Dimensions of Adjustable Autonomy

K. Suzanne Barber[1], Cheryl E. Martin[1], Nancy E. Reed[2], and David Kortenkamp[3]

[1]Laboratory for Intelligent Processes and Systems, Electrical & Comp. Eng.,
The University of Texas at Austin, Austin TX, USA 78712,
{barber|cemartin}@mail.utexas.edu

[2]Dept. of Computer & Info. Sci, Linköping University, Sweden SE-58183,
nanre@ida.liu.se

[3]NASA Johnson Space Center – ER2, Houston TX, USA, 77058
korten@smtp.traclabs.com

Abstract. This panel discussion focused on the dimensions of autonomy that are defined and adjusted by the research of the workshop participants. In particular, the groups were able to explain and relate their work in adjustable autonomy to each other's work, although the topics and application areas are very diverse. We describe three dimensions of an agent's autonomy that are adjustable - the agent's *independence* in how it carries out its goals, its *control* over choosing between options to carry out its goals, and the set of *goals* it has to achieve. Each of these dimensions of autonomy can be adjusted over time.

1 Introduction

The panel discussions at the First Workshop on Teams with Adjustable Autonomy, held at PRICAI 2000, focused on identifying the primary dimensions of agent autonomy that are adjusted by the research work of the various workshop participants. Although this workshop and three previous workshops (Musliner and Pell 1999, Hexmoor 1999, Kortenkamp 1999) have addressed the topic of adjustable autonomy, this topic area is still in its infancy and its boundaries and core ideas have yet to be precisely defined. The work presented at all these workshops has varied widely with respect to the interpretation of “agent autonomy” and “adjustable autonomy.” As a result, it can be difficult to relate the different research approaches in the papers and presentations to one another. The panel discussion at this workshop gave researchers the opportunity to ask and answer the question, “How does each piece of research presented here connect to other pieces of research in the context of adjustable autonomy?” This paper describes the framework that the workshop participants developed to compare and contrast their work on adjustable autonomy.

The concept of agent autonomy is broad, thus no widely accepted definition for this concept exists in the field of agent-based systems. This poses a fundamental challenge for developing a consensus about “adjustable autonomy.” Nevertheless, it is well accepted among workshop participants that “autonomy” can be adjusted and that this endeavor can lead to the development of systems with useful agent behavior. Therefore, the unifying model of adjustable autonomy sought by this panel discussion

R. Kowalczyk et al. (Eds.): PRICAI 2000 Workshop Reader, LNAI 2112, pp. 353–361, 2001.

is developed from descriptions of these useful adjustments rather than from controversial previously proposed definitions of autonomy. The autonomy adjustments considered by this workshop lie in various dimensions. The proposed model of adjustable autonomy attempts to relate these dimensions to one another.

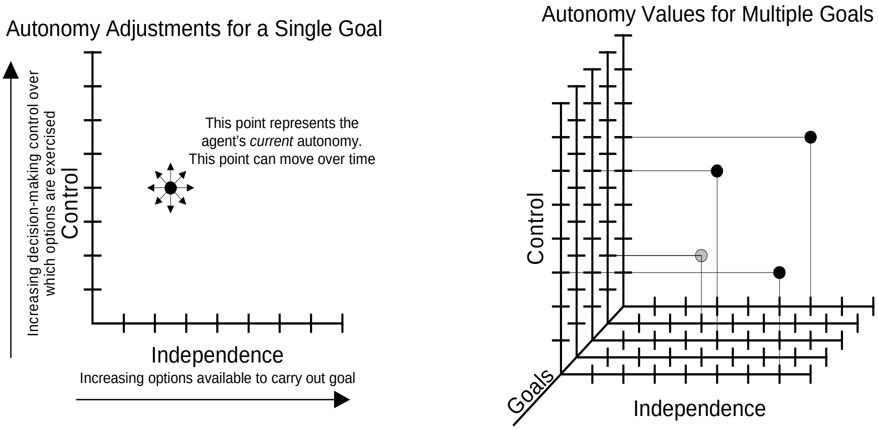

Fig. 1. Relationships of Dimensions of Adjustable Autonomy

As exemplified by the research presented at this workshop, most instances of "adjustable autonomy" research actually adjust different things. This is the fundamental problem faced when relating each piece of research to its counterparts. We identified each of the "characteristics" that were being adjusted as different dimensions of autonomy. Three primary dimensions were identified and related to one another to form a unified context for discussions of adjustable autonomy.

The three primary dimensions of adjustable autonomy we identified are as follows:

(1) **Independence** – changing the number and kinds of options an agent can exercise in carrying out a particular goal. As the number of exercisable options increases, the agent's autonomy increases.

(2) **Control** – changing the degree of decision-making control an agent has over which options it exercises in carrying out a particular goal. As the degree of decision-making control an agent has increases, the agent's autonomy increases.

(3) **Goals** – changing the goals (and/or their priorities) that an agent is pursuing over time. For each goal an agent pursues, the first and second dimensions determine the amount of autonomy the agent has, with respect to that goal.

Each of these three dimensions relates to the others, as shown in Figure 1. An agent is assumed to have one or more goals at all times. For each goal, an agent has a particular degree of control over the decision-making process that determines how that goal should be carried out (autonomy in the decision-making process). For each goal, an agent has a particular set of options or alternatives available for carrying out that goal (autonomy with respect to independence). The following example highlights these three dimensions.

Let *Broker* be an agent whose goal is to “make money for *Investor* over the long-term horizon.” Let *Investor* be a person who uses *Broker* to invest money and who can adjust the autonomy of *Broker* (or on a constituent basis, the autonomy per goal, held by *Broker*). Assume *Broker* has the capability, in general, to invest money in bonds, mutual funds, individual stocks, and precious metals.

Investor can adjust the autonomy of *Broker* by placing constraints on the options *Broker* can exercise to make money. For example, *Investor* can limit *Broker* to investments of only precious metals and bonds. Alternatively, *Investor* can allow *Broker* to invest in bonds, mutual funds, and individual stocks. Here, *Investor* is adjusting *Broker*’s autonomy along the first dimension, independence. *Investor* could allow *Broker* to have maximum autonomy along this dimension by allowing *Broker* to exercise any or all of its possible buying capabilities.

Investor can adjust the autonomy of *Broker* along the second dimension by taking over some or all of the decision-making control for how the money is invested. For example, *Investor* could allow *Broker* to have maximum decision-making control by simply transferring the investment money to *Broker* and letting *Broker* decide where and when to invest it. Alternatively, *Investor* can take away some of the decision-making control from *Broker* and make *Broker* consult with *Investor* about every transaction. At the other extreme, *Investor* could allow *Broker* to have no decision-making control and simply order *Broker* to make each individual transaction that *Investor* wants.

Investor could also give *Broker* another goal or goals, which may or may not conflict with the goals *Broker* already has. For example, *Investor* could give *Broker* an additional goal of “generate monthly income for *Investor*.” The options and decision-making control that *Broker* can exercise for this goal may vary.

These dimensions help describe what is being adjusted as adjustable autonomy is realized. Different dimensions may be adjusted, but in each case, these adjustments occur dynamically, over time, as a system operates. Therefore, in the example, *Investor* may change *Broker*’s goals, control, and independence multiple times in the course of a week or a month. This time-scale may actually become very short for some other types of systems. To realize adjustable autonomy, *Investor* must be able to both increase and decrease *Broker*’s autonomy as the system operates and changes.

The following sections describe research contributions of the workshop participants addressing the realization of these adjustments. Each contribution is discussed within the context of the dimensions of adjustable autonomy introduced above.

2 Independence

Independence refers to the number of different ways that an agent is allowed to try to achieve its goals. The more options an agent has the more autonomous it can be said to be. An example is the 3T architecture developed at NASA JSC (Bonasso *et al* 1997a). The 3T architecture consists of a low-level set of skills that perform actions in the environment, a mid-level reactive planner (Firby’s RAPs system (Firby 1987)) that activates and deactivates sets of *skills* and a high-level planner that places reactive plans (RAPS) on the agenda and monitors resources and time. This architecture allows for independence in many different ways. First, the RAPS system allows the developer to encode different *methods* to accomplish the same tasks. The RAPS system can choose amongst these methods depending on sensed data or previous experience. Typically one of those methods is to ask a human to perform the task (see Bonasso *et al* 1997b). Second, the planner replans when the current plan is failing and attempts to come up with an alternate path to the goal situation. Finally, by breaking the agent's sensing and acting capabilities into low-level primitive skills the architecture allows for easy recombination of these skills to accomplish tasks in novel ways.

Adjusting an agent's level of autonomy with respect to independence can take many different forms. For example, at the planning stage a knowledgeable user can restrict the search space of the planner. Various forms of mixed-initiative planning (Burstein and McDermott 1996) do just that – use human knowledge to guide the planning activities of an agent. This restricts the agent's independence in choosing plans and, therefore, adjusts the agent's level of autonomy. Or a human user can restrict the agent from using a particular method or from using a particular action. For example, we have an autonomous control system for an advanced water recovery system (see Bonasso 2001) and we can tell the control system not to use particular pumps or valves when accomplishing a task.

In general, restricting an agent's independence is a good way to guide it in the right direction without needing to take direct control. This allows the agent to maintain awareness of the situation and react to new data, while still allowing the human to use his or her capabilities. Any adjustably autonomous agent needs to have mechanisms for restricting its actions and guiding its deliberations.

3 Control

The dimension of autonomy described in this section considers autonomy as decision-making control. This viewpoint holds that an agent's degree of autonomy, with respect to some goal that it actively uses its capabilities to pursue, is the degree to which the decision-making process, used to determine how that goal should be pursued, is free from intervention by any other agent (Barber and Martin, 2001a).

The capability of adjustable autonomy with respect to decision-making control is realized in Adaptive Decision-Making Frameworks (ADMF)(Barber et al. 2001a). The following subsection presents a representation of autonomy as decision-making control that is based on the representation of agent decision-making frameworks and decision-making interaction styles. A decision-making framework (DMF) specifies the set of interactions exercised by members of an agent group as the group determines how a goal or set of goals should be achieved. The interactions specified by a decision-making framework are (1) decision-making control relationships and (2) authority-over relationships. A specification of decision-making control dictates which agents make decisions about how to achieve a goal. A specification of authority-over dictates to which agents the decision-makers can assign tasks (i.e. which agents the decision-makers have authority over). An agent's individual decision-making interaction style describes how that agent participates in the overall framework. Agents adopt a distinct decision-making interaction style for each goal they pursue. Agents' decision-making interaction styles can be described informally along a spectrum as shown in Fig. 2.

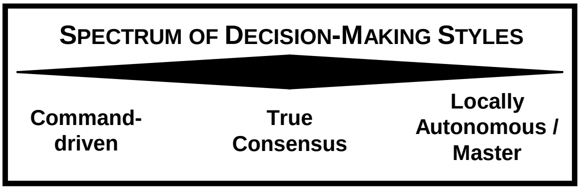

Fig. 2. The spectrum of agent autonomy with respect to decision-making interaction styles.

The three discrete categories of decision-making interaction styles, which define salient points along the spectrum, are labeled in Fig. 2:

- **Command-driven –** The agent does not make any decisions about how to pursue its goal and must obey orders given by some other agent(s).
- **True Consensus –** The agent works as a team member, sharing decision-making control equally with all other decision-making agents.
- **Locally Autonomous / Master –** The agent makes decisions alone and may or may not give orders to other agents.

A decision-making framework is composed of a coherent set of individual decision-making styles for all participating agents (e.g. a Master/Command-driven framework, an All Consensus framework, etc.).

A computational representation of agent decision-making interaction styles is needed in order to assign, evaluate, and modify these concepts in an automated fashion (i.e. to realize the capability of ADMF). Such a representation gives the agent or its designer something to set, a "knob to turn" so to speak, allowing decision-making control to be assigned and adjusted.

Decision-making interactions can be represented by the tuple (*D*, *G*, *C*) (Barber et al., 2000), where *D* identifies the decision-makers and their relative strengths in the decision-making process, *G* identifies the set of goals that are the focus of the decision-making framework, and *C* declares the authority-over constraint. Table 1 presents the specification for this representation, and each component is explained in detail in the following paragraphs. An algorithm for classifying assignments to the (*D*, *G*, *C*) representation according to the decision-making interaction styles defined in Fig. 2 is introduced in (Barber et al., 2000).

The set *D* identifies which agents make decisions about how the intended goals listed in *G* should be pursued. *D* also describes the relative strength of these decision-making agents in the decision-making process. The evaluation of the relative strength of any agent in the decision-making process is based on an analogy to a simple voting process in which every vote must be cast. In this process, every decision-making agent receives an integer number of votes, greater than or equal to one. In *D*, the tuple (a_x , v_{a_x}) represents an agent who is making decisions about how to pursue the goal(s) in *G* along with the number of votes that agent can cast to determine the

Table 1. Specification for Decision-Making Interaction Representation

Decision Making Interaction Representation (*G*, *D*, *C*)		
D	decision-makers	$\{(a_0, v_{a_0}) [, (a_1, v_{a_1}), ..., (a_n, v_{a_n})]\}$, or $\{ (a_1, v_{a_1}) [, ... , (a_n, v_{a_n})]\}$
G	focus	$\{ g_i^{a_0} [, g_j^{a_1}, ... , g_k^{a_n}]\}$ or $\{ g_j^{a_1} [, ... , g_k^{a_n}]\}$
C	authority-over constraint	$\{ a_0 [, a_1, ..., a_n] \}$ or $\{ a_1 [, ..., a_n] \}$

overall decision of the group. Each agent in the set *D* may play an equal part in determining how to pursue *G*, or some agents may have more votes than others.

As the focus of the decision-making framework, *G*, identifies the goal(s) about which agents are making decisions. Any agent may make decisions for goals it intends to achieve as well as for goals that other agents intend to achieve. Additionally, agents may combine their goals for concurrent solution in a “you scratch my back, I’ll scratch yours” fashion. The set *G* may either contain a goal intended by the self-agent (a_0), plus any number of goals intended by other agents (1 per other agent), or *G* may contain only goals intended by other agents and no goal intended by the self-agent. The set *G* must identify at least one goal intended by some agent. If the number of elements in *G* is greater than one ($|G| > 1$), then the decision-making agents must find a solution for all constituent goals concurrently.

The set *C* simply lists the agents who are bound to carry out the decisions made by the decision-makers identified in the set *D*. The decision-makers are said to have authority over the agents in *C* because these agents have previously committed to the decision-making framework, thereby committing to accept task assignments from the decision-makers that are required to carry out the goal(s) identified by *G*. The authority-over constraint, *C*, ensures that some agent(s) will carry out the decisions of the decision-making group. If an agent listed in *C* fails to accept its required task assignment, it must pay a penalty for breaking its commitment to the decision-making framework (Martin, 1997).

Previous experiments exploring the Adaptive Decision-Making Frameworks (ADMF) capability have shown that no one decision-making framework performs best across various situations that may be encountered at run-time (Barber et al., 1999). In fact, the performance of agents operating under a given decision-making framework differs greatly across run-time situations. Given these differences, further research has been undertaken to show that agents capable of ADMF (who implement the decision-making framework that performs best for every different situation encountered) perform better overall than agents who are not able to adapt to various situations in this manner (Barber and Martin, 2001b). Overall these experiments show that implementing the capability of Adaptive Decision-Making Frameworks is a useful method of realizing adjustable autonomy in agent-based systems.

4 Goals

This section considers the adjustment in autonomy of an agent produced by changes to its set of goals or objectives (and/or their priorities). For each goal an agent pursues, the independence and control dimensions described above determine the amount of autonomy the agent has with respect to each goal.

An agent development environment called EASE (End-user Actor Specification Environment) (Scerri and Reed, 2000a, Scerri and Reed, 2000b) allows a user to adjust the autonomy of an actor interacting with a simulation environment by changing the goals the agent is pursuing or can pursue. Examples of actors are a pilot controlling an aircraft in a tactical simulation or a football player in a football simulation. Each actor moves in its environment and its actions are produced by the reasoning of a hierarchy of agents. Each agent in this hierarchy is relatively simple and pursues one goal or objective. An agent accomplishes its goal by contracting other agents (which are then instantiated and fall just below the agent in the hierarchy) to satisfy parts of its goal or objective. A user is able to add and remove agents in the hierarchy, thereby changing the goals or objectives the agent can pursue (adjusting the autonomy of the actor). A user may also change the priorities of goals or change the way a goal is pursued.

The user can make two types of changes in the autonomy of an actor - temporary ones that modify specific agents in the actor, and permanent ones that modify a generic agent or some other variables. The temporary modifications are ones done to the agents that are currently negotiating contracts within an actor. These modifications are adding agents (goals), removing agents (goal and all subgoals), and suspending agents (temporarily inactivating the goal and its subgoals from the reasoning process).

The persistent modifications a user can make include adding new or removing existing types of agents to the pool of available generic agents that can be contracted, adding capabilities to an existing generic agent, removing or modifying the capabilities of an agent, and modifying the importance (priority) of the agent. Since these changes are done to the generic agents instead of to particular instances of agents, the changes affect every agent created from the generic agent in future simulations. A user can also modify the values of constants that are used by agents in their decision-making. This will affect all future decisions of any agent using the constant in its reasoning. This has the potential to affect decision-making control and make alternative solutions available to solve goals.

Certainly there are other ways to change the autonomy of an actor by changing its goals. EASE is a prototype implementation of one set of ways for doing the adjusting.

5 Discussion

Falcone & Castelfranchi (2000) describe levels of delegation and adoption as they relate to adjustable autonomy. They describe two categories of meanings for autonomy. The first is self-sufficiency - not being dependent on others for one's goals. The second is performance or executive autonomy, which the authors divide into 2 subcategories, planning autonomy, how much of the goal planning the agent does for itself, and goal autonomy, whether the agent is allowed to have/find goals.

The dimensions of autonomy described in the previous sections relate to the types and levels of delegation described in Falcone & Castelfranchi (2000) as follows. They describe an n-dimensional space of delegation types to characterize the autonomy of an agent. In their example (n = 3) they show the types: 1) delegation of control, 2) specification-based kinds of delegation and 3) interaction-based kinds of delegation. Their delegation of control dimension ranges from full control to no control. This is approximately the same as the control dimension described above.

Secondly, their specification-based kinds of delegation ranges from closed delegation (complete specification of the task) to open delegation (less complete specification of the task). This is similar to the independence dimension above.

Finally, their interaction-based delegation ranges from strict delegation (contracts) to weak delegation, where an agent believes another agent will accomplish a desired goal on its own. This dimension overlaps both the control dimension and the goal dimensions above. Control: "Who" has control over what goals the agent should achieve? Goals: Can a goal be "forced" upon an agent?

6 Conclusions

The panel discussion was useful in identifying common features within the research presented as well as identifying some differences. This paper also ties the discussion to other work in adjustable autonomy. It seems we have only touched the surface of the definition and use of adjustable autonomy and teamwork in agent systems.

Acknowledgements. Part of this work was supported by the Swedish National Board for Industrial and Technical Development (NUTEK), the network for Real-time education and research in Sweden (ARTES), and the Swedish Foundation for Strategic Research (SSF).

References

Barber, K. S., Goel, A., and Martin, C. E. 1999. The Motivation for Dynamic Adaptive Autonomy in Agent-based Systems. In *Intelligent Agent Technology: Systems, Methodologies, and Tools. Proceedings of the 1st Asia-Pacific Conference*

Barber, K. S., Goel, A., and Martin, C. E. 2000. Dynamic Adaptive Autonomy in Multi-Agent Systems. *The Journal of Experimental and Theoretical Artificial Intelligence, Special Issue on Autonomy Control Software* 12(2): 129-147.

Barber, K. S. and Martin, C. E. 2001a. Dynamic Adaptive Autonomy in Multi-Agent Systems: Representation and Justification. To appear in the *International Journal of Pattern Recognition and Artificial Intelligence. Special Issue on Intelligent Agent Technology*.

Barber, K. S. and Martin, C. E. 2001b. Dynamic Reorganization of Decision-Making Groups. *Fifth international Conference on Autonomous Agents(Agents 2001)*. J. Muller, E. Andre, S.Sern and C. Frasson, Eds., Montreal, Canada, May 30 - June 1, ACM Press, New York, 513-520.

Bonasso, R.P., 2001. "Intelligent Control of a NASA Advanced Water Recovery System," *Proceedings of the Sixth International Symposium on Artificial Intelligence in Space* (i-SAIRAS 2001).

Bonasso, R. P., Firby, R.J, Gat, E., Kortenkamp, D., Miller, D., and Slack, M. 1997a. "Experiences with an Architecture for Intelligent, Reactive Agents," *Journal of Experimental and Theoretical Artificial Intelligence*, Vol.9, No. 1, 1997.

Bonasso, R.P., Kortenkamp, D. and Whitney, T. 1997b. "Using a Robot Control Architecture *on IAT, Hong Kong, December 14-17, 1999*, Liu, J. and Zhong, N., Eds. Singapore: World Scientific, 131-140.

Burstein, Mark and McDermott, Drew 1996. "Issues in the Development of Human-Computer Mixed-Initiative Planning," *Cognitive Technology: In Search of a Humane Interface*, eds. B. Gorayska and J. L. May, Elsevier Science, Amsterdam, 1996.

Falcone, R. and Castelfranchi, C. 2000. "Levels of Delegation and Levels of Adoption as the basis for Adjustable Autonomy". *Advances in Artificial Intelligence '99*, Lamma, E. and Mello, P., Eds., LNAI 1792, Springer-Verlag, Berlin, 273-284.

Hexmoor, H., 1999, Editor, Workshop on Autonomy Control Software, The Fourth International Conference on Autonomous Agents (Agents 99), Seattle, WA, May 1.

Kortenkamp, D., 1999. Editor, Workshop on Adjustable Autonomy Systems, The Sixteenth International Joint Conference on Artificial Intelligence (IJCAI-99), Stockholm, Sweden, August 1.

Martin, C. E. 1997. Representing Autonomy in Sensible Agent-based Systems. Master's Thesis, Electrical and Computer Engineering, University of Texas at Austin.

Musliner, D. and Pell, B., 1999, Editors, AAAI Spring Symposium on Agents with Adjustable Autonomy, Stanford, CA, March 22-24.

Scerri, P. and Reed, N., 2000a, "Creating Complex Actors with EASE", *Fourth International Conference on Autonomous Agents (Agents 2000)*, C. Sierra, M. Gini and J. S. Rosenschein, Eds., June 3-6, ACM Press, New York, 142-143.

Scerri, P. and Reed, N., 2000b, "Real-time Control of Intelligent Agents", Technical Summaries of the Software Demonstration Sessions, *Fourth International Conference on Autonomous Agents (Agents 2000)*, Josep Puyol-Gruart, Ed., June 3-6, ACM Press, New York, 28-29.