol. 2, 2002, pp. 1–12.

4. H.H. Clark and S.E. Brennan, "Grounding in Communication," *Perspectives on Socially Shared Cognition*, L.B. Resnick, J.M. Levine, and S.D. Teasley, eds., Am. Psychological Assoc., 1991.

5. J.F. Allen and G. Ferguson, "Human-Machine Collaborative Planning," *Proc. NASA Planning and Scheduling Workshop,* NASA, 2002.

6. J.M. Bradshaw et al., "Dimensions of Adjustable Autonomy and Mixed-Initiative Interaction," *Agents and Computational Autonomy: Potential, Risks, and Solutions*, M. Nickles, M. Rovatsos, and G. Weiss, eds., LNCS 2969, Springer-Verlag, 2004, pp. 17–39.

7. M. Tambe et al., "Teamwork in Cyberspace: Using TEAMCORE to Make Agents Team-Ready," *Proc. AAAI Spring Symp. Agents in Cyberspace*, AAAI Press, 1999, pp. 136–141.

8. P.R. Cohen and H.J. Levesque, "Teamwork," *Nous*, vol. 25, 1991, pp. 487–512.

9. J.M. Bradshaw et al., "Making Agents Acceptable to People," *Intelligent Technologies for Information Analysis: Advances in Agents, Data Mining, and Statistical Learning*, N. Zhong and J. Liu, eds., Springer-Verlag, 2004, pp. 361–400.

10. G. Klein, *The Power of Intuition*, Currency Book/Doubleday, 2004.

11. K. Myers and D. Morley, "Directing Agents," *Agent Autonomy*, H. Hexmoor, C. Castelfranchi, and R. Falcone, eds., Kluwer Academic Publishers, 2003, pp. 143–162.

12. D.D. Woods and N. Sarter, "Learning from Automation Surprises and Going Sour Accidents," *Cognitive Engineering in the Aviation Domain*, N. Sarter and R. Amalberti, eds., Lawrence Erlbaum, 2000, pp. 327–254.

13. P.J. Feltovich et al., "Social Order and Adaptability in Animal and Human Cultures as Analogues for Agent Communities: Toward a Policy-Based Approach," *Engineering Societies in the Agents World IV*, LNAI 3071, Springer-Verlag, 2004, pp. 21–48.

14. D.A. Norman, "The 'Problem' with Automation: Inappropriate Feedback and Interaction, Not Over-Automation," *Philosophical Trans. Royal Soc. London*, vol. 327, 1990, pp. 585–593.

15. R.R. Hoffman and D.D. Woods, *The Theory of Complex Cognitive Systems*, tech. report, Inst. for Human and Machine Cognition, Pensacola, Fla., 2004.

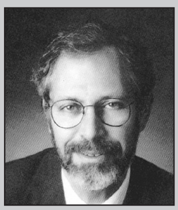
16. C.E. Billings, *Aviation Automation: The Search for a Human-Centered Approach*, Lawrence Erlbaum, 1996.

17. D.D. Woods, "Steering the Reverberations of Technology Change on Fields of Practice: Laws That Govern Cognitive Work," *Proc. 24th Ann. Meeting Cognitive Science Soc.*, Lawrence Erlbaum, 2002, pp. 14–17; http://csel.eng.ohio-state.edu/laws.

18. J.M. Bradshaw et al., "Teamwork-Centered Autonomy for Extended Human-Agent Interaction in Space Applications," *Proc. AAAI Spring Symp. Interaction between Humans and Autonomous Systems over Extended Operation*, AAAI Press, 2004, pp. 136–140.

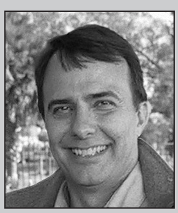
19. N. Sarter and D.D. Woods, "Team Play with a Powerful and Independent Agent: A Full Mission Simulation," *Human Factors*, vol. 42, 2000, pp. 390–402.

20. C.-Y. Ho et al., "Not Now: Supporting Attention Management by Indicating the Modality and Urgency of Pending Tasks," to be published in *Human Factors*.



Gary Klein is chief scientist at Klein Associates. Contact him at Klein Associates, 1750 Commerce Center Blvd. North, Fairborn, OH 45324; gary@decisionmaking.com.

David D. Woods is a professor at the Institute for Ergonomics at Ohio State University. Contact him at the Cognitive Systems Eng. Laboratory, 210 Baker Systems, 1971 Neil Ave., Ohio State Univ., Columbus, OH 43210; woods.2@osu.edu.



Jeffrey M. Bradshaw is a senior research scientist at the Institute for Human and Machine Cognition. Contact him at IHMC, 40 S. Alcaniz St., Pensacola, FL 32502; jbradshaw@ihmc.us.

Robert R. Hoffman is a research scientist at the Institute for Human and Machine Cognition. Contact him at IHMC, 40 Alcaniz St., Pensacola, FL 32501; rhoffman@ihmc.us.

Paul J. Feltovich is a research scientist at the Institute for Human and Machine Cognition. Contact him at IHMC, 40 S. Alcaniz St., Pensacola, FL 32502; pfeltovich@ihmc.us.