

Human-Agent-Robot Teamwork

Jeffrey M. Bradshaw, *Florida Institute for Human and Machine Cognition*

Virginia Dignum and Catholijn Jonker, *Delft University of Technology*

Maarten Sierhuis, *Ejenta*

Researchers and developers continue to pursue increasingly sophisticated roles for autonomous systems. Whether they are working within networked systems as software agents or embedded in robots and unmanned vehicles, what makes these systems valuable is their intelligent, active, and adaptive nature.

These qualities are often characterized in intelligent systems literature by the word “autonomy”—a catch-all label that highlights the qualities of self-directedness and self-sufficiency in task performance.

Thinking about Autonomy

Strictly speaking, though, the term “autonomous system” is a misnomer. Autonomy is not a property of a system, but rather the result of an interaction between the system, the task, and the situation. No system—and, for that matter, no person—can perform

autonomously in every task and situation. On the other hand, even the simplest machine can function autonomously if the task and context are sufficiently constrained.

Much of the early research on autonomous systems was motivated by situations in which autonomous systems had to “replace” human participation, thus minimizing the need for considering the human aspects of such solutions. For example, one of the earliest high-consequence applications of sophisticated agent technologies was in NASA’s Remote Agent Architecture (RAA),

designed to direct the activities of unmanned spacecraft engaged in distant planetary exploration. RAA was expressly designed for use in human-out-of-the-loop situations where response latencies in the transmission of round-trip control sequences would have impaired a spacecraft's ability to respond to urgent problems or to take advantage of unexpected scientific opportunities.

Since those early days, most autonomy researchers have continued to pursue their work in a technology-centric fashion, as if full autonomy—complete independence and self-sufficiency of each system—were the holy grail. However, reflection on the nature of human work reveals the shortsightedness of such a singular focus: what could be more troublesome to a group of individuals engaged in dynamic, fast-paced, real-world collaboration than a colleague who is perfectly able to perform tasks alone but lacks the skills required to coordinate his or her activities with those of others?

In view of these shortcomings, interest has grown in the topic of “cooperative” or “collaborative” autonomy. Unfortunately, this research usually imagines collaboration only among the autonomous systems themselves, regrettably excluding humans as potential teammates. For example, the United States Department of Defense Unmanned Systems Roadmap established the goal of pursuing “greater autonomy in order to improve the ability of unmanned systems to operate independently, either individually or collaboratively, to execute complex missions in a dynamic environment.” Similar briefs have complained that because unmanned vehicles are not truly autonomous, their operation requires substantial input from remote operators. They ask whether additional research in cooperative

autonomous behavior—referring to cooperation between the autonomous systems without any human element—could address this “problem.”

Of course, there are situations where the goal of minimizing human involvement is appropriate. However, virtually all of the significant deployments of autonomous systems to date—for example, military unmanned aerial vehicles, NASA rovers, unmanned undersea vehicles for oil spill work, and disaster inspection robots—have involved people in important role. Such involvement was not merely to make up for the current inadequacy of autonomous capabilities, but also because jointly coordinated efforts with humans were—or should have been—intrinsically part of the mission planning and execution itself. Although continuing research to make machines more active, adaptive, and functional is essential, the point of increasing such proficiencies is not merely to make the machines more independent during times when unsupervised activity is desirable or necessary (in other words, to provide autonomy), but also to make them more capable of sophisticated interdependent joint activity with people and other machines when such is required (in other words, to participate in teamwork). The mention of joint activity highlights the need for autonomous systems to support not only fluid orchestration of task handoffs among different people and machines, but also combined participation on shared tasks requiring continuous and close interaction—that is, coactivity.

Historical Perspectives: HABA-MABA to HART

Why has the notion of human-agent-robot teamwork (HART) taken so long to catch on? Some of the reasons are historical. The concept of

automation—which began with the straightforward objective of replacing any task currently performed by a human with a machine that could do the same task better, faster, or cheaper—attracted the notice of early human-factors researchers. Pioneers such as Paul Fitts attempted to systematically characterize the general strengths and weaknesses of humans and machines. 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, known as the “humans are better at/machines are better at” (HABA-MABA) approach (see Figure 1).

Obviously, however, the suitability of a particular human or machine for a particular task might vary over time and in different situations. So, early researchers in adaptive function allocation (in the human-factors community) and adjustable autonomy (in the software agents and robotics communities) hoped to make the shifting of responsibilities between humans and machines dynamic. Of course, machines couldn't take on certain tasks, such as those requiring sophisticated judgment, and humans couldn't do others, such as those requiring ultraprecise movement. But for tasks where human and machine capabilities overlapped—the area of variable task assignment—a series of software-based decision-making schemes were proposed to allow tasks to be allocated according to the availability of the potential performer.

Eventually, it became plain to researchers that things were not as simple as they first appeared. For example, humans and machines share many functions in complex systems; hence the need to consider synergies and conflicts among the various performers of joint actions. Moreover,

it has become clear that function allocation isn't a simple process of transferring responsibilities from one component to another. Automated assistance of whatever kind doesn't simply enhance our ability to perform the task, it changes the nature of the task itself. It's like asking a five-year-old child to help do the dishes—from the point of view of an adult, such "help" doesn't necessarily reduce the effort involved, it merely transforms the work from the physical action of washing the dishes to the cognitive task of monitoring the child.

Historically, the HABA-MABA approach naturally led to research programs that divided up work between humans and machines rather than considering how they could work together. This was all right so long as machines remained simple. However, as automation has become more sophisticated, the nature of its interaction with people has begun to change in profound ways. In envisioning the increasingly substantive interaction of the future, the point is not to think so much about which tasks are best performed by people and which by machines, but rather how tasks can best be shared by both humans and automation working in concert. As far back as 1960, J.C.R. Licklider, the first director of the Information Processing Technology Office of the US Advanced Research Projects Agency (now DARPA), called

this concept *man-computer symbiosis*.¹ To counter the limitations of the Fitts list, which is clearly intended to summarize what humans and machines each do well on their own, Robert Hoffman has summarized the findings of David Woods proposed an "un-Fitts" list (see Table 1), which emphasizes how the competencies of humans and machines can be enhanced through appropriate forms of mutual interaction.²

One important exception to the US funding emphasis on minimizing human involvement in autonomous

systems was the now-completed DARPA Cognitive Assistant that Learns and Organizes (CALO) program. CALO ran from 2003 to 2008 and inspired development of the types of capabilities we see in Apple's Siri personal assistant. Successive versions of Siri may eventually incorporate a range of autonomous capabilities and work more effectively with people to answer questions and take simple actions. However, it's not designed to address the challenging requirements of teamwork and coactivity in complex and large-scale multiagent systems, such as coordinated operations of people with heterogeneous unmanned vehicles, or sensemaking applications such as cyber-situation awareness, where software agents and analysts engage coactively in a progressively converging process to identify emerging threats.

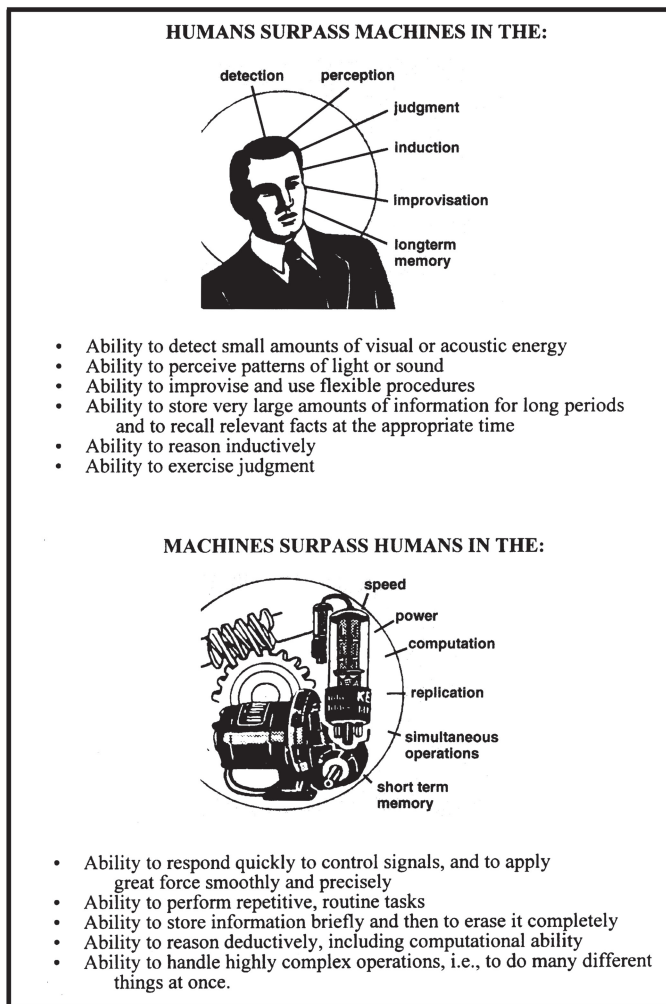


Figure 1. The Fitts HABA-MABA (humans-are-better-at/machines-are-better-at) approach. Reprinted with permission from *Human Engineering for an Effective Air Navigation and Traffic Control System, 1951*, by the National Academy of Sciences, courtesy of the National Academies Press, Washington, D.C.

HART Research Challenges

Today's autonomous systems come in two major varieties:

1. Software agents and networked multiagent systems that help address data-to-decision problems (for example, course-of-action evaluation, sensor integration, and logistics planning), that provide intelligent user interface functionality (personal assistants, sophisticated natural-language query processing, advanced visualization), and that assist in

Table 1. An “un-Fitts” list.

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

monitoring, analyzing, and making sense of complex, uncertain, high-tempo events (cyberdefense, disaster management);

2. Robots and autonomous vehicles with software agent technology embedded in specialized hardware for military, space, security, and commercial applications requiring sophisticated sensors and effectors, and physical mobility. Examples of such applications include complex search-and-rescue activities in dangerous environments, such as battle zones packed with improvised explosive devices or contaminated by nuclear, biological, or chemical agents. Effective human-agent-robot teamwork has important applications off the battlefield as well, such as robots that work alongside doctors on surgical teams, with researchers in labs, or with physically or cognitively challenged populations.

Future applications will increasingly require combinations of both kinds of autonomous systems. Regrettably, the efforts of the research communities for software agents and robotic agents are relatively disjoint,

despite the fact that many research challenges are common to both fields, such as coordinating interdependent activity, establishing and maintaining common ground among team members, and recovering gracefully from individual or team breakdowns.

HART research seeks to bring together the best thinking from these and other allied research communities to advance current and anticipated applications of intelligent human-machine collaboration. Addressing the technology gap created by the past emphasis on making machines self-sufficient, we’re seeing modest efforts to understand and develop capabilities that would allow the participation of humans as first-class citizens in collaboration with autonomous systems. Such capabilities would enable autonomous systems not merely to do things *for* people, but also to work together *with* people and other systems.

To date, autonomous-system designers haven’t sufficiently appreciated the essential role of interdependence in joint human-machine activity. While some approaches to cooperative interaction have become widely known (for example, dynamic function allocation, supervisory

control, adaptive automation, and adjustable autonomy), each of them shares a common flaw: they rely on some notion of “levels of autonomy” as a basis for their effectiveness. The problem with such approaches is their singular focus on managing human-machine work by varying which tasks are assigned to an agent or robot on the basis of some (usually context-free) assessment of its independent capabilities for executing that task. However, decades of studies have shown that successful teamwork in everyday human interaction is largely a matter of managing the context-dependent complexities of interdependence among tasks and teammates. This requirement for interdependence affects not only when and how tasks need to be done but also the sometimes subtle properties of team interaction such as observability, directability, predictability, and the maintenance of common ground. And because the capabilities for teamwork and coactivity interact with autonomy algorithms at a deep level, system design must embed them from the beginning, not layer them on with a thin veneer of user interface widgets after the fact. Systems designed without these considerations are almost always difficult

to repurpose without significant reengineering.

The most tantalizing claim for HART research is that computational frameworks for autonomy that incorporate well-founded sociocognitive theories will demonstrate greater effectiveness, robustness, resilience, and safety in the face of dynamic real-world complexity than will frameworks focusing on autonomy alone. To reduce the overall human footprint in deployment, these systems would take full advantage of capabilities for autonomy when appropriate while also having the additional sense needed to be able to work well with people.

In This Issue

Six articles in this special issue discuss the importance of incorporating HART in the development of autonomous computational actors. Earlier versions of some of these articles were previously presented at the third HART workshop that was held in December 2010 at the Lorentz Center in Leiden, Netherlands.

Each article highlights the importance of interdependence of computational team members and human team members. Rather than viewing the human as a “user,” they treat the human as a member of a team of intelligent actors engaging in joint, coactive tasks. And just as people do, the computational actors need a sociocognitive model of their team members in order to be aware of the context. These six articles describe different approaches to the creation and deployment of a teamwork model incorporating human members.

Machine learning is important in the automatic development of HART models. People learn how to work together from past experience. They learn contextually over time as part of everyday activity, be it

conscious or unconscious, playful or part of an educational activity. The question is: how can computational agents learn most efficiently? We are more willing to accept human errors than errors from our computational team members. For these and other reasons, machine learning is less commonly applied during real-life daily teamwork activities than in research settings. A good example of this is the way Google, Mercedes, BMW, Audi, Toyota and others are creating the autonomous car. “Collaborative Programming by Demonstration in a Virtual Environment” discusses the application of machine learning techniques in a virtual environment that provides the human team member with training scenarios that let a computational agent learn the human’s behavior.

One of the difficulties in the development of HART systems is testing the entire system. “Mixed-Reality Testbeds for Incremental Development of HART Applications” describes an environment in which developers of HART systems can test team interaction. The focus isn’t on learning teamwork models but on more easily testing a complex distributed system of mixed human-automation teams, enabling the inclusion of combinations of simulated human, agent, and robot models.

“Situated Communication for Joint Activity in Human-Robot Teams” discusses what it takes for robots to understand human communication. The approach models robot “experience” as a collection of representations that bridge the gap between low-level sensing and high-level representations. The authors see communication as part of coactivity, connecting what is being said to situations, plans, tasks, capabilities, and roles.

“The Social Landscape: Reasoning on the Social Behavioral Spectrum”

goes one step further to discuss, from a theoretical perspective, how the models of joint activity in robots and computational agents need to go beyond teamwork to include a range of other kinds of engagement. Real teamwork includes the ability of team members to reason about the overall spectrum of social behavior, ranging from altruistic behavior at one end of the spectrum, through cooperation, individualism, and competition, to aggression at the other end.

The last two articles discuss the results of specific HART experiments. “Autonomy and Interdependence in Human-Agent-Robot Teams” describes students playing a version of the well-known Blocks World game, adapted to explore hypotheses about autonomy and teamwork. The authors argue that contrary to common assumptions, increased autonomy in the computational team member doesn’t always lead to better overall performance: instead, it sometimes results in greater opacity and more frequent coordination breakdowns.

Finally, “Incorporating a Robot into an Autism Therapy Team” reports on the use of a robot as part of a real-life therapeutic team for autistic children. The authors show that when designing a robotic team member, making it more autonomous is less important than making it understand the interdependence between the team members and the coactivity of the team as a whole.

The field of HART is still young. The goal of this special issue is to make agent and robot researchers aware of the importance of human-centered design and teamwork. Much remains to be done, and this is but a small selection of topics that are relevant and important to the future of increasingly sophisticated autonomous agent and robots. ■

THE AUTHORS

Jeffrey M. Bradshaw is a senior research scientist at the Florida Institute for Human and Machine Cognition (IHMC), where he leads the research group developing the KaoS policy and domain services framework. He also co-leads the group developing IHMC's Sol Cyber Framework. He is coeditor of *IEEE Intelligent Systems' Human-Centered Computing* department. Bradshaw has a PhD in cognitive science from the University of Washington. Contact him at jbradshaw@ihmc.us.

Virginia Dignum is an associate professor of technology, policy and management at the Delft University of Technology. Her research focuses on the interaction between people and intelligent systems—in particular, the behavior of artificial systems as social actors—and on agent-based models and simulations of organizations. Dignum has a PhD in AI from Utrecht University. Contact her at m.v.dignum@tudelft.nl.

Catholijn Jonker is a professor of electrical engineering, mathematics and computer science at the Delft University of Technology, specializing in man-machine interaction. Her recent publications address cognitive processes and concepts such as trust, negotiation, and the dynamics of individual agents and organizations. Jonker received her PhD in computer science from Utrecht University. Contact her at c.m.jonker@tudelft.nl.

Maarten Sierhuis is the founder and CTO of Ejenta, Inc. His research interests include blending methods of social science and computer science in the engineering of human-centered systems, and the development and application of agent-oriented languages for modeling and simulating work practice. Sierhuis has a PhD in social science informatics from the University of Amsterdam. Contact him at sierhuis@ejenta.com.

References

1. J.C.R. Licklider, "Man-Computer Symbiosis," *Trans. Human Factors in Electronics*, vol. HFE-1, IEEE, 1960, pp. 4–11.
2. 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, Mahwah, N.J., 2002, pp. 14–16.

 Selected CS articles and columns are also available for free at <http://ComputingNow.computer.org>.

“All writers are vain,
selfish and lazy.”

—George Orwell, “Why I Write” (1947)

(except ours!)



The world-renowned IEEE Computer Society Press is currently seeking authors. The CS Press publishes, promotes, and distributes a wide variety of authoritative computer science and engineering texts. It offers authors the prestige of the IEEE Computer Society imprint, combined with the worldwide sales and marketing power of our partner, the scientific and technical publisher Wiley & Sons.

For more information contact Kate Guillemette, Product Development Editor, at kguillemette@computer.org.

 **CS Press**
www.computer.org/cspress