Adjustable Autonomy, Norms and Pronouncers

Harko Verhagen and Magnus Boman
The DECIDE Research Group
Department of Computer and Systems Sciences
The Royal Institute of Technology and Stockholm University
Electrum 230, SE-16440 Kista, Sweden
verhagen, mab@dsv.su.se

Abstract
Discussions of agent autonomy levels largely ignore the function of norms in the agents' decision making process. Allowing artificial agents to use norms makes smooth interaction in the social space less problematic. Agent autonomy will also increase if one grants agents normative support, making accurate and consistent models of other agents feasible. This will result in better behavior predictions and simulations of not only physical but also social systems.

Introduction
The description of degrees of artificial agent autonomy is problematic from a conceptual point of view. In the field of multi-agent systems (MAS), several authors, e.g., (Conte & Castelfranchi 1995), (Carley & Newell 1994), (Werner 1996) and (Dennett 1981) have developed more or less overlapping hierarchies of autonomy based on the types of processes for decision making that the agents use. This problem will be dealt with in Section 2 below. Our focus will be on control, distributed between human and artificial agents belonging to the same MAS. We will refer to a MAS populated by humans as well as artificial agents, in which the agents model all other agents that they are aware of, as a social space.

One aspect of agent autonomy is largely ignored in the current literature, viz. the level of norms. Designing agents that respect social norms is not only helpful in achieving rational agent behavior, but also enhances the possibility of using accurate models. We will discuss the function and learning of norms and also the results of a simulation study of the learning of norms in Section 3.

Norms can also be used by external counseling devices such as pronouncers (Boman, Davidsson, & Verhagen). A pronouncer is an entity providing normative advice to agents. It is external and usually not wrapped into an agent itself. By contrast, internal procedures are usually called decision modules. The use of pronouncers by idle or disoriented agents reduces the need for human intervention. It also makes the social space less dependent on any outside control, such as an earth-based mission control center (Dorais et al. 1998). Pronouncers are discussed in Section 4.

The need for, and implementation of, adjustable autonomy for governing overriding through human commands is well-documented in the agent literature. One domain in which a MAS with relatively autonomous agents must expect frequent human intervention is intelligent buildings (Boman, Davidsson, & Verhagen). Agent intervention in human activity is less studied but occurs e.g. in surveillance, alarm, and tutoring systems (Rickel & Johnson to appear). Requirements on the simplicity of the agents involved in such systems have so far kept designers from implementing sets of norms. We indicate below why this is likely to change in future designs, and present our conclusions in Section 5.

Levels of Autonomy
In (Lange & Lin 1996), adjustable autonomy was first defined as integrated system management (in the space exploration domain) being “able to be overridden by the operators at any chosen level of shared or traded automatic control”. In (Erickson 1997), adjustable operations are defined more narrowly as “operations which are usually autonomous, but provide shared and/or traded control whenever the crew chooses. Crew-centered adjustable autonomy is defined as giving the crew user of the system control over and improved situational awareness insight into the current, past (in-sight only), and possible future operations of the system whenever the user chooses.” In a more recent report (Dorais et al. 1998), the level of autonomy is defined in this narrow manner, viz. as being dependent on:

- the complexity of the commands
- the number of autonomously controlled subsystems
- the circumstances that will allow for override of manual control
- the duration of autonomous operation

The autonomy models developed in MAS research are summarized and extended in (Verhagen & Smit 1997). In short, decision making in MAS is made at four separate connected levels:

- the level of actions
- the level of plans
• the level of norms

These two definitions of levels of autonomy are quiet different. The basic difference lies in their view on autonomy:

• levels of autonomy as abstraction levels, or the control the agent has over its behavior and decision making process
• level of autonomy as level of independence of coalition

The work on adjustable autonomy is concerned with the second type of autonomy whereas MAS research is primarily concerned with the first. Reconciling these two is non-trivial. The complexity of the commands (from humans to agents) can be measured in terms of the transfer of control (or delegation) that the communication yields in terms of the levels of actions, plans, goals, and norms. Autonomy in the social space increases with the number of subsystems under agent control. This is akin to the action repertoire of the agent. Decision making at the level of norms (or in more simple cases applying norms) is to decide on overriding manual control or for that matter any outside control. The introduction of time as a separate autonomy level is an important addition, although it in fact is merely making explicit the role time plays in the MAS view on autonomy. The levels of decision making are closely linked to time. Both the time span of execution and time span used for self-reflection increase with the levels mentioned.

Simulation of Learning of Norms

In human societies, norms have a dual role in that they both serve as filters for unwanted actions, plans, or goals and as help in predicting the behavior of other members of society. Implemented norms for artificial agents can be used for predicting the behavior of humans in the social space, and thus helps agents maintain a good domain model. It will also enable the agents to become reactive: they can recognize other agents as agents instead of just objects. The use of norms in artificial agents is a fairly recent development in MAS research (c.f. e.g., (Shoham & Tennenholtz 1992), (Verhagen & Smit 1997), (Boman 1999)).

The learning of norms can be divided in two types, viz. the emergence of norms (Ullman-Margalit 1977) and the acceptance of norms (Conte, Castelfranchi, & Dignum 1998). These two types of learning express learning at different levels. The emergence of norms is learning at the level of the social system while the acceptance of norms is learning at the level of the individual agent. In (Conte, Castelfranchi, & Dignum 1998) reasons for the acceptance of norms are discussed. We are not primarily interested in why agents accept norms since we presuppose that membership of a coalition implies the agents accept the norms of the coalition. Instead we are interested in how acceptance of norms changes the decision making behaviour of the coalition (norm-spreading) and by the adaption of the agents' own norms (norm-internalization).

We have conducted simulation studies of the spreading and internalizing of norms as a function of autonomy towards the coalition (at the level of actions). Agents forming a coalition roam around in a two-dimensional world where two resources are randomly distributed. Agents have two decision trees, one containing the subjective evaluation of all alternatives (i.e., a self-model) and one containing the subjective view on the coalition's evaluation of all alternatives (i.e., a coalition model). An agent chooses one of the alternatives and tells the other agents about its choice. Feedback of these agents is used to update the coalition model. The self-model is updated based on the feedback from the environment (i.e., if the chosen alternative is realized or not). Deciding which alternative to choose entails balancing the self-model and coalition model. If the agent is completely autonomous with respect to the coalition, it only evaluates its self-model. If an agent has no autonomy with respect to the coalition, it uses its coalition model. To be able to use the coalition model to predict the behavior of other agents, each agent should have the same coalition model (i.e., share the norms of the coalition). We thus measured the spreading of norms in the coalition.

The spreading of norms is measured as the differences in the coalition utility bases over the agents (i.e., the mean value of the standard deviation per alternative of the coalition utility of that alternative of each agent). We can imagine two situations in which agents totally comply with the group norms. The agents may have no autonomy, or the agents may have adapted to the coalition model to the extent that their self-model equals the coalition model. For this purpose we measured the internalizing of norms as the difference between an agent's own utility base and the coalition utility base it has. A hypothesis was formulated:

Hypothesis: the higher the degree of autonomy, the lower the predictability of behavior will be.

The simulations showed that an increase of the agents' autonomy resulted in a decreased norm-spreading. The norm-internalizing factor did not have such a straightforward relationship as we had hypothesized. We suspect that this is due to the second-order type of learning involved in the internalizing of norms. Further simulation studies (Verhagen & Boman in preparation) will be conducted to clarify this.

Pronouncers

When an intelligent agent has to decide on what action to take, it might ask for advice. The base case is the agent asking itself what to do next. The even more difficult case is when the precarious agent asks someone (or something) else. This case can in turn be analysed by considering two sub-cases. Firstly, the agent may ask other agents in its MAS. This situation
on different grounds, e.g., to get a better bargaining

The agent in need of advice may instead feed a pronouncer with a description of a decision situation, including its subjective assessments of relevant utilities and probabilities. How the pronouncer accesses these assessments is not important for our discussion. The pronouncer also has access to a norm base containing all norms. Each agent coalition has its set of norms, a subset of the norm base, and an agent can belong to many coalitions. Regardless of the coalition structure, the agents turn to the same pronouncer for advice. It is the involvement of norms that makes a pronouncer more than just a decision rule aimed at maximizing an agent’s expected utility. The norm base can be used to disqualify agent actions, plans, or goals if they fail to adhere to the norm set applying to the agent. It can also be used for calculating punishments and rewards, if agent feedback is used.

Naturally, one can imagine a simple MAS in which each agent has the same responsibility towards a group. Then it would suffice to store norms globally, as part of the pronouncer. The realistic and most general case, however, is where each agent has unique obligations towards each and every one of the other agents. For instance, a MAS might consist of 200 agents in which a particular agent has obligations towards the entire population (including itself), but also towards two overlapping strict subsets of, say, 20 and 25 agents that constitute coalitions. These coalitions might be dynamically construed, something which will affect the nature of obligations heavily over time. The control of coalition membership for human agents is different from that for artificial agents, as are the reasons for choosing to join a coalition. In the intelligent building MAS described in (Boman et al. 1998), for instance, a so-called personal comfort agent might be instructed by its human owner to join a coalition of agents representing people working on the same floor, or to join a coalition of agents that have the same temperature and lighting preferences for the conference room. A particular personal comfort agent may then join other coalitions on different grounds, e.g., to get a better bargaining

Conclusions

Acknowledgements

References


