Dynamic Robot Autonomy: Investigating the Effects of Robot Decision-Making in a Human-Robot Team Task

Paul Schermerhorn
Cognitive Science Program
Indiana University
Bloomington, IN, USA
pscherme@indiana.edu

Matthias Scheutz
Cognitive Science Program
Indiana University
Bloomington, IN, USA
mscheutz@indiana.edu

ABSTRACT

Robot autonomy is of high relevance for HRI, in particular for interactions of humans and robots in mixed human-robot teams. In this paper, we investigate empirically the extent to which autonomy based on independent decision making and acting by the robot can affect the objective task performance of a mixed human-robot team while being subjectively acceptable to humans. The results demonstrate that humans not only accept robot autonomy in the interest of the team, but also view the robot more as a team member and find it easier to interact with, despite a very minimalist graphical/speech interface. Moreover, we find evidence that dynamic autonomy reduces human cognitive load.

1. INTRODUCTION

Suppose a robotic team member in a mixed human-robot team has noticed that the mission objectives are compromised due to unexpected events and has worked out a sequence of actions that is likely to correct the problem and ensure the accomplishment of the task goals. However, repeated attempts at contacting the team leader to obtain permission to execute the plan have failed. What should the robot do? Should it simply assume that the team leader knows what is best and wait, or should it act on its own? A more drastic rendering of the situation would have the robot confronted with a counterproductive command from its team leader that would jeopardize the success of the team mission and prevent it from pursuing its corrective plan. Should the robot warn the human team leader about the command being problematic, act on its own plan to ensure the achievement of the mission goals (thus blatantly disregarding the humans command), or simply comply with the counterproductive command despite the risk?

Situations like the above are likely to occur during task performance in mixed human-robot teams (for they certainly occur in human teams). If a robotic team member is truly *autonomous*–a notion which we will discuss in more detail below–it will have to weigh the importance of two conflicting goals: that of acting based on instructions from the team leader versus that of achieving the overarching goals to which the "obey instructions" goal is subservient. Subsequently, it will have to make a decision based on the comparison that is in the best interest of the team *and* also acceptable to the team leader.

While there is a large body of work in multi-agent systems on agent autonomy that focuses on the effectiveness of distributed decision-making (e.g., [10, 6, 15]), this research is typically not concerned with mixed human and non-human agent groups, and therefore does not address the acceptability of non-human agents making decisions "autonomously" for human team members. Past work on autonomy in HCI and robotics (e.g., [9, 4, 1]), while concerned with evaluating the human factor–in particular, the extent to which varying degrees of autonomous robot behavior are acceptable, desirable, or useful to humans–does not involve or address the robot's capacity to make "independent decisions" other than "decisions" in the service of a human command (see below for an elaboration), thus effectively eliminating what we take to be a *critical aspect* of robot autonomy.

We believe that both aspects of robot autonomy, its contribution to the overall team performance and its acceptability for human team members, need to be investigated together, as their effects and impact might be mutually dependent. Hence, in this paper, we will use a version of the above scenario to empirically examine both objective and subjective components of robot autonomy in human-robot experiments. Specifically, we will use a human-robot team task to investigate whether independent decisions made by robots based on their knowledge of the mission and the overall team goals alone will (1) lead to better team performance and (2) be acceptable to human team leaders.

The rest of the paper is organized as follows. We start with a discussion of the notion of autonomy, which critically underlies our investigation, and specify our meaning of "robot autonomy" (based on notions of autonomy in humans). We then formulate hypotheses about possible effects of robot autonomy in mixed human-robot teams and describe ways to test them experimentally. Following is a detailed description of the experiment conducted in this study and the results we obtained. We also compare our investigation to related work and make suggestions for follow-up experiments.

2. ROBOT AUTONOMY

As with many widely used notions (like "agent", "action", "behavior", etc.), there is no agreed-upon definition of "robot autonomy". We do not propose a novel systematic classifi-

ICMI ICMI-MLMI 2009, Cambridge, MA, USA

cation of different notions; instead, we distill four different, non-exclusive notions of autonomy that can be found in the recent literature. The first (A1) applies when a robot operates outside the direct control of a human: "automation refers to the full or partial replacement of a function previously carried out by a human operator" [13]. The second (A2) describes the case where the robot follows orders, but those orders may leave open exactly what steps should be taken to achieve the task [7]. The third (independent) sense of robot autonomy (A3) comes closest to the notion of autonomy as applied to humans: "robot autonomy" as an "agent's active use of its capabilities to pursue its goals, without intervention by any other agent in the decision-making processes used to determine how those goals should be pursued" [3]. The (A3) sense stresses the idea of decision-making by the agent to pursue its goals, thus requiring the agent to at least have mechanisms for decision making and goal representations, and ideally, additional representations of other intentional states (such as desires, motives, etc.) and non-intentional states (such as task representations, models of teammates, etc.). A fourth, orthogonal, aspect of autonomy (A4) is concerned with a human's perception of the (level of) autonomy of a robot (whatever "robot autonomy" is).

2.1 Dynamic Robot Autonomy

The four notions of "robot autonomy" introduced above are not mutually exclusive, but can all be true of a robot at the same time. However, we believe that (A3) captures aspects of "autonomy" (close to the human sense) that are essentially different from the other three notions and might be important for the performance of mixed human-robot teams. Probably the most striking difference is that (A3) robots "choose" to follow the commands of a human team leader, but essentially never give up their authority to act based on their own decision making. The human command to act in a particular way may be a verbal command for an (A3) robot that is processed, understood, and then used in a decision-making process based on the robot's goals (one of which may be to follow human commands), in which the robot, critically, can decide whether it should follow the command. This is different from robots not capable of (A3) autonomy (but that may be capable of (A2) or (A1) autonomy), which will always follow the command. It is also likely that the (A4) autonomy of (A3) robots is different than that of (A2) or (A1) robots.

Another aspect of (A3) robot autonomy is related to the notion of "adjustable autonomy"–the idea that there are degrees of autonomy (as in (A2)). Clearly, there is a difference between a human operator adjusting the robot's level of (A2) autonomy (e.g., by virtue of pressing buttons on an interface, or giving verbal commands such as "please find a way to move to location X and meet me there at 18:00 hours") and the robot adjusting its level of autonomy itself. But note that in (A3) robots, it is not really the robot's autonomy that is adjusted–strictly speaking, the robot is always autonomous–but rather it is the degree to which the robot will (autonomously choose to) follow commands *in the light of other goals*. In other words, the robot independently decides whether to take initiatives to improve chances of meeting the team goals or to wait for human instructions, in which case it might or might not follow human commands based on whether they will advance the team goals, modulo precautionary measures (e.g., the extent to which the robot can actually evaluate the consequences of a command with respect to team goals). It is the dynamics of decision making under the robot's goals and the environmental conditions that might give the appearance of "adjustment of robot autonomy" in the sense of (A4) to the human observer. Hence, by "dynamic robot autonomy" we will refer to both (A3) autonomy and ways to improve the resultant (A4) (i.e., the human perception of (A3) robot autonomy) to make it acceptable to humans (e.g., because the situation-dependent dynamics are what humans might expect of an autonomous system).

3. ROBOT AUTONOMY IN TEAM TASKS

While any of the above discussed forms of autonomy is relevant to HRI in general, and for the interactions of humans and robots in mixed human-robot teams in particular, dynamic robot autonomy has some unique properties that could be a virtue or a vice in the context of mixed human-robot teams–for example, the question arises: do we really want robots to "choose" to obey, and thus potentially "disobey" commands given by humans, rather than have to follow them at all times?

It seems obvious that nobody in her right mind would want a robot on her team that generally disobeys commands. However, the freedom to make decisions about how to act brings advantages beyond what (A2) autonomy affords. Robots that can make informed, rational evaluations (e.g., based on decision-theoretic principles) might be in a better position to weigh different options in a given situation and make informed decisions about what is best for the team, for many reasons. For one, they might be able to consider a large set of options more quickly, or they might have access to information that human team members do not. More importantly, they might be able to make strictly rational decisions using probabilities of action outcomes for utility calculations, while it is known that people have difficulties reasoning with probabilities and are, furthermore, biased by affect in their decision making [11]. Moreover, communicating decisions to human team members in ways that are accessible to humans that allow humans to understand the robot's reasoning might improve (A4) autonomy and human trust (e.g., see [8] for an empirical study demonstrating that performance accuracy can result in more appropriate levels of trust when users are regularly updated about an autonomous system's successes and failures).

We formulate the above considerations as two empirically testable hypotheses: (H1) *dynamic robot autonomy can lead to better team performance*, and (H2) *people will accept dynamic autonomy when the robot makes autonomous decisions in the interest of team goals*. In particular, (H2) is a critical component of the evaluation of dynamic robot autonomy, for it is not only important to verify (A4) (i.e., that the robot appears to be autonomous to the human), but also to ensure that (A4) autonomy (brought about by (A3) autonomy that might involve overriding human commands) is acceptable (and possibly even desirable) to humans.

3.1 A Team Task for Evaluation

Tasks that can be used to test the utility of dynamic robot autonomy as well as the degree to which people feel comfortable with it, need to meet a minimal set of requirements:

(R1) at least one robot and one human are needed for the task and neither robot nor human must be able to accomplish the task alone

(R2) robot and human must exchange information in order to accomplish the task

(R3) there are measures that can be evaluated based on objective task performance alone and on subjective ratings

(R4) the task must include activities where the robot acts as a teammate and follows the human commands

(R5) the task must include activities that the robot can choose to perform without being directed by the human ("autonomy condition")

(R6) a control condition ("no autonomy condition") is needed where the robot is not allowed to make decisions about what to do unless ordered by the human team leader

Note that while the first three items are common to many joint human-robot tasks, the second three are specific to testing the utility of the dynamic adjustment of robot autonomy.

The specific team task that we will use is a modification of the task previously used by [17]. It takes place against the backdrop of a hypothetical space scenario, where a mixed human-robot team has to investigate rock types on the surface of a planet as quickly as possible within a given amount of time and transmit the information to an orbiting space craft before the time is up. Failure to transmit any data within the allotted time results in an overall task failure. Unfortunately, the electromagnetic field of the planet interferes with the transmitted signal and, moreover, the interference changes over time. Hence, transmission locations shift and need to be tracked over time. Only the robot can detect the field strength, and only in its current position. The performance of the team is evaluated objectively in terms of the total number of rocks inspected.

In this task, the human team leader has the responsibility to (1) find and measure particular types of rocks, classifying them based on their volume into two categories ("small" and "large"), and (2) direct the robot in its search for a good transmission point, also telling it to transmit the data before time runs out. The robot's responsibilities are to (1) follow human commands (e.g., to move through the environment for exploration, to measure the field strength and find a transmission point), and (2) to ensure the data collected by the team leader is transmitted in time.

Note that the task meets the above six requirements: the data can only be transmitted to the orbiting space craft via the built-in transmitter on the robot, and only the human can inspect and measure rocks, hence both human and robot have unique capabilities and both are required to complete the task (R1). Moreover, both human and robot need to exchange information via spoken natural language to transmit data (R2). The *number-of-rocks-measured* and the degree to which people accept the robot's behavior provide objective and subjective performance measures (R3). The robot takes commands from the human concerning measuring the field strength, in which direction to go, or when to transmit data (R4). Moreover, in the autonomy condition, the robot can decide to find transmission locations instead of waiting for the human team leader to direct its search, and can also take the initiative to begin data transmission if time is pressing (R5). In the no-autonomy condition, the robot will always wait for commands from the team leader before it performs an action (R6).

3.2 The Goal Selection Mechanism

While it is impossible for space reasons to provide a detailed overview of the functionality of all components relevant to the robot's goal and action processing in the employed DIARC architecture (see Figure 2 for a diagram), we describe the robot's goal management component, which is responsible for computing and updating the priorities of goals in some detail in order to show how decisions about actions are made in the employed architecture. This is essential for making the argument that the robot is autonomous in sense (A3) (it is clearly autonomous in senses (A1) and (A2), and likely also in sense (A4), which will be tested as part of the experiment). Specifically, to be autonomous in sense (A3), the robot needs to have representations of its goals, which are used to make decisions about what actions to perform. We will briefly describe the representation of *actions* and *goals*, as well as their role in *decision making* in the employed architecture.

Actions are either *simple* (e.g., initiating movements) or *complex* (e.g., whole tasks) and are represented in the form of *scripts*. Each script (i.e., a simple or complex action) has a *goal* associated with it, which is accomplished if script execution succeeds. Goals are thus represented as post-conditions of their associated scripts (a goal can thus have multiple scripts associated with it). Scripts that represent complex actions have subscripts, which, in turn, have associated subgoals. Some goals do not have associated scripts, in which case a problem solving or planning process will attempt to establish an appropriate sequence of actions that can accomplish the goal (i.e., satisfy the post-conditions)–in the present study no planning is required as the robot already has scripts for all its goals and subgoals.

Decisions about what actions to perform or what goals to pursue are established via computations of *goal priorities*, which are determined for each goal based on its *importance* to the robot and its *urgency*, a measure of the time remaining within which to accomplish the goal. *Action selection* is then performed to allow for a high degree of parallel execution while respecting the priorities of goals–roughly, actions in the service of higher priority goals will always have precedence and are able to preempt conflicting actions in the service of lower priority goals.

Let $t_{G,start}$ be the time when a goal G is created and let $t_{G,tot}$ be the time limit for achieving the goal (making the deadline $t_{G,end} := t_{G,start} + t_{G,tot}$). Then the *importance* $i_G(t)$ of a goal G at time t is given by

$$i_G(t) = bE_G - cE_G(t, t_{G,end}) - c_G(t_{G,start}, t)$$

where bE_G is the *expected benefit*, $cE_G(t, t')$ is the *expected cost* from t to time t', and $c_G(t, t')$ is the *actual cost* accrued for G from t to time t'. The *urgency* $u_G(t)$ of goal G at time t is given by

$$u_G(t) = (t - t_{G,start}) \cdot (u_{G,max} - u_{G,min})/t_{G,tot} + u_{G,min}$$

where $u_{G,min}$ and $u_{G,max}$ are the *minimum* and *maximum* urgency ($0 \leq u_{G,min} \leq u_{G,min} leq 1$). Note that $u_G(t)$ is

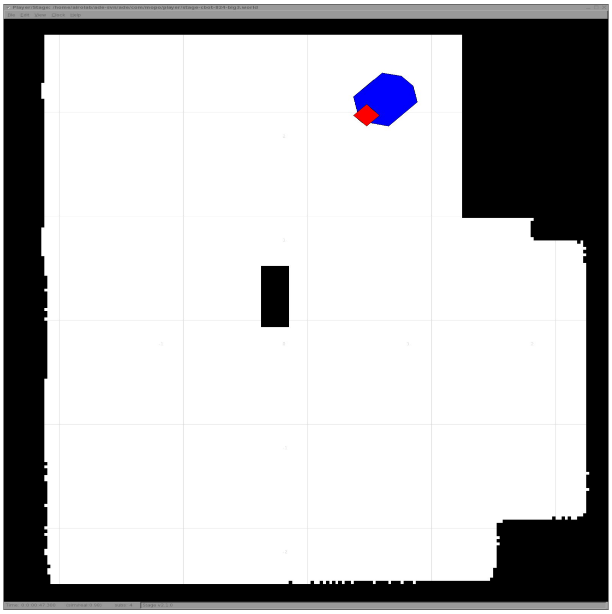

Figure 1: The simulation environment used in the experiment and a map representation of the room.

defined only for $t_{G,start} \leq t \leq t_{G,start} + t_{G,tot}$ and undefined otherwise. The *priority* $p_G(t)$ of a goal at time t is then defined as $p_G(t) = i_G(t) \cdot u_G(t)$.

4. METHODS

We employ a within-subjects design with two robot conditions: the *autonomy* (A) and *no autonomy* control condition (N). In (N), robot will always and only follow human commands and thus not take any actions independently unless so instructed (e.g., the team leader will have to tell the robot to a particular location and take a reading). In (A), on the other hand, the robot will independently take actions based on the priorities of its goals and its action selection mechanism (e.g., it might take the initiative to explore the environment to find a transmission point). Specifically, the robot has an overall *Mission Goal* with three subgoals: to accept and execute commands from the team leader (*Obey Commands*), to find and track transmission regions (*Tracking Goal*), and to transmit the rock information obtained from the team leader in time (*Transmit Goal*). The robot will regularly update the priorities of all of its goals, which, in our case, will effectively be determined by the urgency of the goal (as we will keep benefits and costs constant throughout for simplicity). While the *Obey Command* subgoal always has the same priority (as the urgency to follow commands does not change in this task), the urgency, and thus the priorities of the other subgoals will increase over time to a point where they will exceed the priority of the *Obey Command*.

Participants: 10 subjects were recruited from the general student population at Indiana University; most were undergraduate students.

Experimental Setup: We know from several past experiences with HRI experiments, where humans interacted with physical robots in the same (laboratory) environment, that the physical presence and appearance of the robot matters critically in people's perception of the capabilities of the robot. Since we did not want subjects to be distracted or influenced by the robot's appearance in their evaluation of its abilities to make decisions autonomously, we needed to find a way to remove physical characteristics while keeping the setup intuitive for subjects. We accomplished this using a multi-modal "remote interaction setup" (see Figure 1) where subjects would interact via a big LCD display with a remotely located robot that was depicted in a very generic fashion (i.e., without any particular physical attributes other than a square in the front of an hexagonal body to indicate the robot's heading). The robot was located in a virtual room of approximately 5mx6m identical to the room in which subjects were located (in fact, we used the real robot to build a map of the room for the simulation environment). As a result, subjects could determine that the layout of the room with the robot depicted on the screen looked just like theirs, which helped them imagine where the robot was (as they could easily project the robot's position on the screen into their physical environment). The robot was controlled by the DIARC architecture (in Figure 2) connected to the simulation environment (see the description in Equipment below). During the experiment, a map of the area is maintained, where the virtual field and robot's own location are represented.[1] The field is computed based on a peak location (unknown to the robot) with strength of 450 that decreases proportionally with distance at a rate of one unit per cm. To learn about the field strength at the current location, the robot checks a "simulated field sensor," which effectively returns the field value for the robot's current location based on the peak location.

Procedure: Ten subjects with no prior robotics experience were recruited from the undergraduate student population.[2] The experimenter read the "background story" (summarized in the above task description). Then subjects were told that there were two experimental conditions: *local* and *remote*, and that they were assigned to the *remote* condition where they had to control a remote robot in an environment identical to the room in which they would be performing the measurement task (whereas subjects in the local condition would control the robot in the same room). This was done to strengthen the subjects' beliefs that they were controlling an actual robot. Before attempting the actual task, subjects went through a *practice phase* during which they became acquainted with the robot by interacting with it in natural language. They were encouraged to use commands such as "go forward", "turn right", "take a reading", etc., but were not explicitly instructed about the limits of the natural language processing system. The practice phase consisted of a trial run of the task with the robot in an obstacle-free environment. Subjects were told that they were going to perform three 4 minute runs each in two blocks using two different robot architectures with similar functionality to test the effectiveness of these two architectures. This was to allow us to collect subjective information on the post-survey for each block. Subjects are not informed of the nature of the difference. The two blocks corresponding to the two conditions

[1]Note that the map in this experiment is not a proper part of the robot's architecture.

[2]Although the sample size is small, statistical analysis finds a number of significant (or nearly significant) trends, which we expect ongoing supplemental experimental runs to corroborate.

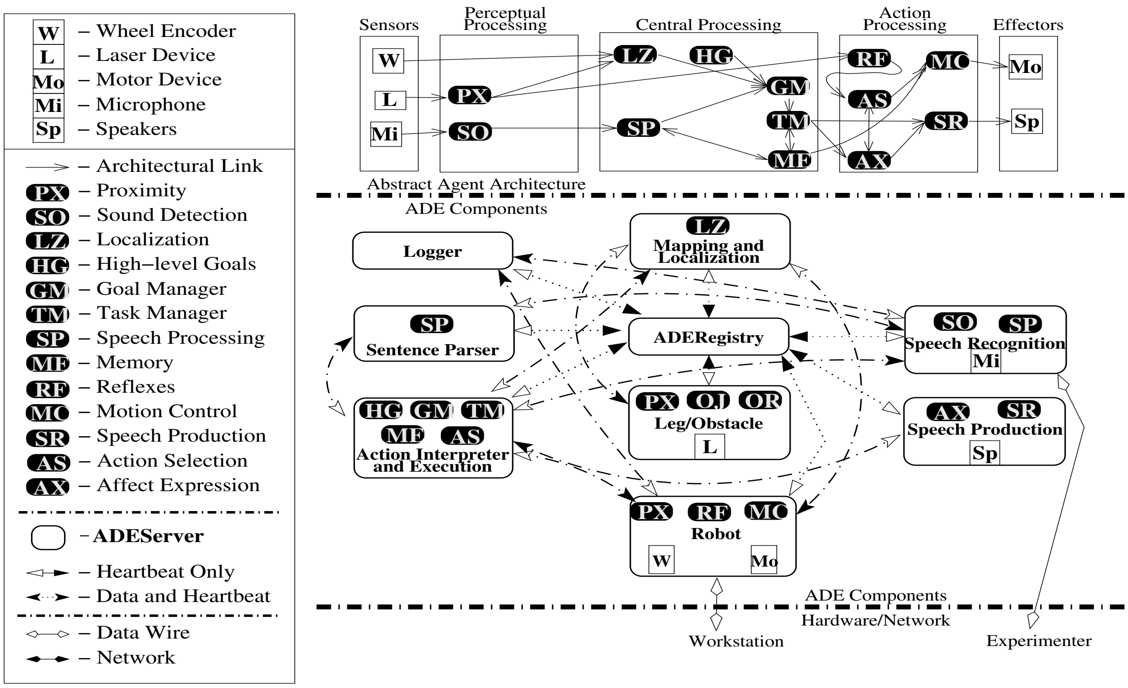

Figure 2: An overview of the DIARC architecture used for the team task components (top), its implementation in the ADE framework (middle), and mapping onto computing hardware (bottom). Boxes depict concurrently running components of varying complexity and arrows indicate the information flow through the architecture.

were counterbalanced across subjects.

In addition to the robot component of the task, subjects were instructed to take "measurements" of "rock formations" in the environment. The measurement component of the task required subjects to locate boxes of a particular color in the environment and solve (two-digit by two-digit) multiplication problems printed on sheets of paper inside the boxes. Subjects were given a clipboard with paper and pencil on which to perform the multiplication. In the final minute of each task run, subjects were required to transmit the data to the orbiting ship. Transmission consisted of reporting an abbreviated version of the multiplication results: when transmission was initiated, the robot would ask the subject how many formations were measured (i.e., multiplications completed) and how many of the products were above a predetermined threshold.

The robot announced the remaining time every 30 seconds. Transmission of data took 15 seconds and was only complete thereafter. The overall task lasted for exactly 4 minutes. Task parameters (in particular the duration of the task) were chosen to make it very difficult to complete all measurements and transmit the data successfully to avoid ceiling effects. The whole experiment lasted for about 45-60 minutes. Throughout the experiment, the robot's motion trajectory, speech produced and detected, and the state of the goal stack were recorded. This allowed us to examine multiple objective performance measures in addition to the subjective survey responses (e.g., the number of items measured or the number of commands issued).

After the experimental run, subjects were asked to fill out the post-survey with questions about their impressions of the interaction. Some of the questions in the post-survey were designed to assess subjects' attitudes toward robots in general. Others asked about the subjects evaluation and comparison of the particular architectures to which they were exposed in the experiment. The survey was administered via computer, with care taken to minimize influencing factors in the interface (e.g., "sliders" were employed for answers on a scale with no initial default slider position; subjects had to click to determine the initial position of the slider and then adjust the slider accordingly). Finally, subjects were debriefed about the simulation setup.

Equipment: The robot model simulated in these experiments was a Pioneer P3AT within the Player/Stage environment (see Figure 1). We employed the goal management system described above as part of the distributed integrated affect, reflection, cognition architecture DIARC used successfully in previous HRI experiments. An overview of the main functional components of the architecture and their mapping onto components in the ADE infrastructure is shown in Figure 2. To eliminate speech recognition errors and their influence on subject performance and thus their experience of the robot (especially regarding their impressions of robot autonomy which strongly correlate with the subjects' impressions of how well the robot understands them, e.g., [17]), we employed a "human speech recognizer": a human confederate was placed in a neighboring room with a headset and instructed to transcribe the human instructions as soon as they were understood (using a simple graphical interface that provided the most common words and a text field for words not depicted on buttons). Note that the graphical interface was run within DIARC as a "speech recognizer" component in exactly the same way as other speech recognizer components are usually run within the architecture (e.g., SPHINX or SONIC (e.g., [17, 5]), hence the robot still had to do all the parsing, natural language understanding, goal generation, decision-making, action selection and execution, etc. (different from full-fledged "Wizard-of-Oz" studies in which the robot is completely controlled by a human wizard). In sum, the "human speech recognizer" allowed us to eliminate the problem of having to sort out the difference between a robot that did not understand a command and one that understood but chose to ignore it, while leaving the robot maximally autonomous (given that the human speech recognizer only functioned as a close-to-perfect speech recognition component in the architecture).

Note that there are no explicit provisions for robot autonomy or "adjustable autonomy" in the architecture. Rather, dynamic robot autonomy is achieved with explicit representations of goals and subgoals, and mechanisms to decide which to prioritize based on circumstances. Table 1 shows the relevant parameters for the three goals used in the autonomy condition in the experiments. The "Obey commands" goal requires the robot to follow the team leader's orders when they are received. The "Tracking" goal requires the robot to move to and stay at a good transmission location. The "Transmit" goal requires the robot to request the data to be transmitted from the team leader and then attempt to initiate transmission. The parameters are specifically chosen such that the robot will engage in autonomous exploration after 150 seconds (if it is not already at a transmission location), and will initiate transmission after 195 seconds (if it has not already been initiated) in an effort to ensure achievement of the overall mission goal. Hence it is possible that the robot might not comply with a team leader command if the command is inconsistent with its current actions during the last 90 seconds of the task. The robot's (initially) subordinate role and particular goal parameters were specifically chosen to create *simple replicable experimental situations* in which it would nevertheless be clear to subjects that the robot was choosing its own actions. [3] In an attempt to make the "disobedience" palatable to the human team leader (i.e., to achieve acceptable (A4) autonomy), the robot will announce changes in its goals ("We are running out of time. I need to find a transmission location.") and give a rationale when it chooses not to comply with human commands ("Sorry, I cannot stop right now, I am trying to find a transmission location.").

5. RESULTS AND DISCUSSION

As part of the task description read to subjects, the importance of finding the transmission location and successfully transmitting data was stressed. Given that it very difficult to complete all measurements *and* transmit the data successfully, we expected subjects in the non-autonomous condition to fail to transmit the data more frequently than in the autonomous conditions simply because they overlooked the time due to the cognitive load imposed on them by the measurement subtask. Hence, we first confirmed that the autonomy condition does indeed help them to get the data transmitted, as expected. A paired t-test comparing the average number of successful transmissions from each subject's autonomy ($M_a = 2.6, sd_a = .7$) and non-autonomy ($M_n = .8, sd_n = 0.8$) blocks finds that the difference is highly significant ($t_{paired} = -6.2, p < .001$), as expected.

Hence, the question arises whether the robot autonomy helped subjects perform their task better, i.e., whether robot autonomy led larger number of measured formations and/or to greater accuracy in the measurements. If (H1) is correct, then subjects should perform better when working with the autonomous robot, due to the robot "taking over" the tracking and transmission aspects of the team task. The results in this regard are mixed. In the number of measurements attempted in autonomy ($M_a = 8.9, sd_a = 3.7$) and non-autonomy ($M_n = 10.1, sd_n = 3.4$), there was no difference ($t_{paired} = 1.5, p = 0.9$). Similarly for measurements completed ($M_a = 6.4, sd_a = 4.0, M_n = 6.4, sd_n = 4.5, t_{paired} = 0.0, p = .5$). Thus, autonomy did not make or allow subjects to work faster. However, there is evidence that subjects were more accurate (i.e., produced more correct measurements) in the autonomy mode ($M_a = 4.5, sd_a = 3.4$) than in the non-autonomy mode ($M_n = 3.5, sd_n = 3.1$), a marginally significant difference ($t_{paired} = -1.5, p = .08$). This is particularly important as the results are within-subjects and thus suggests that individuals were able to devote more cognitive resources to the measurement task with robot autonomy compared to without. It is likely that this result will become highly significant with a larger number of subjects.

Given that subjects were able to transmit data more frequently in autonomous mode and also showed evidence for improved accuracy, it is now interesting to check whether they noticed a difference between the two modes. We examined both objective and subjective measures. For the objective measure, we analyzed the number of commands (motion commands, requests for signal strength readings, etc.) issued by the human team member during the course of the experiment. When in autonomy mode, subjects issued an average of 54.4 commands ($sd_a = 21.7$), whereas in non-autonomy mode, they issued 79.9 on average ($sd_n = 20.0$). This difference is highly significant ($t_{paired} = 6.5, p < .001$), confirming that in autonomy mode, subjects spent fewer cognitive resources directing the robot (thus also lending further evidence to the trend about increased measurement accuracy we described above, which is probably the result of subjects being able to devote more cognitive resources to the multiplications in autonomy mode).

For the subjective measures, we examined the items on the post-experiment survey shown in Table 2, which provide strong support for (H2). Subjects took this survey on a

Table 1: The parameters for all relevant goals in the experimental autonomy condition used in goal prioritization ($b = 1800$ for all goals).

Goal	$u_{G,min}$	$u_{G,max}$	t_{start}	c
Obey Commands	.5	.5	300	600
Tracking	.0	.54	240	400
Transmit	.0	.46	240	200

[3]While the robot's behavior and decision-making can be much more complex with a larger number of goals, which are supported by the architecture, we were aiming at the smallest number of goals that would allow for a thorough human subject evaluation of the mechanism without obscuring the causes of any effects due to too many different potentially contributing behaviors.

Table 2: Survey responses. Subjects were asked the same question for both the autonomous (a) and non-autonomous (n) modes. Shown here are the means and standard deviations for each type and the results of a pairwise t-test comparing the two. Statistically significant results at $alpha = .05$ are printed bold, marginally significant results are printed in italics.

	Performance Measure	M_a	sd_a	M_n	sd_n	t_{paired}	p
	Task Completion	2.6	.7	.8	.8	-6.2	$<$ **.001**
	Measurements Attempted	8.9	3.7	10.1	3.4	1.5	.9
	Measurements Completed	6.4	4.0	6.4	4.5	0.0	.5
	Measurements Correct	4.5	3.4	3.5	3.1	-1.5	*.08*
	Commands Issued	54.4	21.7	79.9	20.0	6.5	$<$ **.001**
	Survey Item	M_a	sd_a	M_n	sd_n	t_{paired}	p
1	The a/n robot was helpful.	7.3	2.6	3.4	2.5	-2.5	**.02**
2	The a/n robot was capable.	7.7	2.1	3.6	2.5	-3.0	**.008**
3	The a/n robot appeared to make its own decisions.	7.8	1.2	3.8	3.0	-3.8	**.002**
4	The a/n robot appeared to disobey my commands.	4.6	3.2	4.7	3.3	.05	.52
5	The a/n robot was cooperative.	7.0	2.8	4.3	3.0	-1.6	*.07*
6	The a/n robot acted like a member of the team.	7.0	2.6	3.5	2.5	-2.2	**.03**
7	The a/n robot was easy to interact with.	7.2	2.7	4.7	2.9	-1.6	*.07*
8	The a/n robot was annoying.	2.7	2.1	4.8	3.3	1.6	*.07*

computer, where they were presented with a series of items to which they responded using a slider. Responses range from 1 (for "strongly disagree") to 9 (for "strongly agree"). Although the survey items are fairly simple and direct, some surprising trends can be found in subjects' answers. From items 1 through 3 in Table 2 we see that subjects found the autonomy mode more helpful and more capable, and also attributed decision-making to the robot in autonomy mode to a much greater degree than to the non-autonomy mode. These items are unsurprising, given the conditions, and the results were as expected, confirming that subjects did, in fact, recognize the difference between the two architectures. However, responses to the remaining items are not as straightforward. Subjects appear to ignore the disobedience (item 4), even though 9 of the 10 issued commands that the robot refused to comply with in autonomy mode. In fact, they rate the autonomy mode as more cooperative (item 5). Hence, subjects do perceive the advantage of autonomy mode and are willing to accept this potentially troublesome aspect of autonomy (disobedience), and on items 6 to 8 seem to even express a preference for autonomy mode: in autonomy mode the robot is viewed more as a team member, easier to interact with, and less annoying.

In sum, subjects do attribute autonomy to the robot in autonomy mode (A4). They accepted robot autonomy and seemed to prefer it, even going so far as to ignore instances of disobedience and attribute greater cooperativeness to the autonomy mode (H2). Moreover, there is evidence suggesting that subjects' willingness to accept autonomy allows them to concentrate more on other aspects of the task, leading to improved performance (H1).

6. RELATED WORK

While it is impossible to do justice to the large literature on robot autonomy in the given space, we will briefly review some related approaches and compare them to our work. [12] propose a robotic architecture that uses visual communication and allows robots to execute various activities autonomously based on decisions made using internal "motive" variables. This is related to the affect variables used in the utility calculation of our decision making system, although making explicit the goal, benefit, and cost representations in our utility calculation allows mechanisms for taking information specific to the situation into account (e.g., urgency and affect). Moreover, communication here is achieved through natural language rather than visual signals.

[1] discuss the potential for improved performance in an urban rescue scenario with six levels of adjustable autonomy using a GUI that automatically makes suggestions as to when a switch in autonomy would likely be beneficial. We take this idea further by defining different types of autonomy, independent of their level, and investigate and empirically quantify the effects on performance when the robot is allowed to take on more (apparent) autonomy without making suggestions to a human.

[10] argue that greater robot autonomy is justified by accounting for greater user neglect; the experiments are designed to determine the appropriate level of autonomy to correct for various lengths of neglect time during a task. We consider "accounting for neglect" to be (A2) autonomy, and furthermore consider other definitions of autonomy.

[16] propose guidelines for an agent architecture that is able to make decisions at an appropriate level of autonomy at a given time based on collected information, reasoning, and actuation of the results. However, the proposal is not empirically evaluated. Moreover, autonomy is adjusted in a centralized fashion, whereas autonomy adjustment in our proposal occurs in a distributed, local fashion in each agent (this has been shown to lead to better performance [2]).

[14] apply adjustable autonomy concepts to an "intelligent assistant" meant to help manage plans and commitments via reminders, conflict detection, etc. The system currently makes suggestions to the user, but is being modified so that the user can allow the program to autonomously implement decisions based on their level of importance. Our experiment applies a similar concept in a more general domain, allowing a robotic agent acting in a dynamic environment to autonomously make decisions and possibly reduce the work-

load of human team members.

7. CONCLUSIONS

This paper presented empirical evaluation of the conjecture that dynamic robot autonomy can significantly improve the performance of mixed human-robot teams while at the same time being palatable to human team members. We find some support for the hypothesis that (A3) autonomy can contribute to performance improvements. Moreover, not only is (A4) autonomy palatable, subjects seem to prefer it over the non-autonomy mode. Although this may seem obvious (for why would they not prefer the mode that they find most competent and helpful?) it is likely that there is a competing desire to have the machine obey commands unconditionally, as that is the behavior we are used to in most machines with which we interact. The results above demonstrate that people can set aside such desires (if they have them to begin with) when the team benefits. Moreover, this is true despite the simple graphical depiction; subjects perceived the autonomy, not because of the interface, but because of the robot's capability.

Future work will explore a third scenario, call it "incompetent autonomy," in which the robot's autonomous behavior is (possibly to varying degrees) *harmful* to the achievement of the team's goals. This will allow us to asses how much the performance improvements affect acceptance of (A4). Similarly, we will explore the use of affect expression in the robot's voice to convey the urgency that the robot is "feeling;" affect expression may make (A4) autonomy more understandable to subjects, making them more likely to accept it (cp. to [17]). Finally, we also plan to replicate these experiments on a physically present robot to determine what effect actual embodiment has on acceptance of (A4); it is possible that people will be much less sanguine about robot disobedience when they share the same physical space.

8. REFERENCES

[1] Michael Baker and Holly A. Yanco. Autonomy mode suggestions for improving human-robot interaction. In *Proceedings of the IEEE International Conference on Systems, Man & Cybernetics*, The Hague, Netherlands, October 2004.

[2] K.S. Barber, A. Goel, and C.E. Martin. Dynamic adaptive autonomy in multi-agent systems. *JETAI*, 1999.

[3] K.S. Barber and C.E. Martin. Specification, measurement, and adjustment of agent autonomy: Theory and implementation. *Autonomous Agents and Multi-Agent Systems*, May 1999.

[4] Jeffrey Bradshaw, Alessandro Acquisti, J. Allen, M. Breedy, L. Bunch, N. Chambers, L. Galescu, M. Goodrich, R. Jeffers, M. Johnson, H. Jung, S. Kulkarni, J. Lott, D. Olsen, M. Sierhuis, N. Suri, W. Taysom, G. Tonti, A. Uszok, and R. van Hoof. Teamwork-centered autonomy for extended human-agent interaction in space applications. In *Proceedings of the AAAI 2004 Spring Symposium on Interaction between Humans and Autonomous Systems over Extended Operation*, Stanford, California, 2004. ACM Press.

[5] Timothy Brick and Matthias Scheutz. Incremental natural language processing for HRI. In *Proceedings of the Second ACM IEEE International Conference on Human-Robot Interaction*, pages 263–270, Washington D.C., March 2007.

[6] Michael Y.K. Cheng and Robin Cohen. A hybrid transfer of control model for adjustable autonomy multiagent systems. In *Proceedings of the 4th International Joint Conference on Autonomous Agents and Multiagent Systems*, The Netherlands, 2005.

[7] Gregory Dorais, R. Peter Bonasso, Daven Kortenkamp, Barney Pell, and Debra Schreckenghost. Adjustable autonomy for human-centered autonomous systems on Mars. In *Mars Society Conference*, August 1998.

[8] Mary Dzindolet, Scott Peterson, Regina Pomranky, Linda Pierce, and Hall Beck. The role of trust in automation reliance. *International Journal of Human-Computer Studies*, 2003.

[9] Michael Fleming and Robin Cohen. A decision procedure for autonomous agents to reason about interactin with humans. In *Proceedings of the AAAI 2004 Spring Symposium on Interaction between Humans and Autonomous Systems over Extended Operation*, pages 81–86, 2004.

[10] Michael Goodrich, Dan Olsen Jr., Jacob Crandall, and Thomas Palmer. Experiments in adjustable autonomy. In *Proceedings of the IJCAI Workshop on Autonomy, Delegation and Control: Interacting with Intelligent Agents*, 2001.

[11] D. Kahneman, P.P. Wakker, and R. Sarin. Back to Bentham? Explorations of experienced utility. *Quarterly Journal of Economics*, 112:375–405, 1997.

[12] Francois Michaud and Minh Tuan Vu. Managing robot autonomy and interactivity using motives and visual communication. In *Proceedings of the 3rd Annual Conference on Autonomous Agents*, pages 160–167, Seattle, Washington, 1999. ACM Press.

[13] Raja Parasuraman, Thomas Sheridan, and Christopher Wickens. A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Sytems and Humans*, 30(3):286–297, May 2000.

[14] Martha E. Pollack and John F. Horty. Adjustable autonomy for a plan management agent. In *Proceedings of the AAAI Spring Symposium on Agents with Adjustable Autonomy 1999*, Stanford, California, 1999.

[15] Paul Scerri, David Pynadath, and Milind Tambe. Why the elf acted autonomously: Towards a theory of adjustable autonomy. In *Proceedings of the 1st International Joint Conference on Autonomous Agents and Multiagent Systems*, Bologna, Italy, 2002.

[16] Paul Scerri and Nancy Reed. Designing agents for systems with adjustable autonomy. In *The IJCAI-01 Workshop on Autonomy, Delegation, and Control: Interacting with Autonomous Agents*, 2001.

[17] Matthias Scheutz, Paul Schermerhorn, James Kramer, and Christopher Middendorff. The utility of affect expression in natural language interactions in joint human-robot tasks. In *Proceedings of the 1st ACM International Conference on Human-Robot Interaction*, pages 226–233, 2006.