# Supporting Adjustable Autonomy in Agent Systems

Nancy E. Reed
Department of Information and Computer Sciences
University of Hawai'i at Manoa, USA
Email: nreed@hawaii.edu

Intelligent agents are being designed to assume more responsibility and for more critical tasks. In some systems, it is infeasible to simply "stop" an agent when it is about to or has done the wrong thing. Rather, we want to reduce the agent's autonomy and give control of the unsatisfactory aspects of the agents task to another agent or person. Adjustable Autonomy (AA) is a solution, meaning that the autonomy of agents and humans in a system can vary dynamically. In this paper we look at human-computer interaction in systems with adjustable autonomy, with the aim of creating more powerful and useful AA in future systems.

## I. Intelligent Agents

Recent developments in the area of artificial intelligence (AI) have centered around the concept of an *agent*. An agent is an autonomous entity that senses its environment and acts intelligently and pro-actively towards its goals [1]–[3]. An important characteristic of an agent is that it has the ability to sense its environment and take actions to affect that environment. Software agents (softbots) sense and act in purely software environments (e.g. an operating system monitoring agent [4]). Other agents have a physical body interacting with a physical environment (e.g. robotic rescue or soccer [5], [6]). Agents can be assigned tasks for which they are faster and tasks that are risky or boring for people. Thus people are freed from menial, dangerous and/or boring tasks. Despite a long list of successful agent applications [7], [8], much has yet to be accomplished for agents to successfully work on more complex tasks, act more intelligently and interact with people, than current systems.

An *intelligent system* is one consisting of intelligent agents and, possibly, humans and/or other conventional software. The assignment of responsibility and authority, i.e. autonomy, is either fixed or may alternate among a small number of fixed configurations. In an intelligent system with AA, the distribution of autonomy among the people and agents may be dynamically configured, thus as the situation changes, an appropriate response can be made that will increase overall system performance.

## II. Adjustable Autonomy

The autonomy agents are given has increased over time, i.e. more capable agents are being developed and given more authority and more responsibility. In most cases, once an agent is designed, its level of autonomy is fixed. In complex environments, an agent will be faced with a vast range of situations, demanding complex reasoning. It is unlikely that any agent has appropriate reasoning mechanisms, sensors and actuators to act effectively in *all* the situations it could potentially face [10]. There will be some situations in which the agent will make unacceptable decisions and, hence, take unacceptable actions – potentially causing harm.

A model of interactions and possible changes in autonomy in a system are shown in Figure 1. There may be one or more agents and one or more people in the system.

There are three types of situations where agents make incorrect decisions. Some situations are so unlikely that the designers ignore them. Hence, the agent is simply not designed to handle them properly. The second type of situation occurs where a situation may be common, however the cost of developing software and hardware to properly handle the situation is deemed too expensive or time-consuming. The third possibility is due due to "bugs" in the agent design.

Most programs are notoriously brittle, e.g. the systems not only fail, but also fail to recognize it has failed or encountered a situation it is not equipped to handle! For example, an inspection robot may be lost when it tries to cross a "puddle" that is in fact several feet deep.

The more autonomy an agent has, over more complex and important tasks, the more potential there is for serious consequences should its behavior be incorrect. Because the agent is autonomous *and* fails to recognize its own inadequacy, termination the agent may be the only means of preventing incorrect actions. However, as agents take on complex tasks with many facets, a person may not be able to perform the task entirely (e.g. a spacecraft). A reasonable solution is to have the ability to stop or change only the *incorrect parts* of its behavior. The useful activities continue to be performed by the agent, while other critical parts are performed by another agent or person.

Adjustable Autonomy (AA) means the ability to *dynamically change* the autonomy of parts of a system, both agent and human. Instead of the responsibility and authority of each one being fixed at design time, they can be changed to best configure the system's autonomy to the current situation. The idea is to dynamically assign autonomy to best leverage the constituent entities' strengths and avoid their weaknesses. Thus, *an AA system is an intelligent system where the distribution of autonomy is dynamically modified to improve overall system performance*.

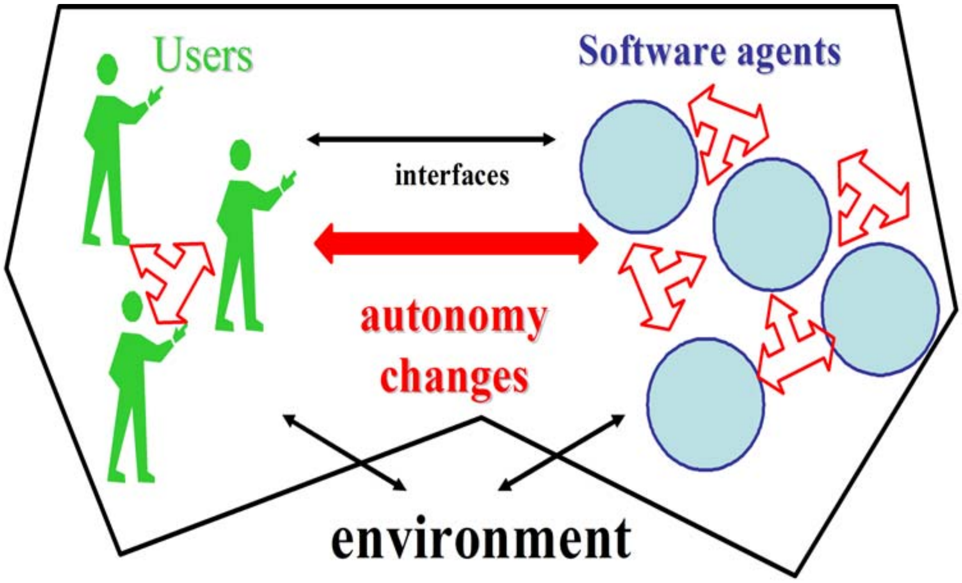


Fig. 1. Adjustable Autonomy in an intelligent system. There may be 1 or more users and one or more software/hardware agents.

## III. Reasoning about AA

A system with AA is a collaborative system. The entities responsible for decision-making, achievement of goals and the allocation of authority are dynamically modified to improve system performance. AA mechanisms manage the changing autonomy by determining appropriate changes in autonomy and implementing those changes.

*A key problem to be addressed when building AA is to determine an appropriate distribution of autonomy and provide mechanisms to realize the autonomy changes.*

The distribution of autonomy should change according to the current system sub-goals, reconfiguring them to best utilize the system's resources. Conceptually the task of changing autonomy can be broken into three parts:

- AA Information (AAI) : Collection of the information relevant to the AA decision making.
- AA Reasoning (AAR) : Reasoning about what autonomy changes could or should be made.
- AA Actuation (AAA) : Realization of the decisions made by the AAR.

AAI provides information on prevailing environmental conditions and the current system state and potential (referred to as *context* in [11]) as are relevant to AA Reasoner. The AA reasoner determines what changes in the autonomy distribution will lead to better system performance with respect to its goals. AA Reasoner might be done either by a human or in software. Finally, the AAA provides the mechanisms for implementing the decisions of the AA reasoning, i.e. it provides the mechanisms for realizing changes in authority or transfer of responsibility.

The AA Information and AA Actuation tightly constrain the design and potential of AA Reasoner. Information not available from the AA Information cannot be used. Similarly, any change that cannot be realized by AA Actuation should not be considered by AA Reasoning. In turn AA Information and AA Actuation are both tightly constrained by the services provided to them by the entities in the system. Hence *the services provided by the entities are critically important to the building of an AA system.* Obviously, we are limited to designing the services that software entities provide (as the services of the human entities are fixed).

Consider an analogy to an architect on a large industrial building. The architects job s to collect information about the needs of the occupants, the characteristics of the site and applicable building codes. Then to decide on the materials and methods used to construct the building, and monitor its construction. No matter how good the architect is at their job they rely on employees in the organization to inform them (either implicitly or explicitly) of the activities that need to be accommodated in the building. If this information is limited or incorrect, the architect's job is significantly harder and the results are likely to be disappointing. Once a decision is made, it must be implemented. No matter how good the decision, if it is not accepted and appropriately implemented, the final result is unacceptable Implementation of AA works in the same way – if the entities do not supply appropriate information and properly implement decisions, the best AAR is useless.

## IV. Discussion

Systems examples are drawn from EASE [12] applications. In EASE, the AA is used to give a user runtime control over the agents in a simulation environment [13]. Fighter pilot agents have been constructed to work with Saab Corporation's commercial flight simulator TACSI [14] on air-combat scenarios.

Agent designs may differ in the way that information is represented and in how easily it can be extracted in an understandable (to the AAR) manner. Whether the AAI services are

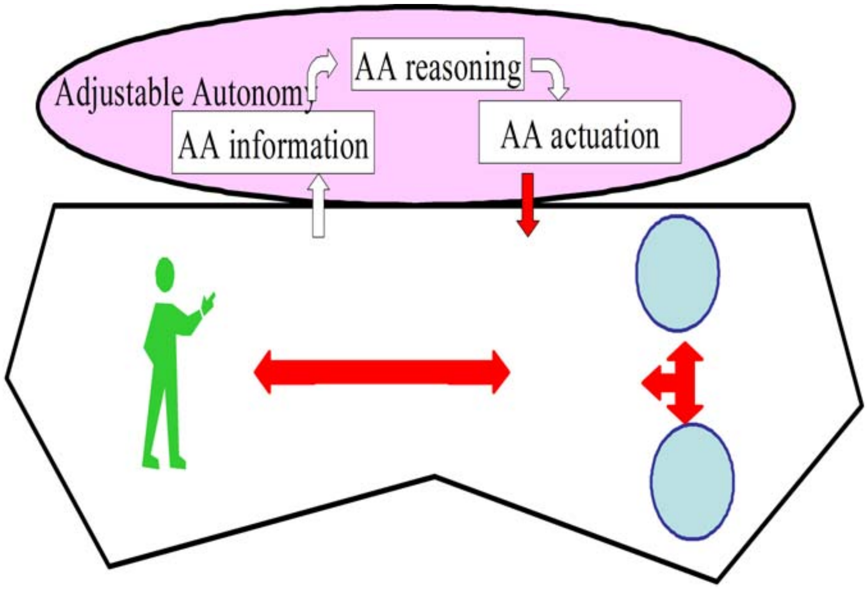


Fig. 2. Adjustable Autonomy functionality includes data collection, reasoning to decide on any changes, and effecting the desired changes.

simple, e.g. providing access to particular parts of the agent's reasoning, or very complex, e.g. needing complex information manipulation, depends on the design of the agent. To have as much understandable information as possible available quickly, presenting unmodified data structures of a running agent to the user. Avoiding the need for complex algorithms to extract information from a running agent increases speed and usefulness of the information for adjustable autonomy.

Furthermore, the more elegantly and smoothly the agent incorporates autonomy changes into its ongoing behavior, the better the behavior of the resulting system. For example, imagine a team of robots playing soccer [6]. The coach wants a player to assist with blocking an opponent. "Blind" reassignment of the nearest player might lead to that agent (with the ball) dropping it! The agent must not terminate a useful action or goal in an unsatisfactory manner, even if it receives an instruction to that effect.

When designing intelligent agents, many requirements, possibly conflicting, need to be weighed. Not all the requirements are related solely to the observable behavior of the agent. For example, there may be particular verifiability, simplicity or computational requirements on an agent [15]. Designers, implicitly or explicitly, follow guidelines when attempting to meet some requirement with a design. A guideline is "a statement or other indication of policy or procedure by which to determine a course of action" [16]. For example, a (simple) guideline to minimize computational requirements for an agent might suggest avoiding algorithms involving significant amounts of search. By following appropriate guidelines designers have a principled, justifiable reason for believing that their design will meet its requirements. For example, if a design avoids extensive use of search algorithms a designer can argue, with justification, that once implemented the design will be computationally efficient (according to the above guideline).

## V. Summary

This paper draws on human-computer interaction (HCI) and autonomous agent research with the aim of aiding in the creation of agents for systems with adjustable autonomy. By looking at human factors and the levels and types of HCI [9], [17], and existing agent systems, we hope to combine the two, often separate areas to move towards achieving more useful Adjustable Autonomy in future systems.

## Acknowledgments

This work has been supported by The Network for Real-Time Research and Education in Sweden (ARTES), Project no. 0055-22, Saab Corporation, Operational Analysis Division, the Swedish National Board for Industrial and Technical Development (NUTEK) under grants 97-09677, 98-06280 and 99-6166, and the Center for Industrial Information Technology (CENIIT) under grant 99.7.

## References

[1] G. Weiss, Ed., *Multiagent Systems*. MIT Press, 1999.
[2] M. J. Wooldridge, *An Introduction to MultiAgent Systems*. West Sussex, England: John Wiley, 2002.
[3] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach, Second Edition*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2002.
[4] H. Song, S. Franklin, and A. Negatu, "Sumpy: A fuzzy software agent," *on line at : http://www.msci.memphis.edu/˜songh/publications/sumpy.html*, 1995.
[5] CRASAR, "Center for robot-assisted search and rescue," http://www.crasar.org.

[6] H. Kitano and M. Asada, "The RoboCup humanoid challenge as the millenium challenge for advanced robotics," *Adv. Robot.*, vol. 13, no. 8, pp. 723–737, 2000.

[7] B. Pell, E. Gamble, E. Gat, R. Keesing, J. Kurien, W. Millar, P. Nayak, C. Plaunt, and B. Williams, "A hybird procedural/deductive executive for autonomous spacecraft," in *Proceedings of the second international conference on autonomous agents*, 1998, pp. 369–376.

[8] P. Rybski, S. Stoeter, M. Erickson, M. Gini, D. Hougen, and N. Papanikolopoulos, "A team of robotic agents for surveillance," in *Proceedings of the fourth international conference on autonomous agents*, 2000, pp. 9–16.

[9] R. Parasuraman, T. B. Sheridan, and C. D. Wickens, "A model for types and levels of human interaction with automation," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 30, no. 3, pp. 286–297, 2000.

[10] B. Shneiderman, *Designing the User Interface*. Addison Wesley, 1998.

[11] H. Hexmoor, "Case studies of autonomy," in *Proceedings of FLAIRS 2000*, 2000, pp. 246–249.

[12] P. Scerri and N. E. Reed, "Real-time control of intelligent agents," in *Technical Summaries of the Software Demonstration Sessions, Fourth International Conference on Autonomous Agents (Agents 2000)*, J. Puyol-Gruart, Ed., June 2000, pp. 28–29.

[13] P. Scerri, "Designing agents for systems with adjustable autonomy," Ph.D. dissertation, Computer and Information Sciences Department, Linköping University, December 2001.

[14] *TACSI - User Guide*, 5th ed., Saab AB, Gripen, Operational Analysis, Modeling and Simulation, September 1998, in Swedish.

[15] P. Scerri and N. E. Reed, "Online control of agents using EASE: Implementing adjustable autonomy using teams," in *Workshop on Teams with Adjustable Autonomy, Sixth Pacific Rim International Conference on Artificial Intelligence (PRICAI 2000)*, N. E. Reed, Ed., Aug. 2000, pp. 25–34.

[16] Houghton-Mifflin, Ed., *The American Heritage Dictionary of the English Language*. Houghton Mifflin Company, 1996.

[17] K. S. Barber, C. E. Martin, N. E. Reed, and D. Kortenkamp, "Dimensions of adjustable autonomy," in *Advances in Artificial Intelligence: PRICAI 2000 Workshop Reader*, R. Kowalczyk, S. W. Loke, N. Reed, and G. Williams, Eds. Berlin: Springer Verlag, 2001, vol. 2112, pp. 353–361.