Human-Agent Interaction

Jeffrey M. Bradshaw, Paul Feltovich, Matthew Johnson
Florida Institute for Human and Machine Cognition (IHMC)

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. Pioneering researchers such as Fitts attempted to systematically characterize the general strengths and weaknesses of humans and machines [28]. 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 (fig. 1).

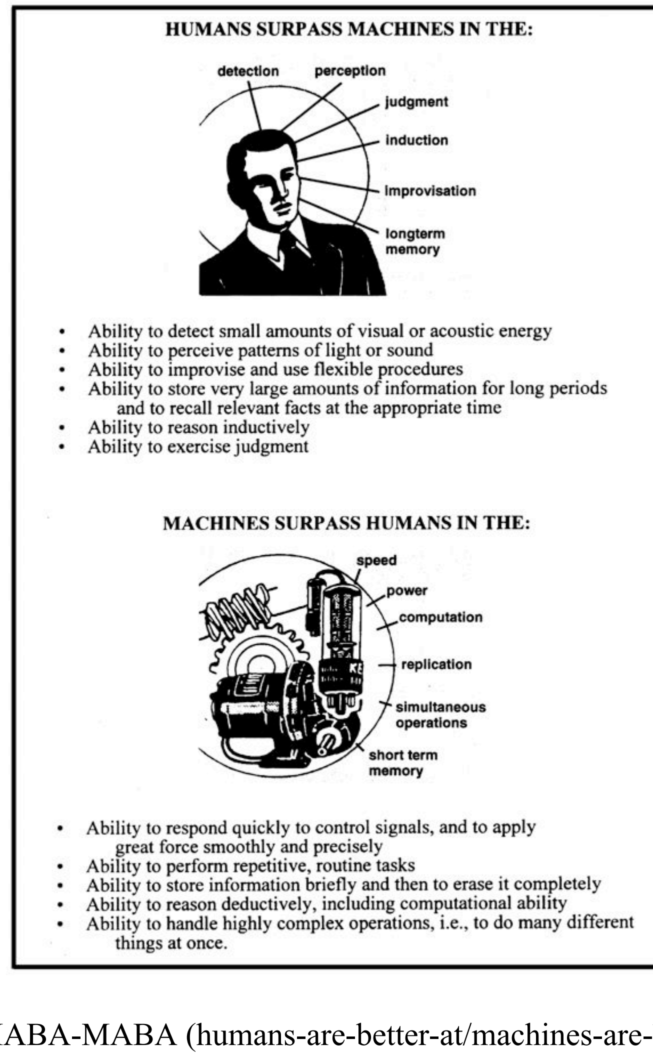

Fig. 1. The Fitts HABA-MABA (humans-are-better-at/machines-are-better-at) approach.

Obviously, however, the suitability of a particular human or machine to take on a particular task may vary over time and in different situations [36]. Hence, early research in adaptive function allocation and adjustable autonomy was undertaken with the hope that shifting of responsibilities between humans and machines could be made dynamic. Of course, certain tasks, such as those requiring sophisticated judgment, could not be shifted to machines, and other tasks, such as those requiring ultra-precise movement, could not be done by humans. But with regard to 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 (fig. 2).

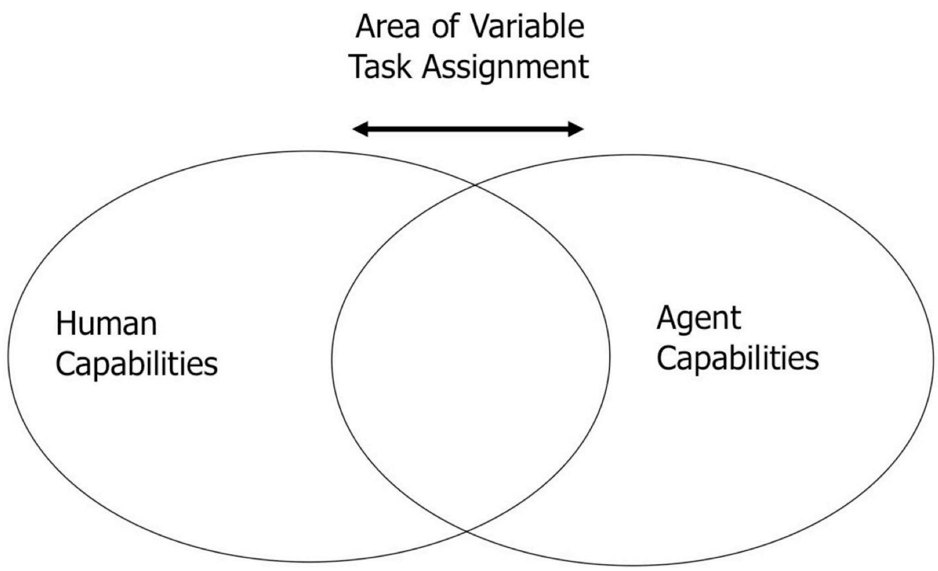

Fig. 2. Perspective of early research in adaptive allocation and adjustable autonomy.

Eventually, 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. 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 [15; 50]. For example, those who have asked 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.

As automation becomes more sophisticated, the nature of its interaction with people will need to change in profound ways. In non-trivial interaction of this sort, the point is not to think so much about which tasks are best performed by humans and which by automation but rather how tasks can best be shared by both humans and automation working in concert [36]. Licklider called this concept *man-computer symbiosis* [43]. In the ultimate form of such symbiosis, human capabilities are transparently augmented by cognitive prostheses—computational systems that leverage and extend human intellectual, perceptual, and collaborative capacities, just as a steam shovel is a sort of muscular prosthesis or eyeglasses are a sort of visual prosthesis [11; 30; 35]. To counter the limitations of the Fitts' list, which is clearly intended to summarize what humans and machines each do well on their own, Hoffman has summarized the findings of Woods in an "un-Fitts list" [38] (table 1), which emphasizes how the competencies of humans and machines can be enhanced through appropriate forms of mutual interaction.

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

Table 1. An "un-Fitts" list [38], © 2002 IEEE.

The Concept of Agents

Though machines that demonstrate superhuman strength, rapid calculation, or extreme agility have proven indispensable in their own right, one of the greatest dreams of scientists and ordinary people has always been a non-human agency whose capabilities might begin to approach those of their fellow beings. The word *robot*, derived from the Czech word for drudgery, captured the public imagination following Karel Capek's 1921 play *RUR: Rossum Universal Robots* (fig. 3).

Fig. 3. Scene from Capek's 1921 play, *Rossum Universal Robots*.

Though automata of various sorts have existed for centuries, it is only since World War II, with the development of computers and control theory, that anything resembling modern agent technology has begun to appear. Computer visionary Alan Kay provided a thumbnail sketch tracing the more recent roots of the idea:

> The idea of an agent originated with John McCarthy in the mid-1950's, and the term was coined by Oliver G. Selfridge a few years later, when they were both at the Massachusetts Institute of Technology. They had in view a system that, when given a goal, could carry out the details of the appropriate computer operations and could ask for and receive advice, offered in human terms, when it was stuck. An agent would be a "soft robot" living and doing its business within the computer's world. [41]

Since the idea of agents was first introduced, people have debated the meaning of the term. The debate on the definition of agenthood will probably never be fully settled: one person's "intelligent agent" is another person's "smart object"; and today's "smart object" is tomorrow's "dumb program" [6, p. 5]. However, by whatever names and definitions we adopt, the systems we interact with a few decades from now will be different in fundamental ways, and will bring new questions to the fore.

"Agents occupy a strange place in the realm of technology," argues Don Norman, "leading to much fear, fiction, and extravagant claims" [51, p. 49]. By their ability to operate independently in complex situations without constant human supervision, agents can perform tasks on a scale that would be impractical or impossible for fully human-in-the-loop approaches to duplicate. On the other hand, this additional autonomy, if unchecked, also has the potential of effecting severe damage if agents are poorly designed, buggy, or malicious. Because ever more powerful intelligent agents will increasingly differ from software that people are accustomed to, we need to take into account social issues no less than the technical ones if the agents we design and build are to be acceptable to people.

Fig. 4. The concept of agents has evoked fear, fiction, and extravagant claims (Kelly Freas: *The Gulf Between*, with permission)

Continues Norman:

> The technical aspect is to devise a computational structure that guarantees that from the technical standpoint, all is under control. This is not an easy task.
>
> The social part of acceptability is to provide reassurance that all is working according to plan.… This is also a non-trivial task. [51, p. 49]

The Emergence of Human-Agent Interaction Research

Much of the early work of researchers in software agents and robotics was motivated by situations in which autonomous systems were envisioned to “replace” human participation, thus minimizing the need to consider the “social” aspects of acceptability. For example, one of the earliest high-consequence applications of sophisticated agent technologies was in NASA’s Remote Agent Architecture (RAA). RAA was designed to be used in situations where response latencies in the transmission of round-trip control sequences from earth would have impaired the satellite’s ability to respond to urgent problems or take advantage of unexpected science opportunities [49]. However, in contrast to autonomous systems that are designed to take humans out of the loop, an increasing number of efforts are being specifically designed to address requirements for close and continuous interaction with people [1; 7; 44; 55].

Specific approaches to human-agent interaction (HAI) have been explored in many forms and with somewhat divergent perspectives. For example, research communities have formed around the topics of interface agents and assistants [16; 21; 39; 44; 45], adjustable autonomy [9; 10; 22; 23; 33; 46; 47], mixed-initiative systems [2; 3; 9; 12; 27], human-agent teamwork [46; 58], and collaboration theory [34; 53].

In this article, we will examine the elements of successful HAI from the perspective of joint activity theory [25; 42], a generalization of Herbert Clark's work in linguistics [18, p. 3]. We will not attempt to provide detailed recommendations or a survey of the voluminous literature, but rather will outline some of the most important principles of HAI based on our own experience.

A Joint Activity Perspective on HAI

The essence of joint activity is interdependence. 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.

Types of Joint Activity

As mentioned previously, interdependence is the essence of joint activity. Thus, it should not be a surprise that different kinds of joint activity can be distinguished according to the types of interdependencies involved.

Co-allocation: This is characterized by interdependence among necessary resources only. Parties have independent goals, and there is no functional coupling of methods. Examples include two groups trying to schedule a conference room they both need to use on a certain day, or simultaneously sharing a wireless network. In sharing, constraints on resource allocation require negotiation.

Cooperation: In cooperation—perhaps better rendered here as "co-operation"—there is interdependence of activities but not of motivations and goals. Often there is also interdependence of resources. Following the last example, two groups trying to conduct their own meetings within the same room at the same time would be a cooperation. So also, interestingly, are competitive games, such as football, where the two teams' actions are clearly interdependent while their aims are not the same and even contrasting.

Collaboration: Shared objectives are the hallmark of collaboration. Teamwork can be seen as a particular form of collaboration. All parties are trying to achieve the same end (mutually defined), and there is also usually interdependence of actions (often involving different roles) and resources. Team members within one team in a football game (or a relay team in track and field) fit this description, as does a group of scholars working together to produce a genuinely multi-authored article on a topic of mutual interest. The more sophisticated collaboration roles—for example, those involving negotiation of complex goals and meanings—are more adeptly handled by humans than by agents. Today's agents, however, have begun to participate in the relatively simpler roles of collaboration support. Notwithstanding the many challenges involved, adult humans and radically less-abled entities (e.g., small children, dogs, video game characters) have shown themselves capable of working together effectively in a variety of situations where a subjective experience of collaboration is often maintained despite the magnitude of their differences [37]. Generally this is due to the ability of humans to rapidly size up and adapt to the limitations of their teammates, an ability we would like to exploit in the design of approaches for HAI.

The Challenge of Human-Agent Coordination in Joint Activity

In a very real sense, the cumulative success of research on agent autonomy can be seen as fueling its own demand for more sophisticated HAI. For example, human interaction with simple teleoperated robotic platforms is confined to whatever actions are necessary to direct the robot from one place to another. The final destination—and, more importantly, the reasons behind the journey—remain completely in the mind of the operator, who stays in more or less continuous contact with the platform. However, the more that navigation and reasoning about how to meet mission objectives are delegated to the robotic platform itself, with the operator providing only intermittent supervisory feedback, the greater the need for effective coordination. This need dramatically increases when there are multiple parties—humans, software agents, and robots—involved.

Clark observes that "a person's processes may be very different in individual and joint actions, even when they appear identical" [18]. For example, he contrasts playing a musical solo versus playing a duet. A major difference between the two is the need for coordination. Malone and Crowston [48] defined coordination as "managing dependencies between activities." For example, any sort of teamwork, which by definition implies interdependence among the players, therefore requires some level of work for each party over and beyond the carrying out of task itself in order to manage its role in coordination. Part of that "extra" work involves each party doing its part to assure that relevant aspects of the agents and the situation are observable at an appropriate level of abstraction and using an effective style of interaction [8].

Although coordination is as much a requirement for joint activity among groups of software agents as it is in HAI, the magnitude of the representational and reasoning gulfs separating humans from agents is much larger [50]. Moreover, because the agent's ability to sense or infer information about the human environment and cognitive context is so limited, agent designers must find innovative ways to compensate for the fact that their agents are not situated in the human world. Brittleness of agent capabilities is difficult to avoid because only certain aspects of the human environment and cognitive context can be represented in the agent, and the representation that is made cannot be "general purpose," but must be optimized for the particular use scenarios the designer originally envisioned. Without sufficient basis for shared situation awareness and mutual feedback, coordination among people and agents simply cannot take place, and, as argued above, this need for shared understanding and feedback increases as the size of the group and the degree of autonomy of the agents increase. This increase in size and complexity changes the very nature of the task and the relationships among participants.

Requirements for Effective Coordination

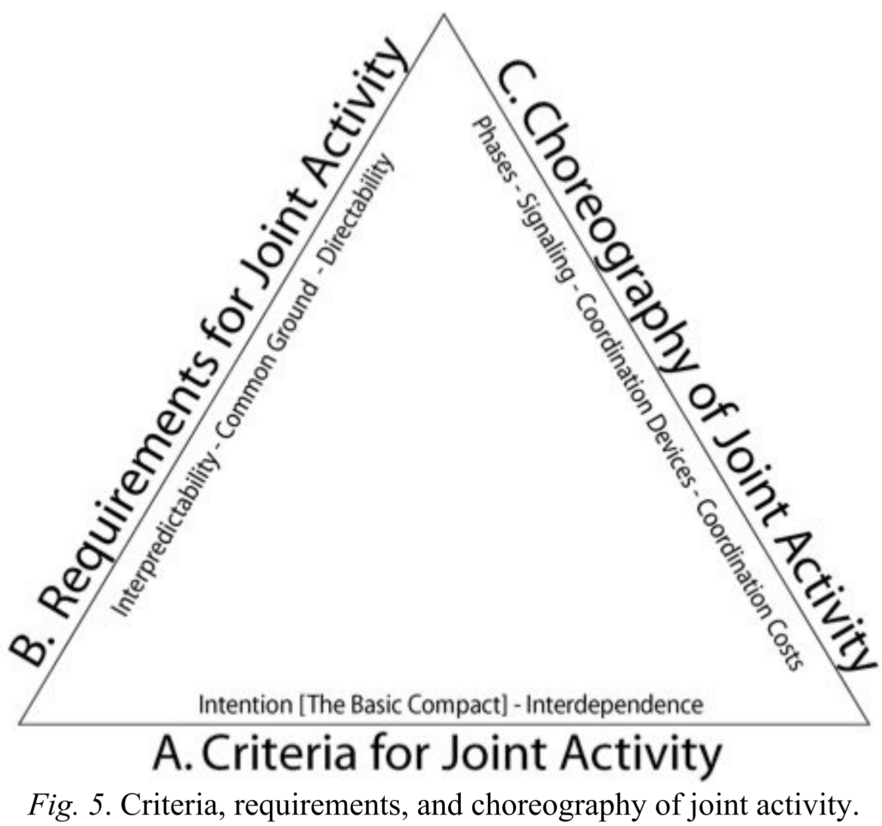

Fig. 5. Criteria, requirements, and choreography of joint activity.

Joint activity theory highlights three major requirements for effective coordination: interpredictability, common ground, and directability [42]:

- *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. On the other hand, bureaucracies with high turnover compensate for such experience by substituting explicit, predesigned structured procedures and expectations relative to formal organizational roles.
- *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 [17]. This includes whatever common ground is already shared 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, 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 [15]. Effective coordination requires adequate responsiveness of each participant to the influence of the others and the requirements of the situation as the activity unfolds. When things go wrong we want to feel assured that there is a mutual commitment to resolve problems in a timely manner.

Following the lead of pioneering researchers such as Geertz [32, pp. 44-46, 67], we have argued that people create and have created cultures and social conventions—albeit in many disparate forms across mankind that can be hard for outsiders to understand—to provide order and predictability that lead to effective coordination [24; 25], including ongoing progress appraisal [26].[1] Order and predictability may have a basis in the simple cooperative act between two people, in which the parties "contract" to engage together in a set of interlinked, mutually beneficial activities. From this simple base, in humans at least, 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). Such diverse regulatory mechanisms can be exploited in HAI to support coordination of complex, interdependent activity [25], as can additional mechanisms discussed next.

The Choreography of Coordination

People coordinate through signals and more complex messages of many sorts (e.g., face-to-face language, expressions, posture). Human signals are also mediated in many ways—for example, through third parties or through machines such as telephones or computers. Hence, direct and indirect party-to-party communication is one form of a "coordination device," in this instance coordination by *agreement*. For example, a group

[1] Even simple forms of animal cooperation seems to bear out such a thesis [56], and we would argue that the more autonomous the agents involved, the more need there is for such regulation and the wider the variety of forms it might take.

of scientists working together on a grant proposal, may simply agree, through e-mail exchanges, to set up a subsequent conference call at a specific date and time. Besides agreement, there are three other common coordination devices [18; 42]:

- *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. 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.

Roles, Regulations, and Organizations in Joint Activity

Roles can be thought of as ways of packaging rights and obligations that go along with the necessary parts that people play in joint activities. Of course, multiple roles can be played by the same actor in a given activity. Knowing one's own roles and the roles of others in a joint activity establishes expectations about how others are likely to interact with us, and how we think we should interact with them. Shoppers expect cashiers to do certain things for them (e.g., total up the items and handle payment) and to treat them in a certain way (e.g., with cheerful courtesy), and cashiers have certain expectations of shoppers. When roles are well understood and regulatory devices are performing their proper function, observers are likely to describe the activity as highly coordinated. On the other hand, violations of the expectations associated with roles and regulatory structures can result in confusion, frustration, anger, and a breakdown in coordination.

Collections of roles are often grouped to form organizations. In addition to regulatory considerations at the level of individual roles, organizations themselves may also add their own rules, standards, traditions, and so forth, in order to establish a common culture that will smooth interaction among parties.

Knowing how roles undergird organizations and how rights and obligations undergird roles helps us understand how organizations can be seen as functional or dysfunctional. Whether hierarchical or heterarchical, fluid or relatively static, organizations are functional only to the extent that their associated regulatory devices and roles generally assist them in facilitating their individual responsibilities and their work in coordinating their actions with others when necessary.

The lesson here for human-agent interaction is that the various roles that different parties assume in their work must include more than simple names for the role and algorithmic behavior to perform their individual tasks. They must also, to be successful, include regulatory structures that define the additional work of coordination associated with that role.

Norms and Policies in Joint Activity

The order needed for agents to engage in joint activity is typically implemented in terms of formalized social regulations. The idea of building strong social regulation into intelligent systems can be traced at least as far back as the 1940s to the science fiction writings of Isaac Asimov [4]. Shoham and Tennenholtz [54] introduced the theme of social "laws" into the agent research community, where investigations have continued under two main headings: *norms* and *policies*. Drawing on precedents in legal theory, social psychology, social philosophy, sociology, and decision theory [60], *norm-based* approaches have grown in popularity [52; 59]. In the multi-agent system research community, Conte and Castelfranchi [20] found that norms were variously described as constraints on behavior, ends or goals, or obligations. For the most part, implementations of norms in multi-agent systems share three basic features:

1. they are designed offline; or
2. they are learned, adopted, and refined through the purposeful deliberation of each agent; and
3. they are enforced by means of incentives and sanctions.

Interest in *policy-based* approaches to multi-agent and distributed systems has also grown considerably in recent years (see, e.g., http://www.policy-workshop.org/). While sharing much in common with norm-based approaches, policy-based perspectives differ in subtle ways. Whereas in everyday English the term *norm* denotes a practice, procedure, or custom regarded as typical or widespread, a *policy* is defined by the American Heritage Online dictionary as a "course of action, guiding principle, or procedure considered expedient, prudent, or advantageous." Thus, in contrast to the relatively descriptive basis and self-chosen adoption (or rejection) of norms, policies tend to be seen as prescriptive and externally-imposed entities. Whereas norms in everyday life emerge gradually from group conventions and recurrent patterns of interaction, policies are consciously designed and put into and out of force at arbitrary times by virtue of explicitly-recognized authority. These differences are generally reflected in the way most policy-based approaches differ from norm-based ones with respect to the three features mentioned above. Policy-based approaches:

1. support dynamic runtime policy changes, and not merely static configurations determined in advance;

2. work involuntarily with respect to the agents, that is, without requiring the agents to consent or even be aware of the policies being enforced; thus aiming to guarantee that even the simplest agents can comply with policy; and
3. wherever possible they are enforced preemptively, preventing buggy or malicious agents from doing harm in advance rather than rewarding them or imposing sanctions on them after the fact.

Fig. 6. Policies constitute an agent's "rules of the road," not its "route plan."

Policy management should not be confused with planning or workflow management, which are related but separate functions. Planning mechanisms are generally *deliberative* (i.e., they reason deeply and actively about activities in support of complex goals), whereas policy mechanisms tend to be *reactive* (i.e., concerned with simple actions triggered by some environmental event) [31, pp. 161-162]. Whereas plans are a unified roadmap for accomplishing some coherent set of objectives, bodies of policy collected to govern some sphere of activity are made up of diverse constraints imposed by multiple potentially-disjoint stakeholders and enforced by mechanisms that are more or less independent from the ones directly involved in planning. Plans tend to be relatively comprehensive, while policies, in our sense, are by nature piecemeal. In short, we might say that while policies constitute the "rules of the road"—providing the stop signs, speed limits, and lane markers that serve to coordinate traffic and minimize mishaps—they are not sufficient to address the problem of "route planning."

Norms and policies can, of course, be combined in agent systems. Typically, however, agent system designers tend to gravitate toward one approach or the other, based on the kinds of agents they are defining and the kinds of problems they are trying to solve. A norm-based approach is always useful when:

- A primary purpose of the agent system is to model the learning and adaptation of norms;
- The norms are not arbitrary constraints, but have a rational basis in repeated experience;
- The results of deliberate violations of regulation is relatively inconsequential.

Implementing regulation through policy is most useful when:

- The application requires predictability and repeatability with respect to the specific agent behavior being regulated;
- The agents themselves are not capable of learning;
- Compliance with regulation within a specified tolerance is essential.

Coactive Design

The term "coactive design" was coined by Johnson as a way of characterizing an approach to HAI that takes *interdependence* as the central organizing principle among people and agents working together in joint activity [40]. Besides implying that two or more parties are participating in an activity, the term "coactive" is meant to convey the reciprocal and mutually constraining nature of actions and effects that are conditioned by coordination. In joint activity, individual participants share an obligation to coordinate, sacrificing to a degree their individual autonomy in the service of progress toward group goals. Below we sketch some of the important considerations that play into coactive design.

- *Teamwork vs. task-work.* Coactive design complements task-focused approaches to HAI such as function allocation, adjustable autonomy, and mixed-initiative interaction. It is more focused on teamwork than task-work. For example, the task-work of playing soccer includes kicking to a target, dribbling, tackling, and tracking the ball and the goal. By way of contrast, the teamwork of soccer focuses on things like allocating players to roles, synchronizing tactics, and sharing information.
- *Mutual affordances and obligations.* Software agents are often described in terms of their role as assistants to people. While this one-way relationship between assistant and the one who is assisted sometimes may be a helpful, it is inadequate for describing joint activity of humans and agents working together. Joint activity, by its nature, implies the greater parity of *mutual* assistance, enabled by intricate webs of complementary, reciprocal affordances and obligations. Human speech would be useless without the complementary affordance of hearing. Likewise a software agent designed to assist with ongoing human needs for navigation help is useless unless its navigation algorithm allows for outside guidance.
- *Soft dependencies*. Coactive design emphasizes the importance of both "hard" and "soft" dependencies in coordinating related activities. Hard dependencies are necessary, or the joint activity could not happen in the first place. An example of a hard dependency is the passing of a baton in a relay race—the second runner simply can't begin until the first runner completes the handoff. Soft dependencies are not strictly necessary but are *helpful*. Attending to soft dependencies is a subtle, but no less significant process—in fact, it is what generally distinguishes great teams from mediocre ones. For instance, the first runner may shout something to the second runner before or during the handoff to convey a warning about a slippery section of track or to share other kinds of relevant information. If the approach of the first runner were difficult to confirm visually, progress

appraisal would be in order ("I'll be there in about five seconds!"). Of course, none of the first runner's signals would be of any use unless the second runner were monitoring for such communications. Soft dependencies may go beyond the sharing of information when, for example, a person or agent suspends its current activity in order to help another member of the group perform their task.

- *Joint goals*. Multi-agent teamwork research typically has held a simple view of joint goals, based on the unification of symbols common to all parties [19; 57]. However, Cartwright and Zander [13] point out the necessity of a more sophisticated view when humans are involved in joint activity. Apart from the problem of establishing and maintaining common ground on complex goals and the best means to achieve them, they emphasize that team goals are sometimes in competition with goals that individuals have for themselves and for the team [55].
- *Mixed-initiative opportunities in all phases of the sense-plan-act cycle.* Mixed-initiative interaction, where the roles and actions of people and agents are opportunistically negotiated during problem solving [2], has typically been limited to the planning and command generation aspects of human-agent interaction. To these, Fong [29] perceptively added the aspects of perception and cognition. Coactive design extends this earlier work in all phases of the sense-plan-act cycle, consistent with Castelfranchi's contention that "any needed resource and power within the action-perception loop of an agent defines a possible dimension of dependence or of autonomy" [14]. Coactive design the mutual interdependence of the all parties instead of merely focusing on the dependence of one of the parties on the other. It recognizes the benefits of designing agents with the capabilities they need to be interdependent.

Summary

With all these considerations in mind, we might formulate the characteristics of a good agent—human or artificial—with regard to joint activity in the following simple maxims:

1. A good agent is *observable*. It makes its pertinent state and intentions obvious.
2. A good agent is attuned to the requirement of *progress appraisal*. It enables others to stay informed about the status of its tasks and identifies any potential trouble spots ahead.
3. A good agent is *informative* and *polite.* It knows enough about others and their situations so that it can tailor its messages to be helpful, opportune, and appropriately presented.
4. A good agent *knows its limits*. It knows when to take the initiative on its own, and when it needs to wait for outside direction. It respects policy-based constraints on its behavior, but will consider exceptions and workarounds when appropriate.
5. A good agent is *predictable* and *dependable*. It can be counted on to do its part.
6. A good agent is *directable* at all levels of the sense-plan-act cycle. It can be retasked in a timely way by a recognized authority whenever circumstances require.
7. A good agent is *selective*. It helps others focus attention on what is most important in the current context.

8. A good agent is *coordinated*. It helps communicate, manage, and deconflict dependencies among activities, knowledge, and resources that are prerequisites to effective task performance and the maintenance of "common ground."

References

1. Allen, J., N. Chambers, G. Ferguson, L. Galescu, H. Jung, M. Swift, and W. Taysom. "PLOW: A Collaborative Task Learning Agent." In *Proceedings of AAAI 2007*. 2007.
2. Allen, J.F. "Mixed-initiative interaction." *IEEE Intelligent Systems*, September-October 1999, 14-16.
3. Allen, J.F. and G. Ferguson. "Human-machine collaborative planning." Presented at the Proceedings of the NASA Planning and Scheduling Workshop, Houston, TX 2002.
4. Asimov, I. "Runaround." In *I, Robot*, edited by I. Asimov, 33-51. London, England: Grafton Books. Originally published in *Astounding Science Fiction,* 1942, pp. 94-103, 1942/1968.
5. Boy, G. *Cognitive Function Analysis*. Stamford, CT: Ablex Publishing, 1998.
6. Bradshaw, J.M. "An introduction to software agents." In *Software Agents*, edited by J.M. Bradshaw, 3-46. Cambridge, MA: AAAI Press/The MIT Press, 1997.
7. Bradshaw, J.M., M. Sierhuis, Y. Gawdiak, R. Jeffers, N. Suri, and M. Greaves. "Teamwork and adjustable autonomy for the Personal Satellite Assistant." Presented at the Workshop on Autonomy, Delegation and Control: Interacting with Autonomous Agents. Seventeenth International Joint Conference on Artivicial Inelligence (IJCAI-2001), Seattle, WA, 6 August, 2001.
8. Bradshaw, J.M., M. Sierhuis, A. Acquisti, P. Feltovich, R. Hoffman, R. Jeffers, D. Prescott, N. Suri, A. Uszok, and R. Van Hoof. "Adjustable autonomy and human-agent teamwork in practice: An interim report on space applications." In *Agent Autonomy*, edited by H. Hexmoor, R. Falcone, and C. Castelfranchi, 243-280. Kluwer, 2003.
9. Bradshaw, J.M., P. Feltovich, H. Jung, S. Kulkarni, W. Taysom, and A. Uszok. "Dimensions of adjustable autonomy and mixed-initiative interaction." In *Agents and Computational Autonomy: Potential, Risks, and Solutions. Lecture Notes in Computer Science, Vol. 2969*, edited by M. Nickles, M. Rovatsos, and G. Weiss, 17-39. Berlin, Germany: Springer-Verlag, 2004.
10. Bradshaw, J.M., H. Jung, S. Kulkarni, J. Allen, L. Bunch, N. Chambers, P.J. Feltovich, L. Galescu, R. Jeffers, M. Johnson, W. Taysom, and A. Uszok. "Toward trustworthy adjustable autonomy and mixed-initiative in KAoS." Presented at the Proceedings of the Autonomous Agents and Multi-Agent Systems (AAMAS) 2004 Trust Workshop, New York City, NY, 19 July, 2004, 9-20.
11. Bradshaw, J.M., N. Suri, M.R. Breedy, A. Canas, R. Davis, K.M. Ford, R. Hoffman, R. Jeffers, S. Kulkarni, J. Lott, T. Reichherzer, and A. Uszok. Updated and expanded version of an article that originally appeared in IEEE Intelligent Systems, July 2001, pp. 49-56. "Terraforming cyberspace." In *Process Coordination and Ubiquitous Computing*, edited by D.C. Marinescu and C. Lee, 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, 2002.

12. Burstein, M.H. and D.V. McDermott. "Issues in the development of human-computer mixed-initiative planning." In *Cognitive Technology: In Search of a Humane Interface*, edited by B. Gorayska and J.L. Mey. Elsevier Science, 1996.
13. Cartwright, D. and A. Zander. *Group Dynamics: Research ant Theory*. Third ed. New York City, NY: Harper and Row, 1968.
14. Castelfranchi, C. "Founding agents' 'autonomy' on dependence theory." Presented at the 14th European Conference on Artificial Intelligence (ECAI 2000), Berlin, Germany, 20-25 August, 2000, 353-357.
15. Christofferson, K. and D.D. Woods. "How to make automated systems team players." In *Advances in Human Performance and Cognitive Engineering Research, Vol. 2*, edited by E. Salas. JAI Press, Elsevier, 2002.
16. Clancey, W.J. "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*, no. 2 (2004).
17. Clark, H.H. and S.E. Brennan. "Grounding in communication." In *Perspectives on Socially Shared Cognition*, edited by L.B. Resnick, J.M. Levine, and S.D. Teasley. Washington, D.C.: American Psychological Association, 1991.
18. Clark, H.H. *Using Language*. Cambridge, UK: Cambridge University Press, 1996.
19. Cohen, P.R. and H.J. Levesque. "Teamwork." Menlo Park, CA: SRI International, 1991.
20. Conte, R. and C. Castelfranchi. *Cognitive and social action*. London, England: UCL Press, 1995.
21. Cypher, A., ed. *Watch What I Do: Programming by Demonstration*. Cambridge, MA: MIT Press, 1993.
22. Dorais, G., R.P. Bonasso, D. Kortenkamp, B. Pell, and D. Schrekenghost. "Adjustable autonomy for human-centered autonomous systems on Mars." Presented at the Proceedings of the AAAI Spring Symposium on Agents with Adjustable Autonomy. AAAI Technical Report SS-99-06, Menlo Park, CA 1999.
23. Falcone, R. and C. Castelfranchi. "Adjustable social autonomy." (2002).
24. Feltovich, P., J.M. Bradshaw, R. Jeffers, N. Suri, and A. Uszok. "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. LNAI 3071*, edited by A. Omicini, P. Petta, and J. Pitt. Lecture Notes in Computer Science, 21-48. Berlin, Germany: Springer-Verlag, 2004.
25. Feltovich, P., J.M. Bradshaw, W.J. Clancey, and M. Johnson. "Toward and Ontology of Regulation: Support for Coordination in Human and Machine Joint Activity." In *Engineering Societies in the Agents World VII. Lecture Notes in Computer Science*, edited by G. O'Hare, A. Ricci, M. O'Grady, A. Ricci, and O. Dikenelli, 175-192. Heidelberg, Germany: Springer-Verlag, 2007.
26. Feltovich, P., J.M. Bradshaw, W.J. Clancey, M. Johnson, and L. Bunch. "Progress appraisal as a challenging element of coordination in human and machine joint activity." In *Engineering Societies for the Agents World VIII. Lecture Notes in Computer Science*, edited by A. Artikis, G. O'Hare, K. Stathis, and G. Vouros, 124-141. Heidelberg, Germany: Springer-Verlag, 2008.
27. Ferguson, G., J. Allen, and B. Miller. "TRAINS-95: Towards a mixed-initiative planning assistant." Presented at the Proceedings of the Third Conference on

Artificial Intelligence Planning Systems (AIPS-96), Edinburgh, Scotland 1996, 70-77.
28. Fitts, P.M., ed. *Human Engineering for an Effective Air Navigation and Traffic Control System*. Washington, D.C.: National Research Council, 1951.
29. Fong, T.W. "Collaborative Control: A Robot-Centric Model for Vehicle Teleoperation. Technical Report CMU-R1-TR-01-34." Carnegie Mellon University, 2001.
30. Ford, K.M., C. Glymour, and P. Hayes. "Cognitive prostheses." *AI Magazine 18*, no. 3 (Fall 1997): 104.
31. Fox, J. and S. Das. *Safe and Sound: Artificial Intelligence in Hazardous Applications*. Menlo Park, CA: AAAI Press/The MIT Press, 2000.
32. Geertz, C. *The Interpretation of Cultures*. New York, NY: Basic Books, 1973.
33. Goodrich, M.A., D.R. Olsen Jr., J.W. Crandall, and T.J. Palmer. "Experiments in adjustable autonomy." Presented at the Proceedings of the IJCAI_01 Workshop on Autonomy, Delegation, and Control: Interacting with Autonomous Agents, Seattle, WA, August, 2001.
34. Grosz, B.J. "Collaborative systems." *AI Magazine 17*, no. 2 (Summer 1996): 67-85.
35. Hamilton, S. "Thinking outside the box at IHMC." *IEEE Computer* (January 2001): 61-71.
36. Hancock, P.A. and S.F. Scallen. "Allocating functions in human-machine systems." In *Viewing Psychology as a Whole*, edited by R. Hoffman, M.F. Sherrick, and J.S. Warm, 509-540. Washington, D.C.: American Psychological Association, 1998.
37. Helton, W.S. *Canine Ergonomics: The Science of Working Dogs*. Boca Raton, FL: Taylor and Francis, 2009.
38. Hoffman, R., P. Feltovich, K.M. Ford, D.D. Woods, G. Klein, and A. Feltovich. "A rose by any other name… would probably be given an acronym." *IEEE Intelligent Systems*, July-August 2002, 72-80.
39. Horvitz, E. "Principles of mixed-initiative user interfaces." Presented at the Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '99), Pittsburgh, PA, May, 1999.
40. Johnson, M., J.M. Bradshaw, P. Feltovich, C. Jonker, B. van Riemsdijk, and M. Sierhuis. "Coactive design: Why interdependence must shape autonomy. Submitted for publication." 2010
41. Kay, A. "Computer Software." *Scientific American 251*, no. 3 (1984): 53-59.
42. Klein, G., P.J. Feltovich, J.M. Bradshaw, and D.D. Woods. "Common ground and coordination in joint activity." In *Organizational Simulation*, edited by W.B. Rouse and K.R. Boff, 139-184. New York City, NY: John Wiley, 2004.
43. Licklider, J.C.R. "Man-computer symbiosis." *IRE Transactions in Electronics. New York: Institute of Radio Engineers.* (1960): 4-11.
44. Lieberman, H., ed. *Your Wish is My Command: Programming By Example*. San Francisco, CA: Morgan Kaufmann, 2001.
45. Lieberman, H. and T. Selker. "Agents for the user interface." In *Handbook of Software Agents*, edited by J.M. Bradshaw. Cambridge, MA: AAAI Press/The MIT Press, 2002.

46. Luck, M., M. D'Inverno, and S. Munroe. "Autonomy: Variable and generative." In *Agent Autonomy*, edited by H. Hexmoor, C. Castelfranchi, and R. Falcone, 9-22. Dordrecht, The Netherlands: Kluwer, 2002.
47. Maheswaran, R.T., M. Tambe, P. Varakantham, and K. Myers. "Adjustable autonomy challenges in personal assistant agents: A position paper." In *Computational Autonomy*, edited by M. Klusch, G. Weiss, and M. Rovatsos, in press. Berlin, Germany: Springer, 2004.
48. Malone, T.W. and K. Crowston. "What is coordination theory and how can it help design cooperative work systems?" Presented at the Conference on Computer-Supported Cooperative Work (CSCW '90), Los Angeles, CA, 7-10 October, 1990, 357-370.
49. Muscettola, N., P.P. Nayak, B. Pell, and B.C. Williams. "Remote Agent: To boldly go where no AI system has gone before." *Artificial Intelligence 103*, no. 1-2 (August 1998): 5-48.
50. Norman, D.A. "Cognitive artifacts." In *Designing Interaction: Psychology at the Human-Computer Interface*, edited by J.M. Carroll, 17-38. Cambridge: Cambridge University Press, 1992.
51. Norman, D.A. "How might people interact with agents?" In *Software Agents*, edited by J.M. Bradshaw, 49-55 (see also How might people interact with robots? http://www.jnd.org/dn.mss/how_might_humans_int.html). Cambridge, MA: The AAAI Press/The MIT Press, 1997.
52. Penserini, L., H.M. Aldewereld, F.P.M. Dignum, and M.V. Dignum. "A formal specification for organisational adapation." Presented at the Proceedings of the Tenth International Workshop on Agent-Oriented Software Engineering (AOSE-09) 2009.
53. Rich, C., C. Sidner, and N. Lesh. "COLLAGEN: Applying collaborative discourse theory." *AI Magazine* (2001).
54. Shoham, Y. and M. Tennenholtz. "On the synthesis of useful social laws for artificial agent societies." Presented at the Proceedings of the Tenth National Conference on Artificial Intelligence, San Jose, CA 1992, 276-281.
55. Sierhuis, M., J.M. Bradshaw, A. Acquisti, R. Van Hoof, R. Jeffers, and A. Uszok. "Human-agent teamwork and adjustable autonomy in practice." Presented at the Proceedings of the Seventh International Symposium on Artificial Intelligence, Robotics and Automation in Space (i-SAIRAS), Nara, Japan, 19-23 May, 2003.
56. Smith, W.J. *The Behavior of Communicating*. Cambridge, MA: Harvard University Press, 1977.
57. Tambe, M., W. Shen, M. Mataric, D.V. Pynadath, D. Goldberg, P.J. Modi, Z. Qiu, and B. Salemi. "Teamwork in cyberspace: Using TEAMCORE to make agents team-ready." Presented at the Proceedings of the AAAI Spring Symposium on Agents in Cyberspace, Menlo Park, CA 1999.
58. Tambe, M., D. Pynadath, C. Chauvat, A. Das, and G. Kaminka. "Adaptive agent architectures for heterogeneous team members." Presented at the Proceedings of the International Conference on Multi-Agent Systems (ICMAS 2000) 2000.
59. Vecht, B.v.d., F.P.M. Dignum, J.-J.C. Meyer, and M. Neef. "Coordination, organizations, institutions, and norms in agent systems." In *Proceedings of COIN*.

Lecture Notes in Computer Science 4870, 83-96. Berlin, Germany: Springer Verlag, 2007.

60. Verhagen, H. "Norms and artificial agents." Presented at the Sixth Meeting of the Special Interest Group on Agent-Based Social Simulation, ESPRIT Network of Excellence on Agent-Based Computing, Amsterdam, Holland, 4-5 December, 2001.

61. Woods, D.D. "Steering the reverberations of technology change on fields of practice: Laws that govern cognitive work." Presented at the Proceedings of the Annual Meeting of the Cognitive Science Society, 10 August, 2002.