Coactive Design

Why Interdependence Must Shape Autonomy

Matthew Johnson[1,2], Jeffrey M. Bradshaw[1], Paul J. Feltovich[1],
Catholijn M. Jonker[2], Birna van Riemsdijk[2], Maarten Sierhuis[3],

[1] Florida Institute for Human and Machine Cognition (IHMC), Pensacola, Flordia, USA
[2]EEMCS, Delft University of Technology, Delft, The Netherlands
[3]Carnegie Mellon University, Silicon Valley Campus, Moffett Field, California, USA
{mjohnson, jbradshaw, pfeltovich}@ihmc.us
{c.m.jonker and m.b.vanriemsdijk}@tudelft.nl
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 the critical design feature. When designing the capabilities that make an agent autonomous, the process should be guided by an understanding of interdependence within the joint activity. This understanding can then be used to shape the implementation of agent capabilities so as to 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 interdependent activity.

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 teleoperated tools, to more sophisticated partners or teammates [1, 2].

Full teleoperation[1] and full autonomy[2] 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 interaction [3], adjustable autonomy [4], collaborative control [5], and sliding autonomy [6]. 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 dynamically along the spectrum as context and resources change. 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.* [7]. 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 agents 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 with people and with other agents. 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 how this approach relates to prior work, highlighting the new areas of focus. We then address the concept of interdependence itself, which, like autonomy, is a highly nuanced term. In doing so, we provide some insights into both interdependence and the underlying dependencies. We briefly discuss some relevant prior work and then close with a discussion of the importance of this novel approach.

[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."

2 Autonomy

Autonomy has two basic senses in everyday usage. The first sense, self-sufficiency, is about the degree to which an entity is able to take care of itself. Bradshaw [8] refers to this as the *descriptive dimension* of autonomy. Similarly, Castelfranchi [9] referred to this as one of the two aspects of *social autonomy* that he called *independence*. People usually consider robot autonomy in this sense in relation to a particular task. For example, a robot may be able to navigate autonomously, but only in an office environment. The second sense refers to the quality of self-directedness, or the degree of freedom from outside constraints (whether social or environmental), which Bradshaw calls the *prescriptive dimension* of autonomy. 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.

For the purposes of this paper, we will use the terms *self-sufficiency* and *self-directedness* to avoid the ambiguity that often exist in the literature in discussions about autonomy.

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 ones. Parts of our discussion of this topic are adapted from [8].

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 [10]. 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 the concept of *Supervisory Control* [11], in which a human oversees one or more autonomous systems, statically allocating tasks to them. Once control is given to the system, it is ideally expected to complete the tasks without human intervention. Both of these approaches are clearly 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. Therefore, this approach to achieving autonomy is shaped by a system's self-sufficiency.

2.1.2 Adaptive, Sliding, or Adjustable 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. For example, Dorais [12] defines adjustable autonomy as "the ability of autonomous systems to operate with dynamically varying levels of independence, intelligence and control." Dias [13] uses the term sliding autonomy, but defines it similarly. Sheridan discusses adaptive automation in which the system must decide at runtime which functions to automate and to what extent. We will use the term *adjustable autonomy* to refer to all three concepts, with reference to the capability of a system to automatically adjust the robot's level of autonomy, in this case the self-directedness aspect, to some appropriate level, based on the situation. The action of 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 still being on task assignment, control and level of independence. Autonomy, in this case, is shaped by the self-directedness. One very important concept emphasized by these approaches is adaptivity, a quality that will be increasingly important in the operation of intelligent 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" [3]. In Allen's work, the system is able to reason about which party should initiate action with respect to a given task. In a similar vein, Myers and Morley describe a framework called "Taskable Reactive Agent Communities (TRAC) [14], which supports the directability of a team of agents by a human supervisor by modifying task guidance." Directability or task allocation is once again the central feature of the approach. Murphy [15] 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 the other approaches discussed above, mixed-initiative interaction is essentially autonomy-centered. Its usual focus is on task assignment or the authority to act and, as such, the self-directedness shapes design of the autonomous system. 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 [5] 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... allows robots to benefit from human assistance during perception and cognition, and not just planning and command generation" [5]. Collaborative Control is a first step toward Coactive Design, introducing the idea that both parties may participate simultaneously in the same action. Here the ongoing interdependence of the human and the robot in carrying out a navigation task is used to shape the design of autonomous capabilities.

The robot was designed to enable the human to provide assistance in the perceptual and cognitive parts of the task. This assistance is not strictly 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

One way to gain insight into the predominant perspectives in a research community is to review how it categorizes and describes its own work. This provides a test of our claim that prior work in agents and robots has been largely autonomy-centered.

The general drift is perhaps most clearly seen in the work of researchers who have tried to describe different "levels" of autonomy. For example, Yanco [16] characterized autonomy in terms of the amount of intervention required. For example, full teleoperation is 100% intervention and 0% automation. On the other hand, tour guide robots are labelled 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 a very subjective matter except when one is at the extreme ends of the spectrum.

Similarly Parasuraman [17] provides a list of levels of autonomy shown in Figure 1.

HIGH	10. The computer decides everything, acts autonomously, ignoring the human.
	9. informs the human only if it, the computer, decides to
	8. informs the human only if asked, or
	7. executes automatically, then necessarily informs the human, and
	6. allows the human a restricted time to veto before automatic execution, or
	5. executes that suggestion if the human approves, or
	4. suggests one alternative
	3. narrows the selection down to a few, or
	2. The computer offers a complete set of decision/action alternatives, or
LOW	1. The computer offers no assistance: human must take all decisions and actions.

Fig. 1. Levels of Automation [17].

Parasuraman's scale is clearly autonomy-centered, as noted by Goodrich and Schultz [18]. Specifically it focused on the self-directed aspect of autonomy. Goodrich and Schultz [18] provided a scale which attempts to focus on interaction instead of autonomy, shown in Figure 2.

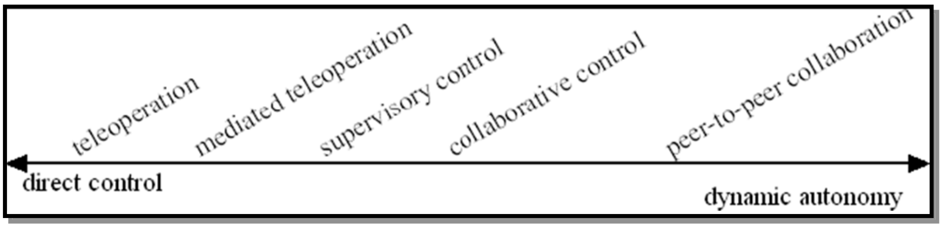

Fig. 2. Levels of autonomy with an emphasis on human interaction [18].

The desire of the authors was to capture something more than the previous autonomy-centered characterizations of the field, but in reality the left-to-right progress of the scale provides little more than a historical summary, with peer-to-peer collaboration as a future direction for research. The label of the right end of the spectrum, “dynamic autonomy,” reveals that this scale is, like the others discussed previously, autonomy-centered.

Bradshaw has characterized autonomy in terms of multiple dimensions rather than a single one-dimensional scale of levels [8]. The descriptive and prescriptive dimensions discussed above capture the two initial senses of autonomy. He also argues that the measurement of these dimensions should be specific to task and situation, since an agent may be self-directed or self-sufficient in one particular task or situation, but not in another.

Castelfranchi suggested dependence as the complement of autonomy [9] and attempts to capture several 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 measured with respect to different aspects of the agent. Castelfranchi put it this way: “any needed resource or power within the action-perception loop of an agent defines a possible dimension of dependence or autonomy” [9].

3 Coactive Design

Coactive Design takes *interdependence* as the central organizing principle among people and agents working together in joint activity. Our sense of joint activity parallels the one explained by Clark [19], especially in its feature that, what one party does depends on what another party does (and usually vice-versa), often over a sustained cycle of actions [20]. Hence, their activity is "interdependent." Their work, thus entwined, needs to be coordinated.

Certainly joint activity of any consequence requires a measure of autonomy (both self-sufficiency and self-directedness) in its participants. Without a minimum level of autonomy, an agent will simply be a burden on a team, as noted by Stubbs [21]. However, we contend that simply adding autonomy is insufficient. The means by which that agent realizes the necessary capabilities of self-sufficiency and self-directedness must be guided by an understanding of the interdependence in the types of joint activity in which it will be involved. This understanding of interdependence can be used to shape the design and implementation of the agent's autonomous capabilities, thus enabling appropriate interaction with people and other agents.

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 [22] from dependence to independence to interdependence. The process is a continuum because at least some level of independence of agents through autonomous capabilities is a prerequisite for interdependence. However, independence is not the supreme achievement in human-human interaction [22], nor should it be in human-agent systems. A completely autonomous human possessing no skills for coactivity would be a useless pariah in ordinary society.

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.

The dictionary gives three meanings [23] to the word "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. These three meanings capture the essence of our approach and we translate these below to identify the three minimum requirements of a coactive system. Our contention is that for an agent to effectively engage in joint activity, it must at a minimum have:

1) *Awareness of interdependence in joint activity*
2) *Consideration for interdependence in joint activity*
3) *Capability to support interdependence in joint activity*

We are not suggesting that all team members must be fully aware of the entire scope of the activity, but they must be aware of the interdependence in the activity. Similarly, all team members do not need to be equally capable, but they do need to be capable of supporting their particular points of interdependence. We now address each requirement in more detail.

3.1 Awareness of Interdependence in Joint Activity

In human-machine systems, like today's flight automation systems, there is a shared responsibility between the humans and machines, yet the automation is completely unaware of the human participants in the activity. Joint activity implies mutual engagement in a process extended in space and time [19, 24]. Previous work in human-agent interaction has focused largely on assigning or allocating tasks to agents that may know little about the overall goal of the activity or about other tasks on which its tasks may be interdependent. However, the increasing sophistication of human-machine systems depends on a mature understanding of the requirements of joint activity.

Consider the history of research and development in unmanned aerial vehicles (UAVs). The first goal in its development was a standard engineering challenge to make the UAV self-sufficient for some tasks (e.g., stable flight, waypoint following, etc.). As the capabilities and robustness increased, the focus shifted to the problem of self-directedness (e.g., what am I willing to let the UAV do autonomously). The future directions of UAVs indicate a another shift, as discussed in the Unmanned Systems Roadmap [25] which states that unmanned systems "will quickly evolve to the point where various classes of unmanned systems operate together in a cooperative and collaborative manner to meet the joint warfighers' needs." This suggests a need to focus on interdependence (e.g., how can I get multiple UAVs to work effectively as a team with their operators?). This pattern of development is a natural maturation process that applies to any form of sophisticated automation. While awareness of interdependence was not critical to the initial stages of UAV development, it becomes an essential factor in the realization of a system's full potential. We are no longer dealing with individual *autonomous* actions but with group *participatory* actions [19]. This is a departure from the previous approaches discussed in section 2.1, with the exception of Collaborative Control [5] which aimed to incorporate all parties into the activity through shared human-agent participation in perceptual and cognitive actions.

3.2 Consideration for Interdependence in Joint Activity

Awareness of interdependence is only helpful if requirements for interdependence are taken into account in the design of an agent's autonomous capabilities. As Clark states, "a person's processes may be very different in individual and joint actions even when they appear identical" [19]. Clark's example is playing, for the same piece of music, a musical solo versus a duet. Although the music is the same, the processes involved are very different. This is a drastic shift for many autonomous robots, most of which were designed to do things as independently as possible.

In addition to the processes involved being different, joint activity is inherently more constraining than independent activity. Joint activity may require participating parties to assume collective obligations [26] that come into play even when they are not currently "assigned" to an ongoing task. These obligations may require the performance of certain duties that facilitate good teamwork or they may limit our individual actions for the good of the whole. For example, we may be compelled to provide help in certain situations, while at the same time being prevented from hogging more than our share of limited resources. Identifying the interdependence in the activity can help us formally capture the implied collective obligations.

3.3 Capability to Support Interdependence in Joint Activity

While consideration is about the deliberative or cognitive processes, there is also an essential functional requirement. We have described self-sufficiency as the capability to take care of one's self. Here we are talking about the capability to support interdependence. This means the capability to assist another or be assisted by another. The coactive nature of joint activity means that there is a reciprocal requirement in order for interdependence to be supported, or to put it another way, there is the need for complementary capabilities of those engaged in a participatory action. 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 for outside guidance. Simply stated; one can only give if the others can take and vice versa. The abilities required for good teamwork require reciprocal abilities from the participating 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.

Most previous systems that have tried to include teamwork-like behavior 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. This is in line with mixed-initiative approaches that aim to allow each participant "contribute to the task what it does best. [3]" However, we are not limiting the contribution of participants merely to a decision about who should perform a given task, but instead allow each participant to contribute to any dimension of the activity, as described at the end of section 2.2. Furthermore, in teamwork, it is not what an individual is best at alone that counts, but what that individual can do that is best for the team. This includes considering situations in which some team members will, of need, perform tasks they are not ideally suited for or would in some other way be suboptimal from an individual perspective.

3.4 Where Coactive Design Is Applicable

Interdependence is a challenging issue for both machines and humans. Coactive Design aims to provide a way to make the design of sophisticated agents (e.g., care taker, medical assistant, coworker, or servant) and human-agent teams easier. 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 more challenging roles mentioned above. The envisioned roles, if properly performed, have a greater 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.

4 VISUALIZING THE NEW PERSPECTIVE

So how does the coactive design perspective change the way we see the agent design problem? In section 2.2, we showed several ways autonomy has been characterized. These ways included ranges, lists, levels and a spectrum (single dimension). Since the capability to perform a task and the authority to perform a task are orthogonal concepts, we separate these two dimensions onto separate axes. Together these two axes represent an autonomy-centered plane of agent capabilities. Coactive Design adds a third orthogonal dimension of agent capability: support for interdependence (Figure 3).

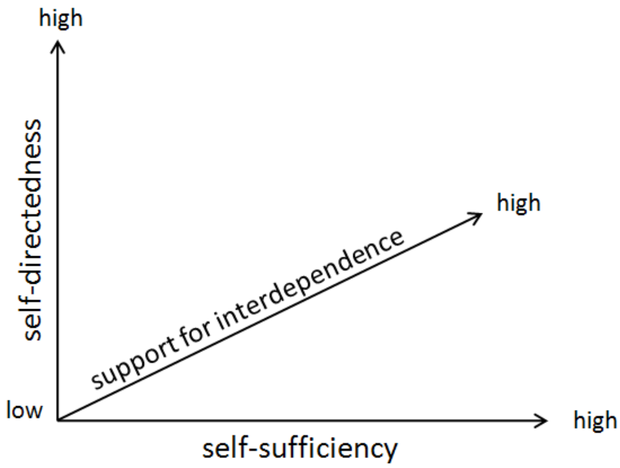

Fig. 3. Support for interdependence as a third, orthogonal dimension of agent capability

In Figure 3, the *self-sufficiency* axis represents the degree to which an agent can perform a task by itself. Low indicates that the agent is not capable of performing the task without significant help and high indicates that the agent can perform the task reliably without assistance. The *self-directedness* axis is about freedom from outside control. Though an agent may be sufficiently competent to perform a range of actions, it may be constrained from doing so by a variety of social and environmental factors. Low indicates that although possibly capable of performing the task, the agent is not in control. High indicates the agent has the authority over its own actions, though it does not necessarily imply sufficient competence. The *support for interdependence* axis characterizes an agent in terms of its capability to coordinate and carry out joint activity. This axis is specifically about the capability to be interdependent, *not* the need or requirement to *be* dependent which are captured by the other axes.

We can now map examples of prior work in autonomy onto this space (Figure 4). The Function Allocation problem of determining what to automate involves a judgment about where a particular agent falls on the self-sufficiency axis. Adjustable autonomy and the most common approaches to mixed-initiative interaction regulate the degree of self-directedness of a given agent. We can also look at how autonomy is characterized in this new model. Yanco's intervention level [16] relates to the self-sufficiency axis while Parasuraman's scale [17] corresponds to the self-directed axis. Bradshaw [8] and Castelfranchi [9] address both axes in their descriptions of autonomy.

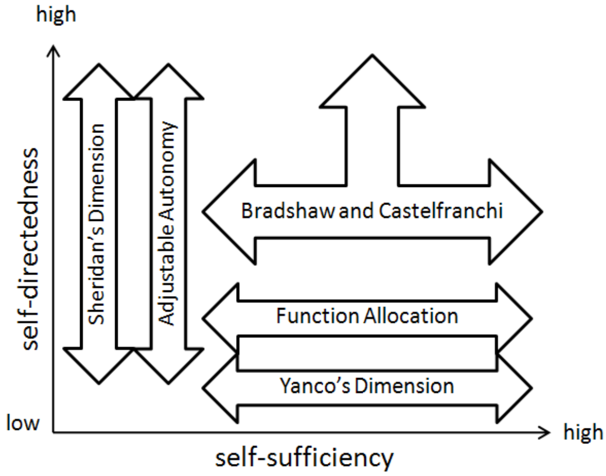

Fig. 4. Mapping Prior Work

Coactive Design presents the unique perspective of the *support for interdependence* dimension. Recall, however, the reciprocal nature of joint activity, which requires that a given participant's capabilities be matched with those of other participants. With this in mind, we have split the third dimension into two parts: one part characterizing the degree to which an agent provides support for others' dependence on it, and the other part characterizing its abilities to deal with its own dependence on others (Figure 5).

With this new perspective, we can now map the concept of Collaborative Control [5] onto this space. As we have already discussed, the most important innovation of this approach was in accommodating a role for the human in providing assistance to the robot at the perceptual and cognitive levels. The key insight of Collaborative Control was that tasks may sometimes be done more effectively if performed jointly. Coactive Design extends this perspective by providing a complementary side to the interdependence axis, accommodating the possibility of machines assisting people.

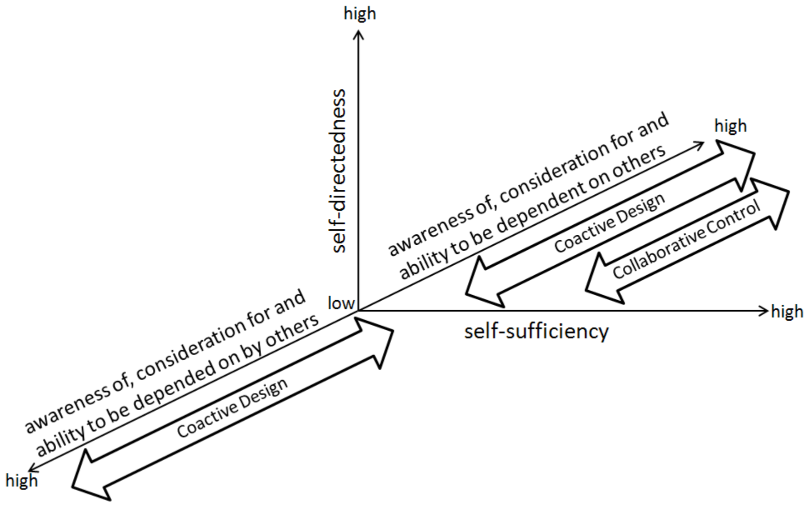

Fig. 5. Mapping to the interdependence axis

Although we are showing a single set of axes for simplicity, The Coactive Design perspective considers all aspects of an agent's sense-act loop. This is directly in line with Castelfranchi's [9] 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.

This article will not be able to address all aspects of Coactive Design, so we will focus the remainder on the first requirement: awareness of interdependence in joint activity. As such, we start with trying to understand different types of interdependence, as well as the specific dependencies that underlie them.

5 Interdependence

Coactive Design has joint activity at its core. Coordination is foundational to joint activity and is required largely because of interdependencies among activities and actors in a working group [27]. Understanding the nature of the interdependencies involved, leads to the kinds of coordination that will be required. This is the first step in awareness and is an important part of determining the capability requirements of agents when designing a solution. Thompson [28] 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. Pooled interdependence has a low level of challenge from interdependence because while team member contributions can be beneficial, they are not essential. Sequential interdependence is about one entity directly depending on the output of another in some way. Reciprocal interdependence is a bidirectional sequential interdependence. These both represent high levels of interdependence because they are completely essential.

All of Thompson's types of interdependence are about the output or product of agents as it affects others. However, there are more subtle types of interdependence that may not be considered core, essential elements of a task, such as progress appraisal (how one's task "is going") and notifying others of unexpected events [29]. These types of elements are usually connected with "teamwork," because they are not directly producing results like "taskwork" does. We will call this type of interdependence *Supportive Interdependence*. Supportive Interdependence is particularly interesting because, unlike the other types suggested by Thompson, it is not easy to generically categorize as involving "low" or "high" interdependence. In fact, depending on particulars, it seems to be able to fit anywhere along the spectrum of interdependence. We will provide more specific examples in the following sections.

5.1 Types of Dependency

We now look a little closer at the types of dependence that make up the interdependence in order to be more aware of interdependence as we design systems. First we will look at the work of Malone and Crowston, who summarize nicely the current work on dependency. Then we will introduce two new types of dependency.

5.1.1 Malone and Crowston

In their interdisciplinary study of coordination, Malone and Crowston [30] summarized prior work on coordination, in which they drew on Computer Science, Organization Theory, Management Science, Economics, Linguistics and Psychology. Like us, they view coordination as required for managing interdependencies. They also characterize 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. We believe Malone's categories can be represented by two basic types of dependency: *resource and temporal*.

Resource dependency can involve a variety of elements including tools, space, the product of a process, or the capability to perform some action. It 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:

$$x \Rightarrow A$$

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

Temporal dependency is the time relation between events or actions. While it is conceivable to view time constraints as a resource as well, this usually 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 $\{A_0 \ldots A_n\}$, where A_0 is the start of the activity and An 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":

$$B_n \blacktriangleright\!\blacktriangleright A_0$$

Table 1 lists a few more examples of temporal dependencies.

Table 1. Examples of Temporal Dependency

Examples of Temporal Dependency	
$B_n \blacktriangleright\!\blacktriangleright A_0$	Process A must start after B is complete
$A_0 \blacktriangleright\!\blacktriangleright B_0$	Process A must start before B starts
$A_0 = B_0$	Processes A must start at the same time as B starts
$A_n = B_0$	Process A must end when B starts
$A_n = B_n$	Processes A must end at the same time as B ends
$Some\ A_{0\text{-}n} = Some\ B_{0\text{-}n}$	Process A must temporary intersect B
$B_0 \blacktriangleright\!\blacktriangleright A_0$ & $A_n \blacktriangleright\!\blacktriangleright B_n$	Processes A must be performed while B is active

Now we will show how Malone's original listing can be represented as combinations of these two types of dependencies. Assuming B generates x, hence x depends on B ($B \Rightarrow x$), 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:

$$B \Rightarrow x \Rightarrow A \quad \textit{and} \quad B_n \blacktriangleright\!\blacktriangleright A_0$$

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:

$$B \Rightarrow A \textit{ and } A_0 \blacktriangleright\!\blacktriangleright | = B_0 \textit{ and } B_n \blacktriangleright\!\blacktriangleright | = A_n$$

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 dependency on the other agent. Lastly, Malone's transfer dependency is similar to the producer/consumer dependency, with the addition of a potentially time dependent exchange of information, which we will call activity C:

$$B \Rightarrow C \Rightarrow A \textit{ and } C_n \blacktriangleright\!\blacktriangleright A_0 \textit{ and } B_n \blacktriangleright\!\blacktriangleright C_n$$

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.

5.1.2 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.

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 [29]("I'm running late"), warnings ("Watch your step"), helpful adjuncts ("Do you want me to pick up your prescription when I go by the drug store?") 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. Supportive Interdependence tends to involve soft dependencies and understanding them can help shape an agent's autonomy to better support interdependent roles.

5.1.3 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 interdependent agents are 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 it 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, leading to a Supportive Interdependence.

6 SUPPORT FOR COACTIVE DESIGN

We provide supporting evidence for our claims from three sources; a preliminary study of our own, results from others' recent work, and some other 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) [31], 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 a particular order. To keep things simple, the blocks are un-stacked to begin with, so un-stacking is not necessary. The degree of interdependence that is embedded in the task is represented by the complexity of color orderings within the goal stack. The task environment is composed of nine rooms containing a random assortment of blocks 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. Teams were composed of two or more players. There are basically two basic tasks in this domain; find a block and deliver a block 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 their interdependencies) 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 pilot study. 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 ordering of the goal stack), we noted an increase in the number 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*. An example of a soft dependency is the exchange of world state information. Since players could only see the status of their current room, they would exchange information about the location of specific colors. Although the task could clearly be completed without this communication, the importance of this soft dependence is demonstrated by the frequency of its use. Progress appraisals about a player's status on a particular task were an example of monitoring dependence. Players very often provided or requested an update when a colored block was picked up. Both progress updates and world state updates are examples of *supportive interdependence*. These types of exchanges typically accounted for approximately 60% of the overall communication and increased in prominence as interdependence increased. 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. For example, during tasks with low interdependence, world state and task assignment were the dominant communications. As interdependence in the task (complexity in the ordering of the goal stack) increased, they both diminished in importance and progress updates became dominant.

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 Human-Robot Interaction work that support our approach. Fong's [5] work demonstrated the support of frail autonomy by making the obstacle avoidance activity a participatory one with matching reciprocal functionality. Stubbs [21] 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 coactive 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

More supporting evidence comes from 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). One cannot overcome failed autonomy with autonomy, but one can possibly do so with teamwork (e.g. Fong's collaborative control [5]).

Additionally, Christofferson and Woods [32] 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, but rather a means to supporting productive interaction [18].

6.4 The Natural Maturation Process

As with the development process of UAVs discussed earlier, 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) [3].”

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 [33].” 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 to have input to affect it. To paraphrase Kidd [34], 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 implementation 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 interdependent activity.

References

1. Bradshaw, J. M., Paul Feltovich, and Matthew Johnson. "Human-Agent Interaction." In *Handbook of Human-Machine Interaction,* edited by Guy Boy, in press. Ashgate, 2011.
2. Klein, Gary, David D. Woods, J. M. Bradshaw, Robert Hoffman, and Paul Feltovich. "Ten challenges for making automation a "team player" in joint human-agent activity." *IEEE Intelligent Systems 19*, no. 6 (November-December 2004): 91-95.
3. Allen, J.E., C.I. Guinn, and E. Horvtz, Mixed-Initiative Interaction. IEEE Intelligent Systems, 1999. 14(5): p. 14-23.

4. Kortenkamp, D., Designing an Architecture for Adjustably Autonomous Robot Teams, in Revised Papers from the PRICAI 2000 Workshop Reader, Four Workshops held at PRICAI 2000 on Advances in Artificial Intelligence. 2001, Springer-Verlag.
5. Fong, T.W., Collaborative Control: A Robot-Centric Model for Vehicle Teleoperation. 2001, Robotics Institute, Carnegie Mellon University: Pittsburgh, PA.
6. Brookshire, J., S. Singh, and R. Simmons. Preliminary Results in Sliding Autonomy for Coordinated Teams. in Proceedings of The 2004 Spring Symposium Series. 2004.
7. Bradshaw, J.M., et al., Teamwork-centered autonomy for extended human-agent interaction in space applications, in In Proceedings of the AAAI Spring Symposium. 2004, AAAI Press. p. 22-24.
8. Bradshaw, J.M., et al., Dimensions of Adjustable Autonomy and Mixed-Initiative Interaction, in Agents and Computational Autonomy. 2004, Springer. p. 17-39.
9. Castelfranchi, C., Founding Agents' "Autonomy" on Dependence Theory, in ECAI. 2000. p. 353-357.
10. Fitts, P.M., Human engineering for an effective air-navigation and traffic-control system. 1951, Washington,: National Research Council, Division of Anthropology and Psychology, Committee on Aviation Psychology. xii, 84 p.
11. Sheridan, T.B., Telerobotics, automation, and human supervisory control. 1992, Cambridge, Mass.: MIT Press. xx, 393 p.
12. Dorais, G. and D. Kortenkamp, Designing Human-Centered Autonomous Agents, in Revised Papers from the PRICAI 2000 Workshop Reader, Four Workshops held at PRICAI 2000 on Advances in Artificial Intelligence. 2001, Springer-Verlag.
13. Dias, M.B., et al., Sliding Autonomy for Peer-To-Peer Human-Robot Teams. 2008, Robotics Institute: Pittsburgh, PA. Myers, K.L. and D.N. Morley. Directing Agent Communities: An Initial Framework. in Proceedings of the IJCAI Workshop on Autonomy, Delegation, and Control: Interacting with Autonomous Agents. 2001. Seattle, WA.
14. K. L. Myers and D. N. Morley, Human directability of agents, Proceedings of the 1st international conference on Knowledge capture, ACM, Victoria, British Columbia, Canada, 2001.
15. Murphy, R., et al. (2000) Mixed-initiative Control of Multiple Heterogeneous Robots for USAR.
16. Yanco, H.A. and J.L. Drury. A Taxonomy for Human-Robot Interaction. in AAAI Fall Symposium on Human-Robot Interaction. 2002.
17. Parasuraman, R., T. Sheridan, and C. Wickens, A model for types and levels of human interaction with automation. Systems, Man and Cybernetics, Part A, IEEE Transactions on, 2000. 30(3): p. 286-297.
18. Goodrich, M.A. and A.C. Schultz, Human-robot interaction: a survey. Found. Trends Hum.-Comput. Interact., 2007. 1(3): p. 203-275.
19. Clark, H.H., Using language. 1996, Cambridge [England] ; New York: Cambridge University Press. xi, 432 p.
20. G. Klein, P. J. Feltovich, J. M. Bradshaw and D. D. Woods, Common Ground and Coordination in Joint Activity, in K. R. B. William B. Rouse, ed., Organizational Simulation, 2005, pp. 139-184.
21. Stubbs, K., P. Hinds, and D. Wettergreen, Autonomy and common ground in human-robot interaction: A field study. IEEE Intelligent Systems, 2007(Special Issue on Interacting with Autonomy): p. 42-50.
22. Covey, S.R., The 7 Habits of Highly Effective People. 1989, New York: Free Press.
23. coaction, in http://dictionary.reference.com/browse/coactive.
24. Sierhuis, M., "It's not just goals all the way down" - "It's activities all the way down" in Engineering Societies in the Agents World VII, 7th International, Workshop, ESAW 2006. 2007: Dublin, Ireland.
25. Office of the Secretary of Defense, Unmanned Systems Roadmap, 2007-2032.

26. Diggelen, J.v., et al., Implementing Collective Obligations in Human-Agent Teams using KAoS Policies, in Proceedings of Coordination, Organization, Institutions and Norms in Agent Systems (COIN). 2009, Springer.
27. Feltovich, P.J., et al., Toward an Ontology of Regulation: Socially-Based Support for Coordination in Human and Machine Joint Activity, in Engineering Societies in the Agents World VII G. O'Hare, et al., Editors. 2007, Springer: Heidelberg, Germany. p. 175-192.
28. Thompson, J.D., Organizations in action; social science bases of administrative theory. 1967, New York,: McGraw-Hill. xi, 192 p.
29. Feltovich, P.J., et al., Progress Appraisal as a Challenging Element of Coordination in Human and Machine Joint Activity, in Engineering Societies in the Agents World VIII A. Artikis, et al., Editors. 2008, Springer: Heidelberg, Germany. p. 124-141.
30. Malone, T.W. and K. Crowston, The interdisciplinary study of coordination. ACM Comput. Surv., 1994. 26(1): p. 87-119.
31. Johnson, M., et al., Joint Activity Testbed: Blocks World for Teams (BW4T) in Engineering Societies in the Agents World X. 2009.
32. Christoffersen, K. and D.D. Woods, How to Make Automated Systems Team Players. 2002.
33. Norman, D.A., The invisible computer : why good products can fail, the personal computer is so complex, and information appliances are the solution. 1998, Cambridge, Mass.: MIT Press. xii, 302 p.
34. Kidd, P.T., Design of human-centered robotic systems, in Human-Robot Interaction, M. Rahimi and W. Karwowski, Editors. 1992, Taylor & Francis. p. 225-241.