Abstract— The majority of robotic systems today are designed by first building a robot that can perform some tasks, adding an interface, and then trying to figure out why the interaction is unnatural and the collaboration is non-existent. Collaboration must be designed into the system from the start. There has been a lot of work on both the interface and the autonomy ends of such systems, but the critical component to facilitate coordination lies in the middle and has had only limited attention in the robotics world. This collaborative middle layer should drive the design of both interface and autonomy. In this paper we will provide a detailed description of the type of collaboration envisioned and the characteristics associated with this type of joint activity. These will be used to establish the design requirements for collaborative human-robot systems and present a cohesive portrait of the essential components necessary to achieve the ideal goal. In addition, some critical challenge areas are highlighted.

I. INTRODUCTION

When one considers the ideal robot it is natural to picture the ones that have appeared in science fiction. Arguably one of the most popular robots ever imagined is R2-D2 from the Star Wars movies. This little robot captured the hearts and minds of many children (and a few adults too). What was so great about this robot? It was not fast or big or strong or heavily armed or exceptionally talented. It had no arms, it could not speak English, it could not walk or fly (until they went crazy in episode II) and was relegated to merely rolling slowly along and beeping. There were two things that made this robot great. First, it had some autonomy. It could get around on its own and perform some simple tasks without assistance. The second, and we argue the most important, capability was interaction; the ability to communicate and collaborate with humans in a variety of ways. This ability not only made the robot’s meager functions useful, it also made the robot predictable, trusted, and endearing.

So what is stopping me from having my own personal R2 unit that I can work with in this ideal manner? Many of today’s entertainment robots have high level behaviors far superior to R2, but still do not provide the same connection one gets when observing R2. The Robosapien performs an array of astounding mechanical maneuvers, and although interesting in the short term, even the most curious child will become bored with the lack of interaction as it only responds to commands. So what about the Roomba? It provides a nice, highly desired function; vacuuming. It also has the required sensors to perform this task autonomously for the most part and provides some beeping as feedback of task completion or failure. So why is this not filling my R2 void that has existed these 30 years? Simply put, you could interact with R2, or more specifically, it could collaborate with you. The difference has been explained [1] as follows; interaction involves action on someone or something and collaboration involves working with others. As Woods has stated, “it isn’t collaboration if you do it all or I do it all.”[12]

The majority of robotic systems today are designed by first building a robot that can perform some tasks, adding an interface, and then trying to figure out why the interaction is unnatural and the collaboration is non-existent. As pointed out by Grosz [1], collaboration must be designed into the system from the start. There has been a lot of work done on both the interface and autonomy ends of the system, but the critical component to facilitate coordination lies in the middle and has had only limited attention in the robotics world, with a few exceptions [2][3][24]. This collaborative middle layer should drive the design of both interface and autonomy. In this paper we provide a detailed description of the type of collaboration envisioned and the characteristics associated with this. These will be used to establish the design requirements for collaborative human-robot systems. We then present a cohesive portrait of the essential components necessary to achieve the ideal goal and highlight some critical challenge areas.

II. ENVISIONED COLLABORATIVE SCENARIO

Imagine you have a robot assistant available to help you in your home. You are a scientist, diligently working on a research paper that is due in a few hours, but the faucet nearby keeps dripping and you are having trouble concentrating. You decide you have ignored it long enough and are finally going to fix it. You send an Immediate Message (IM) to your faithful robot asking to help you fix the sink. Your robot replies by stating that it is vacuuming the upstairs as directed and queries as to whether you would like to postpone that task in lieu of fixing the sink. You affirm the query and task the robot to get your toolbox from the garage, turn off the water to the kitchen, and let you know when all of this is complete. Your robot inquires as to which toolbox you need, the red one or the black one. You clarify that you need the black one. You continue writing your paper and a while later your robot informs you that it is not able to verify that the water to the kitchen has been turned off successfully. You go to the sink, turn on the faucet and confirm that action has been complete and shout “It’s good” down the hallway. The robot joins you in the kitchen with your toolbox and states that everything has
been completed. You begin work under the sink and ask the robot to shine a light on the joint you are working on. The robot retrieves a flashlight from the toolbox and obliges, illuminating the area you gestured toward. You need a second set of hands to tighten the fixture, so you ask the robot to keep the nut from turning. The robot expresses unfamiliarity with the task, so you instruct it how. The robot holds a wrench on the nut as you tighten the fixture. The robot detects some minor slippage and informs you that it may be losing its grip. You adjust the tightness of the wrench for the robot and continue. After fixing the sink, you direct the robot to turn the water to the kitchen back on when you signal, but be ready to turn it off in a hurry. The robot informs you when in position, you signal, the water is turned on and everything works fine. You let the robot know the task is complete and it returns to the vacuuming task while you wonder how you will get your paper done in time.

III. CHARACTERISTICS OF GOOD COORDINATION

The scenario above covers a broad range of activity, both local and remote. It also shows a variety of levels of control. The main feature of the scenario, however, is the collaborative joint activity (interdependent activity—what one party does depends on what others do, and visa versa [25][4]). Indeed, robots can respond to orders: they can vacuum, retrieve objects, and turn wrenches, but it is the collaboration that is truly challenging. So what are the characteristics that would make a good collaborative (successfully interdependently acting) Human-Robot system? In this regard, we pattern our designs after models of human joint activity to enhance naturalness, take advantage of human expectations, and so forth. The pros and cons of such an approach are taken up briefly in the conclusions.

A. Collaborative Communication

Collaboration is intrinsically dependent on communication to support the coordination needs of interdependent activity. Various modes of communication were illustrated in the kitchen sink scenario, including graphical interface (IM), voice, and gesture. Communication is also conveyed through observable action in the scenario when the team member responds to a request. For example, there is no need to verbally respond to the request for a flashlight since the observation of the light being turned on is sufficient. Signaling team members and being able to interpret signals from team members is vitally important to the coordination required in collaborative work and underlies many other functions. The team members need to be able to establish the team and set a common goal. They need to be able to share state information and intentions in order to effectively construct their plan and their own individual actions. Common Ground is a construct investigated by Clark [4]. It refers to the knowledge, pertinent to the task, shared by collaborating parties. The term itself is neutral with regard to quality; our relevant shared knowledge can be bad as well as effective for conducting our joint activity. The goal is building and maintaining good common ground. This is a critical feature in maintaining effective situational awareness and teamwork. Useful interaction to maintain this common ground includes being able to query situational awareness and team members, and provide feedback during the task, alert others when there is a problem, ask about previous events, and ask about previous decisions. Another useful feature is the ability to clarify ambiguity as was the case in selecting the correct toolbox.

B. Mutual Directability

Another important characteristic is to be able to get teammates to do things for you and for your teammates to be able to get you to do things for them. Both the human and the robot in our scenario directed the other’s actions at various stages of the work. Effective coordination entails more flexibility in the level of control during activities then just straight tasking. It includes the ability to interrupt a task, resume a task, abort a task, and redirect the teammate to a new task. This is typically the first step in human robot projects, and frequently is the extent of the interaction, with the addition of some simple feedback on success or failure. Some more advanced features control the flow of activity and include being able to execute a task in a step by step manner, handle conditional execution, iterative execution, and synchronized execution. The ability to direct a teammate at varying levels of control (teleoperation to autonomy) is another important characteristic for collaboration and has been demonstrated by several systems [13][24]. Another advanced characteristic of good interdependent work is the ability to set bounds on the actions of another to allow the other the freedom to accomplish some subtask, while not giving up authority of the task as a whole. This has been demonstrated by some recent policy based approaches [3][26]. A robot must also be able to take the initiative when it needs assistance or has observed a deviation, and direct the human as necessary, as demonstrated in several systems [14][15][16][24]. It is also desirable for robots to be able to suggest alternatives and point out deviations from the plan, or note issues that might impact the team’s decision to continue with the plan. Another important feature of collaborative interaction is the ability to synchronize tasks. These features all involve mutual directability, and there has been much work in this area, with the cited examples being only a small sample.

C. Team Modeling

In order to work with something or even on it, one must have an idea what it can do. In our scenario, the human knew what the robot was capable of doing (retrieval, holding a light, informing on completion, etc.) and what it is not capable of doing or not trusted to do (fixing the sink by itself). The scenario also included a flaw in the model, when the human was unaware that the robot did not know how to keep the nut from turning. Norman has stated that people develop internal, mental, conceptual models of the way the device works, and they form those models from
their expectations and experience with the device itself. For this reason, the device must project an image that is effective in helping the human develop this conceptualization accurately [7]. It seems clear that robots would need to project an image through their physical embodiment, or other interface for remote operations, that provides this information. This can be done through appearance and/or action, but is critical in order to effectively coordinate. It is important to note that this modeling process needs to occur for all members of the team, the humans and the robots, although we acknowledge that robots will not have the same capabilities as humans in such transactions, and we need to account for this in the design process.

D. Mutual Predictability

With appropriate models of the other parties in hand, and an idea of how to direct the behavior of teammates, it is important that team members behave predictably. If asked to get the toolbox from the garage and the robot instead went upstairs, this would probably raise concern in the human team member. The human might wonder if the request was received or if it was understood incorrectly. Mutual predictability is another essential component for effective coordination. Without it, it would be very difficult to perform joint tasks. Predictability requires a model of the teammate, shared knowledge about current state and intentions, and coordination devices. Predictability can apply to the coordination itself. One model of coordination has to do with the kinds of exchanges that should take place to support effective common ground. It has four basic parts and is called the "joint action ladder."[4] This model provides a simple predictable framework for communication and addresses both what communication steps are unnecessary and when extra steps are required to maintain effective common ground. Predictability also includes knowledge about self and partners, such as the ability to obtain information about constraints, capability, availability to help with joint tasks, and authorizations. These abilities enable the formation of plans based on models of all the pertinent team members. To function effectively in coordination, the robot must also provide feedback on progress; hindrances and failure consistent with context [23]. This feedback plays many roles, including establishment and refinement of the team model, support for predictability and maintenance of common ground by providing explanation for any activity that may seem counter to current team expectations. In our scenario, the robot may explain that it needs to turn off the lights upstairs before proceeding, eliminating the human’s confusion and providing an opportunity for redirection.

E. Learning

Although not necessarily essential to collaboration, in order for the interaction to be natural and acceptable to humans, the robot teammates will need to exhibit some learning. Learning is a huge research area and it is outside the scope of this paper to address all the ways it can impact collaboration. The one area we would like to address involves coordination devices. Learning how, when and what to exchange with others may be initially identified for a give task, but it would be beneficial if these could also be learned or taught during the course of work. This will also be critical to humans tolerating their automated assistants. That is, human teams’ interactions will change over time as they gain experience with both the task and each other, and robots will need to adapt similarly. Rigid communication protocols will not be acceptable or adequate. For example, teammates might increase their interaction as drop-dead points are being approached in critical areas or reduce interaction when things are going smoothly to avoid becoming disruptive. In our scenario, the human may want to know when the robot starts and completes the vacuuming of each room to allow for personal inspection. After developing trust in the system, the human may not want such interruptions. There has been some investigation into the social aspects of coordination [17][25][27] but not much on how to learn such social graces.

IV. HOW THESE NEEDS IMPACT THE DESIGN OF THE INTERFACE AND THE ROBOT

To fulfill these design criteria each area will need to be addressed from both the human and robot side of development.

A. Collaborative Communication

A common language is helpful in achieving graceful interaction. There has been some work on open extensible XML based languages to describe robot actions [10][11]. These types of semantic languages have the advantage of allowing the attachment of context as well as content to information exchanges. It is important that these languages be used to capture not just a set of robot actions and properties, but also include extension for communication and display actions and properties necessary for collaborative tasks. We have developed some of these extensions in our demonstration of coordination on a mixed team of humans and robots [24]. It included semantic language to describe coordination devices like responding to non-observable requests and providing progress appraisal. Just as operators will need to know the functional capability of a robot in order to command a robot, they will likewise need to know the collaborative capabilities.

Communication mechanisms are also a critical component. These need to be described in the robot model so humans will have a better idea of how to communicate with the robot, as well as what type of communication to expect from the robot. The robot will need a similar model of the human. These models will need to be context sensitive as the situation changes, for example, when the human and robot are remote as opposed to when they are collocated.

A big challenge is the use of natural language and gesture as a communication mechanism. While these are the most natural modes for the human, the machine task of recognition and interpretation is daunting. There have been several projects working in this area, for example [18][2][24], but the problem is far from solved, particularly
as we move out into the noisy and unpredictable world in which humans typically operate.

Another challenge is how to achieve display behavior from non-humanoid robots. There has been plenty of work in animation, entertainment, and some research fields [2] using humanoid robots, but most robots being used today are not humanoid. Can we achieve similar communication effectiveness with a less capable robot? Alternatively, is there a cheap and easy way to augment non-humanoid systems to provide the display behavior vital to the coordination inherent in joint work? Can we develop a basic set of interaction/coordination devices, having a theoretical basis? A start might come from examination of displays and signals that are common to many animals across species, indicating some special instrumentality [9]. Signals such as attentiveness, warning, and availability enable animals to participate in simple forms of joint activity, and to support coordination. Interestingly, they are also important to human joint activity and coordination. The challenge becomes even more complex when forced onto a remote graphical interface. How can we avoid losing all of the physical cues available during local coordination of team members with effective physical coordination devices when operating remotely? One possibility is the combination of effective modeling and the utilization of coordination policy to ensure coordination occurs through the appropriate channels given the specific context [3]. We have demonstrated some context sensitive coordination in our work [24], where we enforce acknowledgement of requests based on whether the resultant action is observable or not. For example, asking a robot to “turn left” does not need an acknowledgement if requestor can observe the robot turning. However, asking a robot to “join a team” might require an acknowledgement since there may be no outwardly observable cue that the action has taken place.

B. Mutual Directability

There have been great strides in robotic system autonomy. Ironically, the more flexible and autonomous robots become, the less predictable they become. As roboticists develop more complex behaviors it is vital to include coordination mechanisms in the design. Coordination is a key component of situational awareness and maintaining common ground during joint, interdependent activity. Despite the fundamental relationship between enhanced autonomy and the need for more capable coordination, it seems autonomy has received more research attention than the corresponding coordination needs (except for occasional failure messages). There has been significant work devoted to humans directing robots at varying levels of control and some effort to use uncertainty in perception as a cue to direct the need for human intervention [14]. The mixed-initiative interaction area has also provided some valuable insight [18]. Despite these advances, we need to investigate further how we can leverage the robot’s abilities and perception to proactively collaborate when appropriate. The fragile nature of both robotic ability and perception make this a difficult problem, as is the intelligence required for the robot to make such decisions.

C. Team Modeling

Modeling has always been a challenging issue in robotics. There are issues of model inaccuracy and frailty, and these highlight the need for maintenance of effective common ground. For coordination to be successful, assumptions about the other members of the team will need to be made, based on some knowledge and experience. Modeling of humans has been a topic of research in many fields, especially in the software agent community. It will be important to consider lessons learned from these areas and extend them to include models of robots as well. It will also be necessary to consider how the models must change as the roles of team members change [27]. Several important social roles have been identified by Scholtz [22]. In our recent work [24], we have shown dynamic team formation and dynamic role assignment/reassignment, highlighting the impact of these roles on the interaction of the participants, both human and robotic [27]. The frailty of the model can be included as part of the model itself, to guide collaborative interaction. For example, using metrics provided by Olsen and Goodrich [19], coordination frequency can be matched to an evaluation of neglect time. Ultimately, only real world evaluation of systems will determine the accuracy of models developed for interdependent human-machine work. Such work is being conducted in a few places, e.g., at NASA [20].

Robot designers need to take on the challenge of creating robots that can convey their capabilities, state and intentions. A similar challenge is put to graphical interface designers for providing the same capability through a two-dimensional display. The point is that people develop internal, mental, conceptual models of the way the devices work, and they form those from their expectations and experience with the device itself. For this reason, the device must project an image that is conducive to promoting this conceptualization realistically [7]. The information required to yield a good, coherent, accurate conceptual model of a device's operation is not so well known [8].

Another challenge is how to model coordination costs and to be able to intelligently make decisions about the effectiveness and cost of coordination in maintaining common ground. For instance, researchers have demonstrated how even people need to be prudent about when and how often they ping a team member, that is, appraising the member's degree of "interruptability." Too much pinging and nobody gets anything done, the messages can get all tangled up and confused, etc. Too little, and the common ground among the team members starts to deteriorate.

D. Mutual predictability

In order to have mutual predictability, each component must have a model of the other. The human’s model of the robot is often overlooked. To paraphrase Norman [7], it is important for people to have a good conceptual model of a robot in order to coordinate with it. Many researchers have validated the personification and expectations people naturally assign to robots. It is important to develop mechanisms and/or protocols that help humans develop an
accurate model. Some possibilities include Norman's suggestion to not have flawless, complex speech output at a level far more sophisticated than can be understood by the robot. Other possibilities involve investigating animal display behaviors to find corollary robotic behaviors [9].

Predictability also includes abilities such as to obtain information about constraints, feasibility, capability, availability and authorization. This information can not be buried inside the autonomy. In particular, constraints on the system need to be transparent to allow for an effective coordination. In this regard, making constraints external and transparent is another benefit of a policy based approached to robot control [3][26].

Having robots with a model of themselves is not a new idea since any planning system requires some version of this. However, these models are usually focused on an agent's functional abilities and do not usually include its collaborative capabilities. There are also examples of modeling the human [2] and human activity [20]. Another important capacity is to be able to translate between models for appropriate team planning. This includes spatial reasoning to be able to account for perspective differences, as demonstrated by projects like those at NRL [21] and IHMC [24].

Common Ground maintenance requires an extensive implementation of joint activity theory. Some very useful existing examples are [2][24]. These constituted a very thorough approach to designing for coordination, using natural affordance and display behavior that was intuitive to humans. These experiences raised the question about whether there is a core set of coordination functionality that is domain independent and can be reused across applications. As noted earlier, an example is the signals and displays for coordination in joint activity that have found to be nearly universal across species in the higher animal kingdom, suggesting their core instrumentality [9]. We are extending these core coordination mechanisms into those used cross-situationally in higher realms of human joint activity [25]. Maintaining common ground will be a core component of any successful collaborative system. Success at complex collaborative tasks will depend not only on the ability to maintain common ground, but it will also require a way to monitor common ground of the coordinating parties to determine when repairs are needed.

One of the biggest challenges of coordination is identifying what information to send, when to send it and how it should be sent. Should this be expressed in the model, learned, or some hybrid of the two? This will require cost analysis to determine how and when to make repairs to mutual understanding. It will also require display conventions to establish, monitor, and repair common ground as necessary. In addition, how do we manage attention for making signaling work? Context-based protocols need to be developed.

Typically, robots attempt to perform a task and, at best, signal success or failure. A big challenge is for them to be able to monitor their own progress, and even harder that of other team members, to anticipate potential problems and interact to provide advanced notice to team members of performance degradation. How do we or the robots determine that a situation is approaching the limits of our robots capabilities?[8] David Woods believes this separates human and automation capability pretty severely. Humans are able to judge "how things are going." For example, "I am having trouble. I don't think I'll be able to be done in time" or "If we do not get that part today, we are in trouble" or "This is taking me longer than usual." Progress appraisal within automation is a critical problem and a major challenge, one that our research program is addressing [23][24].

E. Learning

As human and robotic systems collaborate in work, particularly over extended periods of time, it will be important that the robots function more like human team members do, especially in learning about their partners. This involves such things as learning the groups' past precedents for actions, conventions that apply to the work, and guidance for attention through experience [6]. These make interaction smoother and reduce coordination costs with experience. How can we give robots the ability to learn this type of behavior? Among other requirements, this will include learning and adapting to the social norms and preferences of their human counterparts [25].

V. Conclusion

In order to achieve this ideal view of collaborative human-robot interaction, the requirements for such a system must be considered from the start. We have taken our lead from the study of how humans do joint activity to identify many capabilities required for good collaboration, highlighting specific requirements and identifying the many challenges involved. This human-centered approach is good because it brings along with it naturalness, intuitiveness, expectations, and maybe even a level of trust from the human point of view. But we also discuss the dangers of people projecting more capability onto a robot then it actual possesses. This tendency can be accentuated to the extent that the robotic components seem to be acting sensibly, one of our aims. Hence, there need to be safeguards that help keep the capabilities of the automation "honest." We have addressed some of these in the form of transparency of capability, intent, and degree of progress, as well as various forms of directability, and policy-based control. Many projects are working on significant pieces to the collaboration challenge. It would be of great value to begin to assemble some of the pieces into an open source toolbox for human-machine joint activity. Development of this type of coordination middleware would be of great benefit and provide guidance to both the interface and robotic systems developers. The example robots from science fiction suggest how effective simple robots can be in interaction with humans, even with meager capabilities. R2 had the wonderful beeps and chirps and whistles that captivated every child in the theatre. These noises provided feedback and displayed emotion of our little tin hero, without fancy dialog or any words at all. He also
was hindered by lack of facial expression and body posture, but made up for this with other behavior displays that although non-human, still conveyed the appropriate meaning. One could ask R2 to do something, check on his progress, interrupt him, redirect him, and trust him to make his best effort. R2 would acknowledge our requests, provide feedback on status, and even provide suggestions at times. With all of its inadequacies, this little idealized hero was a team player and exemplified the level of interdependent capability needed for robots to be effective and accepted as our assistants over the long term.

REFERENCES