Coactive Design

Why Interdependence Must Shape Autonomy

Matthew Johnson1,2, Jeffrey M. Bradshaw1 and Paul J. Feltovich1
1Florida Institute for Human and Machine Cognition (IHMC)
Pensacola, Florida, USA
1-850-202-4476
{mjohnson, jbradshaw and pfeltovich}@ihmc.us

Catholijn M. Jonker2 and Birna van Riemsdijk2
2EEMCS, Delft University of Technology
Delft, The Netherlands
+31.15.2781315
c.m.jonker@tudelft.nl and m.b.vanriemsdijk@tudelft.nl

Maarten Sierhuis2,3
3Carnegie Mellon University
Silicon Valley Campus
Moffett Field, California, USA
+1-650-604-4917
maarten.sierhuis@sv.cmu.edu

ABSTRACT
This paper introduces Coactive Design as a new approach to address the increasingly sophisticated roles for people and agents in mixed human-agent systems. The basic premise of Coactive Design is that the underlying interdependence of joint activity is a critical design feature. When designing the capabilities that make an agent autonomous, the process should be guided by an understanding of the interdependence in the joint activity. This understanding is then used to shape the operation of agent capabilities and enable appropriate interaction. The success of future human-agent teams hinges not merely on trying to make agents more autonomous, but also in striving to make them more capable of sophisticated interdependence.

Categories and Subject Descriptors
I.2.9 [Robotics]: Autonomous vehicles

General Terms
Design, Human Factors, Theory.

Keywords
Coactive, autonomy, interdependence, joint activity.

1. INTRODUCTION
This paper introduces the concept of coactivity and presents Coactive Design as a novel approach for designing human-agent systems. Throughout the paper we will use the terms “agent” and “robot” interchangeably to mean any artificial actor. Both robot and agent developers continue to pursue much more sophisticated roles for their machines. Some of the envisioned roles include caretaking assistants for the elderly, medical assistants, day care assistants, coworkers in factories and offices, and servants in our homes. Not only are the agents themselves increasing in their capabilities, but also the composition of human-robot systems is growing in scale and heterogeneity. All these requirements showcase the importance of robots transitioning from common roles of today, where they are frequently no more than teleoperated1 tools, to more sophisticated partners or teammates.

Full teleoperation1 and full autonomy2 are often thought of as two extremes on a spectrum. Researchers have been investigating the middle ground between these extremes under various names including mixed-initiative [1], adjustable autonomy [2], collaborative control [3], and sliding autonomy [4]. Each of these approaches attempts to keep the human-agent system operating at a “sweet spot” between the two extremes. As the names suggest, these approaches understand that the ideal is not a fixed location along this spectrum but may need to vary along the spectrum. These approaches and most traditional planning technologies at the foundation of intelligent robotic systems typically take an autonomy-centered approach, focusing on the problems of control and task allocation.

In contrast to these autonomy-centered approaches, Coactive Design is a teamwork-centered approach. The concept of teamwork-centered autonomy was addressed by Bradshaw, et al. [5]. It takes as a beginning the premise that people are working in parallel alongside autonomous systems, and hence adopts the stance that the processes of understanding, problem solving and task execution are necessarily incremental, subject to negotiation, and forever tentative. The basic premise of Coactive Design is that, in sophisticated human-agent systems, the underlying interdependence of joint activity is the critical design feature. From this perspective, the design of capabilities to make an agent autonomous should be guided by an understanding of the interdependence in the joint activities these agents will undertake. This understanding is then used to shape implementation of agent capabilities, thus enabling appropriate interaction. We no longer look at the problem as simply trying to make agents more autonomous, but, in addition, we strive to make them more capable of being interdependent.

This paper will begin by explaining the different usages of the term autonomy. We will discuss several major approaches to human-agent interaction and show how they are mainly autonomy-centered. We will also discuss the ways in which autonomy has been characterized. Next we present the basic premise of Coactive Design and the concept of interdependence itself, which, like autonomy, is a highly nuanced term. We then discuss how Coactive Design fits in relation to prior work and

1 Teleoperation is manually operating a machine from a distance.

2 Autonomy will be more fully explained in the following sections, but here it can simply be thought of as “without intervention by other actors.”
highlight the new areas of focus. We briefly discuss some preliminary work and then close with a discussion of why this novel approach is important.

2. AUTONOMY

Autonomy has two basic senses in everyday usage. The first sense, self-sufficiency, is about the capability of an entity to take care of itself. Bradshaw [6] refers to this as the descriptive dimension. Similarly, Castelfranchi [7] referred to this as one of the two aspects of “social autonomy” that he called independence. People usually consider robot autonomy in this sense with respect to a particular task. For example, a robot may be able to autonomously navigate in an office environment. The second sense refers to the quality of self-directedness, or freedom from outside control, which Bradshaw calls the prescriptive dimension. Castelfranchi referred to this as autonomy of delegation and considered it another form of “social autonomy.” For robots, this usually means freedom from human input or intervention during a particular task.

Castelfranchi argued that both of the preceding senses are reducible to a single aspect and we agree. However, we will continue to keep them separate because it aids our explanation of previous work and because we occasionally need to make the distinction. Castelfranchi also included a third sense of autonomy that he calls “non-social autonomy.” This refers to independence from environmental stimuli as opposed to the social autonomy which was about freedom from other agents. Since we do not limit the self-directedness to other agents as Castelfranchi does, we feel it is included in self-directedness.

For this paper, we will use self-sufficiency and self-directedness to distinguish the two senses which often confound discussions on autonomy. Self-sufficiency will be used to express an agent’s inherent capability. Self-directedness will be used to express an agent’s freedom from outside control or authority.

2.1 How Prior Work is Autonomy-Centered

There have been many approaches to improve human-robot system effectiveness and we will now discuss several of the more prominent approaches. Parts of the discussion below are adapted from [6].

2.1.1 Functional Allocation and Supervisory Control

The concept of automation—which began with the straightforward objective of replacing whenever feasible any task currently performed by a human with a machine that could do the same task better, faster, or cheaper—became one of the first issues to attract the notice of early human factors researchers. These researchers attempted to systematically characterize the general strengths and weaknesses of humans and machines [8]. 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. Sheridan proposed Supervisory Control [9], in which a human oversees one or more autonomous systems, allocating tasks. Once control is given to the system, it is ideally expected to complete the tasks without human intervention. Both of these approaches are considered autonomy-centered, specifically concerned with the self-sufficient aspect of autonomy. The designer’s job is to determine what needs to be done and then provide the agent the capability (i.e., self-sufficiency) to do it. Autonomy is therefore shaped by self-sufficiency in this case.

2.1.2 Adjustable/Adaptive/Sliding Autonomy

Over time it became plain to researchers that things were not as simple as they first appeared. For example, many functions in complex systems are shared by humans and machines; hence the need to consider synergies and conflicts among the various performers of joint actions. Also, the suitability of a particular human or machine to take on a particular task may vary by time and over different situations; hence the need for methods of function allocation that are dynamic and adaptive. There are many approaches that suggest ways to vary autonomy. Dorais [10] defines adjustable autonomy as “the ability of autonomous systems to operate with dynamically varying levels of independence, intelligence and control.” Dias [11] uses the term sliding autonomy, but defines it similarly. Sheridan discusses adaptive automation in which the system must decide which functions to automate and to what extent. We will use adjustable autonomy to refer to them all and take a general meaning of automatically and appropriately adjusting the robot’s level of autonomy, in this case the self-directedness aspect, based on the situation. The adjustment may be initiated by the human or by the robot itself. Again, it is clear that these approaches are autonomy-centered, with the focus being on task assignment, control and level of independence. Here the self-directedness shapes the autonomy. One very important concept put forth by these approaches is the importance of adaptivity which will be critical to future systems.

2.1.3 Mixed-Initiative Interaction

Mixed-initiative approaches evolved from a different research community, but share similar ideas and assumptions. Allen defines mixed-initiative as “a flexible interaction strategy, where each agent can contribute to the task what it does best” [1]. Although interaction is used in the description, it is mainly used to negotiate which party does which task. Similarly Myers and Morley describe a framework called “Taskable Reactive Agent Communities (TRAC) [12], which supports the directability of a team of agents by a human supervisor by modifying task guidance.” Again directability or task allocation is the central feature. Murphy [13] also uses the term mixed initiative to describe their attention directing system, the goal of which is to get the human to pick up tasks when a robot has a failure. Like all the others, mixed-initiative style approaches are essentially autonomy-centered and frequently focus on task assignment or the authority to act and as such address the self-directedness aspect of autonomy; thus, self-directedness is still the major contributor to shaping the autonomy of the agent. Mixed initiative interaction contributes the valuable insight that joint activity is about interaction and negotiation and that control is not something that is statically assigned, but dynamically shifts as necessary.

2.1.4 Collaborative Control

Collaborative Control is an approach proposed by Fong [3] that uses human-robot dialogue (i.e., queries from the robot and responses, or lack thereof, from the human), as the mechanism for adaptation. As Fong states, “Collaborative control also allows
robots to benefit from human assistance during perception and cognition, and not just planning and command generation” [3]. Collaborative Control is a first step toward Coactive Design, allowing both parties to participate in the same action. Here the interdependence of the navigation task was used to shape the autonomy. The robot was designed to allow the human, who is also involved in the task, to provide assistance in the perceptual and cognitive parts of the task. This assistance is not required, so we are not talking about self-sufficiency, but it is designed for and enabled. Some of the ideas from this approach will be adapted and extended by Coactive Design.

2.2 How Autonomy has been Characterized

A way to gain insight into the focus of approaches in a community is to review how the community categorizes and describes its own work. This also provides a test of our claims that most prior work is autonomy-centered.

Several researchers have worked on describing different levels of autonomy. Yanco [14] made the distinction based on the amount of intervention required. For example, full teleoperation is 100% intervention and 0% automation. Tour guide robots are labeled 100% autonomous and 0% intervention. The assumption in this model is that intervention only occurs when the robot lacks self-sufficiency. However, identifying the percentage of intervention is difficult to quantify particularly when not at the extreme ends of the spectrum. Similarly Sheridan [15] provides a list of levels of autonomy shown in Figure 1.

![Figure 1 Levels of automation [15]](image1)

Sheridan’s scale is clearly autonomy-centered, as noted by Goodrich and Schultz [16]. Specifically it focused on the self-directed aspect of autonomy. Goodrich and Schultz [16] provided a scale which attempts to focus on interaction instead of autonomy, shown in Figure 2.

![Figure 2 Levels of autonomy with an emphasis on human interaction [16]](image2)

Their desire was to capture something more than the previous autonomy-centered views, but it is more of a historical summary. Peer-to-peer collaboration as a term holds promise, but is never clearly defined, as it was a future direction. However, the right most coordinate of Figure 2 is “dynamic autonomy” which sounds a lot like all of the previous autonomy-centered approaches.

Bradshaw describes multiple dimensions of autonomy instead of a single one-dimensional scale of levels [6]. He describes a descriptive dimension and a prescriptive dimension capturing both of the two initial senses of autonomy. He also argues that the measurement of these dimensions should be specific to task and situation.

Castelfranchi suggested dependence as the complement of autonomy [7] and describes dimensions of autonomy in terms of the autonomy/dependence of various capabilities in a standard Procedural Reasoning System (PRS) architecture. These include information, interpretation, know-how, planning, plan discretion, goal dynamics, goal discretion, motivation, reasoning, monitoring, and skill autonomy. Like Bradshaw, Castelfranchi recognizes that autonomy is not a monolithic property, but should be applied to each aspect of the agent. Castelfranchi put it this way: “any needed resource or power within the action-perception loop of an agent define a possible dimension of dependence or autonomy.” [7]

3. COACTIVE DESIGN

Coactive Design takes interdependence as the central organizing principle among people and agents working together in joint activity. Certainly issues of autonomy are still important, for instance, what does an agent needs to do, how well does it need to do it, and how much control do we give it. However, all of these aspects should be guided by an understanding of the interdependence in the joint activity. Then the interdependence can be used to shape the autonomy and enable appropriate interaction. In contrast to autonomous systems designed to take humans out of the loop, we are specifically designing systems to address requirements that allow close and continuous interaction with people.

As we try to design more sophisticated systems, we move along a maturity continuum [17] from dependence to independence to interdependence. The process is a continuum because a small level of independence of agents through autonomy is a prerequisite for interdependence. However, independence is not the supreme achievement [17] in human-human interaction, nor should it be in human-agent systems. Interdependence is much more complex and difficult for both machines and humans. We are no longer dealing with individual autonomous actions but with group participatory actions [18]. This is a departure from the previous approaches discussed in section 2.1, with the exception of Collaborative Control which began to incorporate all parties into the action (at least in the perceptual and cognitive dimensions). As Clark states, “a person’s processes may be very different in individual and joint actions even when they appear identical” [18]. Clark’s example is playing a musical solo versus playing a duet. Although the music is the same, the processes involved are very different.

The term coactive is specifically chosen to highlight the difference in the Coactive Design approach. There are three meanings [19] associated with coactive:

1) Joint action
2) An impelling or restraining force; a compulsion
3) Ecology. Any of the reciprocal actions or effects, such as symbiosis, that can occur in a community.

Joint action is about each participant being engaged in the same action, or more specifically activity; meaning a process
extended in space and time [18, 20]. Previous work in human-agent interaction focused largely on assigning or allocating tasks to individuals. As we move toward more sophisticated human-machine systems, the activity looks more and more “joint-like.” Consider the unmanned aerial vehicle. The first task in development was a standard engineering task to make the vehicle self-sufficient for some tasks (e.g., waypoint following). As the capabilities and robustness increased, the focus shifted to self-directedness (e.g., what am I willing to let this machine do autonomously). Now much of the effort is interdependence focused (e.g., how can I get these vehicles to work effectively as a team with their operators?). It is a natural maturing process and robots and agents are now pushing into new territory.

The compulsion derives from the interdependence inherent in the joint activity. Joint activity means that there is a collective obligation [21] of all parties, even if not currently “assigned” to the task. This obligation includes certain duties and obligations that correlate with good teamwork. These obligations both compel us, for example, to provide help, and restrain us, for example not to hog any limited resources. Capturing these obligations is an essential part of Coactive Design and a departure from most previous approaches that do not address the idea of a collective obligation.

The last key feature of Coactive Design is the idea of reciprocal actions. Most previous systems take a unidirectional view. They either focus on automating tasks to offload work from the operator or they focus on enabling the operator to take over a task to make up for poor robot performance or ability. Coactive Design espouses a bidirectional view. For example, if I need to know your status, you must be able to provide status updates. If you can help me make navigation decisions, my navigation algorithm must allow outside guidance. Simply stated; one can only give if the others can take and vice versa. Many of the abilities required for good teamwork required reciprocal abilities from the other team members. In this way Coactive Design focuses on teamwork-centered autonomy. This is another break from the previous work that tended to focus on individualistic autonomy.

Coactive Design is a framework for addressing the more challenging roles for agents (e.g. care taker, medical assistant, coworker, or servant) and human-agent teams, especially heterogeneous teams. These roles have a much higher commitment than other types of interaction, such as passing in a hallway or making a sales transaction with a grocery clerk. The target for Coactive Design is not current teleoperated systems or systems struggling with basic autonomy. We are specifically addressing what a human-agent system would look like if it were to fill one of the roles suggested above. The envisioned roles, if properly performed, have a high level of interdependence that cannot be addressed solely by adjusting who is in control or who is assigned what task—and necessitate a focus on the coactivity. In contrast to autonomous systems designed to take humans out of the loop, we are specifically addressing the requirements for close and continuous interaction with people.

Coactive Design has joint activity at its core and joint activity is largely about interdependence. To gain more insight into the aspects of Coactive Design, it is necessary to look more deeply at interdependence.

4. INTERDEPENDENCE

Coordination is foundational to joint activity and is required largely because of interdependencies among activities [22]. Understanding the nature of the interdependencies involved in the coordination is an important part of determining the capability requirements of agents and designing a solution.

4.1 Malone and Crowston

In their interdisciplinary study of coordination Malone and Crowston [23] summarize prior work on coordination in which they drew on Computer Science, Organization Theory, Management Science, Economics, Linguistics and Psychology. They view coordination as managing dependencies. They also characterize the types of dependencies and categorized some of the most common; use of shared resources, simultaneity of processes, producer/consumer relationships, task/subtask roles, task assignment and transfer dependency. Thompson [24] suggested three types of interdependence: pooled, sequential and reciprocal. Pooled interdependence is about each entity contributing a discrete part to the whole and that each is supported by the whole. This is more applicable to large organizations then the smaller teams we are considering. Sequential interdependence maps directly to Malone’s producer/consumer category. Reciprocal interdependence is of the bidirectional producer/consumer type. The use of reciprocal here is referring to the cyclical nature of the producer/consumer relation. This is very different from our use of reciprocal which is about the need for complimentary capabilities. We believe Malone’s categories cover Thompsons and that all of these can be represented by two basic types of dependency: resource and temporal.

4.2 Resource Dependency

Resource dependency can involve a variety of things including a tool, space, the product of a process, or the capability to perform some action. Dependency for resources has received much attention in the literature and is the same as Malone’s “shared resource dependency.” We will represent an activity A as being resource dependent on resource x as:

\[ x \in A \]

There are many ways to formally represent dependence, and we are not attempting to provide another formal specification here. Instead a simplified notation is provided to facilitate this discussion.

4.3 Temporal Dependency

Temporal dependency is the time relation between events or actions. While it is conceivable to view time constraints as a resource as well, this only makes sense when discussing time requirements associated with resources (e.g. I need the hammer for the next five minutes). Hence, it is clearer to keep them separate. For temporal constraints we will need a more detailed definition of the activity involved, so we define an activity A as an activity that spans time from \( t=0 \) to \( t=n \) such that A is represented by \( \{A_0...A_n\} \) where \( A_0 \) is the start of the activity and \( A_n \) is the completion of the activity, noting that \( A_0 \) can be the same as \( A_n \) for actions with negligible duration. Now it is easy to define a serial sequencing temporal dependence such as “A must start after B finishes”:

\[ B \rightarrow A \]

Table I lists a few more examples of temporal dependencies.
4.4 Malone Revisited

Now we will show how Malone’s original listing can be represented as combinations of these two types of dependencies. Assuming B generates x (B⇒x), the producer/consumer dependency can be viewed as resource dependence of A on the output of B and a temporal dependence that A must occur after B:

B ⇒ x⇒ A; B₀ ⇒ A₀

Similarly, task/subtask dependence can be viewed as a resource dependence on the subtask B directly and a temporal dependence that requires the subtask to occur within the time span of the task:

B ⇒ A; A₀ ⇒ B₀; B₀ ⇒ A₀

One can generate even more complex time relationships using the types of dependencies discussed in Table I. Task assignment can also be represented this way with the task now being performed by another, thus adding the additional resource dependence of the other agent. Lastly, transfer dependence is similar to the producer/consumer dependency with the addition of a potentially time dependent exchange of information, which we will call activity C:

B ⇒ C ⇒ A; C₀ ⇒ A₀; B₀ ⇒ A₀

In this way, more complex dependencies can be composed from the two basic types; resource and temporal. There are two other types of dependency that we see as critical in Coactive Design that are not captured by Malone’s list; soft dependency and monitoring dependency. We will discuss these next.

4.5 Soft Dependency

Dependency can be “hard” meaning that activity A cannot proceed without x, or it can be “soft” meaning that activity A can potentially involve x, but it is not required. For example, in order to enter a room with one door, a robot would have a “hard” dependence on the one door. If the room had two doors, the robot would have a “soft” dependence on both doors. We will represent the “soft” dependency as:

x → A

Besides redundant or alternative options, “soft” dependency can also refer to information that is not required, but if provided it could potentially alter the behavior of the recipient. Some examples would be progress appraisals [25] (“I’m running late”), warnings (“Watch your step”) and unexpected events (“It has started to rain”). While the planning community and others have contributed a large body of work on the standard “hard” dependencies critical to a functioning human-robot system, the “soft” dependencies have received less attention. These types of dependency can lead to richer and more interesting types of interaction than have typically been implemented and are important aspects of Coactive Design.

These typically fall under the compulsion or collective obligation we have discussed. Although not required (hard dependency), understanding them can help shape an agent’s autonomy to better support interdependent roles.

4.6 Monitoring Dependency

If there is dependence, either resource or temporal, there is also an implied “monitoring dependency,” if joint activity is to be successful. The dependent agent is obligated to monitor the situation appropriately. There are two possible options:

1) Observe the environment (including time or other agents)
2) Wait for a signal or message

If, for example, an agent needs an elevator (resource dependence) the agent can monitor the elevator doors to see when they open. Alternatively, the agent could be notified of availability through signaling (e.g. up arrow light turns on, audible bell, or an elevator operator telling you “going up”). Each option has its challenges but for now we just want to convey that monitoring is an important consideration in Coactive Design. Monitoring dependence also highlights the reciprocal nature of the activity. Not only does the monitoring entity need to monitor, but the monitored entity may need to make certain aspects of its operation transparent.

5. VISUALIZING THE NEW PERSPECTIVE

So how does the coactive design perspective change the way we see the design problem? The first way is to consider joint activity, as explained by Clark [18]. Joint activity highlights the issue of interdependence. We still need to consider autonomy in both its dimensions, but now we must also consider interdependence of the activity, as sketched roughly in Figure 3.
competence. The Interdependability axis is about the level at which the agent is able to depend on others or be depended upon by others. This axis is specifically about the capability, not the need or requirement which are captured by the other axes.

We can now map some of the prior work onto this space, shown in Figure 4. First we can talk about where the approaches fall on the new three-dimensional spectrum. The Functional Allocation problem of determining what to automate is really about determining where one wants to be on the self-sufficiency axis. Adjustable autonomy and much of the mixed-initiative approach are about changing along the self-directedness axis. We can also look at how autonomy is characterized in this new model. Yanco’s intervention level correlates with the self-sufficiency axis while Sheridan’s scale correlates to the self-directed axis. Bradshaw and Castelfranchi address both axes by capturing both aspects of autonomy.

![Figure 4 Mapping Prior Work](image)

A degree of self-sufficiency is essential to contribute to any joint activity and self-directedness is a sign of more maturity, but we need to progress toward interdependence to really excel at joint activity.

So far, being autonomy-centered we have stayed within the two autonomy axes. We now push into the third axis that captures interdependability. In addition to the interdependence of joint activity, the Coactive Design perspective includes the reciprocal nature of joint activity addressing for example, the need to match capabilities among participants. This means our interdependence axis needs more than just low and high. It needs to capture the duality of a sender-receiver relationship. In a broad sense, we are talking about the participants in the activity helping one another. This splits the third dimension into to complimentary dimensions, as shown in figure 5. We can now map Collaborative Control onto this new perspective. The main feature of Collaborative Control, as discussed earlier, was the ability of the human to provide assistance to the robot at the perceptual and cognitive levels. The wonderful insight of Collaborative Control was that tasks can be done more effectively if done collaboratively. The Coactive Design perspective extends this to allowing for the machine to assist the human providing the complimentary side of the axis.

![Figure 5 Mapping Collaborative Control](image)

Although we are showing a single set of axes for simplicity, Coactive Design perspective considers all aspects of an agent’s sense-act loop. This is directly in line with Castelfranchi’s [7] break down of autonomy based on the components of a PRS system. The take away message is not the support of any particular cognitive model, but instead the concept that there are many aspects to an agent as it performs in a joint activity. Just as Castelfranchi argued that autonomy can occur at any of these “levels” or dimensions, Coactive Design argues that the ability to be interdependent exists at each “level” or dimension as well.

Interdependence brings us to the last change in perspective; compulsion. This part of Coactive Design is about focusing on the monitoring and soft dependencies in a joint activity. Good teams distinguish themselves by handling the “soft” dependencies better, improving performance, efficiency, and/or situation awareness. Coactive Design approaches will need to consider “soft” dependencies. Furthermore, identifying areas of dependence in joint activity (both hard and soft) has been stated as a critical part of Coactive Design.

6. SUPPORT FOR COACTIVE DESIGN

We provide supporting evidence for our claims from three sources; a preliminary study of our own, results from other’s recent work, and some observations about autonomy, coordination and people.

6.1 PRELIMINARY STUDY

We have begun to investigate the implications of Coactive Design experimentally. We started with a very simple example domain and intend to increase complexity as we progress. Our first domain, Blocks World for Teams (BW4T)[26] (tasks were done in groups), was chosen to be as simple as possible. Similar to the classic AI planning problem of Blocks World, the goal of BW4T is to “stack” colored blocks in order. To keep things simple, the blocks are un-stacked to begin with, so un-stacking is unnecessary. Degree of interdependence that is embedded in the task is represented by the complexity of color orderings within the goal stack. The task environment was composed of nine rooms containing a random assortment of boxes and a drop off area for the goal. The environment is hidden from each of the players, except for the contents of the current room. There are basically two tasks in this domain; find a box and deliver a box to the drop off area. In some simple cases, the task could be done without any coordination, but it is clear that coordination (i.e. the players
managing the interdependencies among their moves) is highly beneficial.

Although a simple domain, this example demonstrates the complexity of coordination and interdependence even in the simplest domain. We ran twelve subjects in various team sizes (2, 3, 4, 5, 6, and 8). The teams were all human (i.e. no agents) for this first set of tests. The subjects were allowed to talk openly to one another. Although too early to be conclusive, our initial results are interesting and support our claims. As the activity became more interdependent (more complex configurations of blocks as goals), we noted an increase in the amount of coordination attempts, as would be expected. We also noted some interesting aspects of the communication. Although only two basic tasks are involved, we observed a wide variety of communications. Of particular interest were the large number of communications that were about soft dependencies and monitoring dependencies. Progress appraisal was also a common theme. A final observation was that not only the amount of communication changed with the degree of interdependence in the task, but the pattern of communication varied as well.

These initial results come from the first of a sequence of planned experiments of increasing complexity and we cannot make any firm conclusions, but they support the premise of Coactive Design and demonstrate that even in simple tasks, the coordination involved in managing the interdependence can be quite complex.

6.2 Results from Recent Work

There are several examples from recent HRI work that support our approach. Fong's [3] work demonstrated the support of frail autonomy by making the obstacle avoidance activity a participatory one with matching reciprocal functionality. Stubbs [33] noted that as autonomy increases, transparency became the biggest problem in a remote rover. This is a real world example of how autonomy solves some problems, but at the same time creates new issues that we feel are a direct result of the active nature of the task. These examples and our preliminary study highlight the importance of understanding interdependence and using this understanding to shape autonomy.

6.3 The Nature of Autonomy

Autonomy is inherently frailty. Robots, like their creators, will always be imperfect. This underlying truth necessitates human involvement at some level and accentuates the importance of teamwork. Frailty means one will have unexpected events (failures). You cannot overcome failed autonomy with autonomy, but you can with teamwork (e.g. Fong’s collaborative control [3]).

Additionally, Christofferson and Woods [28] describe the “substitution myth”: the erroneous notion that automation activities simply can be substituted for human activities without otherwise affecting the operation of the system. Even if frailty were not an issue, the “substitution myth” reminds us that autonomy is not removing something, but merely changing the nature of it. Humans cannot simply offload tasks to the robots without incurring some coordination penalty. This is not a problem as long as we keep in mind that autonomy is not an end in itself in the field of HRI, but rather a means to supporting productive interaction [16].

6.4 The Nature of Coordination

Once a base level of competence is achieved, coordination of joint activity (teamwork at its simplest form) will take on an ever increasingly important role in the design of a system. This trend was noted by Allen who reported that “the only type of interactions supported by a typical state-of-the-art planning system (namely, adding a new course of action) handled less than 25% of the interactions and that much of the interaction was concerned with maintaining the communication (summarizing and clarifying, for example) or managing the collaboration (discussing the problem solving strategy) [1].” Autonomy centered approaches tend to focus on coordination of content (what they intend to do). Coactive Design also includes coordination of the process (physical and mental systems to carry out the former [18]).

6.5 The Nature of People

As agents move toward greater and greater autonomy, several researchers have expressed concerns. Norman states that “the danger [of intelligent agents] comes when agents start wresting away control, doing things behind your back, making decisions on your behalf, taking actions and, in general, taking over [31].” Simply deciding who is doing what is insufficient, because the human will always need to understand a certain amount of the activity.

Additionally, humans are typically the desired beneficiaries of the fruits of the robot labor. We are the reason for the system and will always want access to the system. Not only do we want access to understand the system, but we also want input to affect it. To paraphrase Kidd [32], it is not that human skill is required, but that human involvement is desired.

7. CONCLUSION

We have introduced Coactive Design as a new approach to address the increasingly sophisticated roles for people and agents in mixed human-agent systems. The basic premise of Coactive Design is that, in sophisticated human-agent systems, the underlying interdependence of joint activity is the critical design feature. We have argued that when designing the capabilities that make an agent autonomous, the process should be guided by an understanding of the interdependence in the joint activity. The understanding of interdependence is then used to shape the operation of agent capabilities and enable appropriate interaction. The success of future human-agent teams hinges not merely on trying to make agents more autonomous, but also in striving to make them more capable of sophisticated interdependence.

8. REFERENCES
