UNCLASSIFIED

Defense Technical Information Center
Compilation Part Notice

ADP012332

TITLE: Autonomy of Decision-Makers in Coalitions

DISTRIBUTION: Approved for public release, distribution unlimited
Availability: Hard copy only.

This paper is part of the following report:
TITLE: KSCO 2002: Second International Conference on Knowledge Systems for Coalition Operations

To order the complete compilation report, use: ADA402533

The component part is provided here to allow users access to individually authored sections of proceedings, annals, symposia, etc. However, the component should be considered within the context of the overall compilation report and not as a stand-alone technical report.

The following component part numbers comprise the compilation report:
ADP012330 thru ADP012354

UNCLASSIFIED
Defining Autonomy

To dynamically form coalitions of decision-makers, the degree of autonomy assumed by each decision-maker must be explicitly agreed upon, beneficial for coalition members and result in productive development of solutions for the goals the coalition is pursuing. Autonomy is a very complex concept. This discussion develops a definition for one dimension of autonomy: decision-making control. The discussion highlights the notion of decision-making control (autonomy) in the context of decision-making groups or coalitions. The development of this definition draws salient features from previous work. Each stage in the development of this definition is highlighted by bold text.

The general concept of agent autonomy is often interpreted as freedom from human intervention, oversight, or control (Beale & Wood, 1994; Etzioni and Weld, 1995; Evans et. al., 1992; Jennings et. al., 1998; Wooldridge and Jennings, 1995). This type of definition corresponds well to the concept of autonomy in domains that involve single-agent-to-human-user interaction. However, in multi-agent systems involving numerous coalitions formed to solve specific goals, a human user may be far removed from the operations of any particular agent. Some researchers have defined autonomy in a more general sense as a property of self-motivation and self-control for the agent (Castelfranchi, 1995; Covrigaru and Lindsay, 1995; Jennings et. al., 1998; Luck and D’Inverno, 1995). This sense of the word autonomy captures the concept of freedom from intervention, oversight, or control by any other agent, including, but not limited to, a human.

Unfortunately, this broad statement fails to account for many characteristics often considered necessary for the realization of autonomous agents. For example, the behavior of autonomous agents is generally viewed as goal-directed (Castelfranchi, 1995; Covrigaru and Lindsay, 1995; Etzioni and Weld, 1995; Luck and D'Inverno, 1995). That is, autonomous agents act with the purpose of achieving their goals. In addition, many researchers consider pro-activeness to be a defining property of autonomous agents (Beale & Wood, 1994; Etzioni and Weld, 1995; Jennings et. al., 1998). Autonomous agents must consider their goals, make decisions about how to achieve those goals, and act on these decisions. Incorporating these properties, autonomy becomes an agent’s active use of its capabilities to pursue its goals without intervention, oversight, or control by any other agent.

No agent can be completely free from all types of intervention with respect to any goal. This discussion distinguishes among three types of intervention as illustrated in the figure and described below:

1. modification of an agent’s environment – other agents modify the environment in which agent $a_0$ operates,
2. influence over an agent’s beliefs – other agents assert facts or, in general, provide information to agent $a_0$ in order to change or influence beliefs held by agent $a_0$, and
3. control over the decision-making process determining which goals, sub-goals, or intentions the agent will pursue – other agents participate to a greater or lesser degree in telling agent $a_0$ how to pursue its higher-level goals.

Extending and modifying the argument presented in (Castelfranchi, 1995), the figure on the right depicts these three ways that other agents (automated or human) may intervene in the operation of agent $a_0$. The solid arrows in the figure represent interventions that primarily affect an agent’s environment, belief base, or goals, respectively. The dotted arrows represent effects of secondary interactions. This discussion suggests that agent designers attempt to classify each agent interaction as one of the three types of intervention based on its primary effect, as pictured in the figure. For example, a task assignment message from agent $a_j$ to agent $a_0$ should be classified as an intervention of type “goal/task determination” because its most salient effect is to change agent $a_0$’s goals. Certainly, such a message would also affect agent $a_0$’s beliefs (agent $a_0$ first believes agent $a_j$ wants agent $a_0$ to perform the new task) and environment (the sending, propagation, and reception of the message imply environmental change). However, these other effects do not capture the nature of the interaction as completely.
Due to the interplay among an agent’s goals, its beliefs, and its environment (pictured in the figure by dotted arrows), it can be difficult to ascribe causality for any particular internal agent modification to a specific intervention occurrence. Establishing this causality becomes especially difficult if the internal agent implementation is unknown. This discussion argues that task assignments creating internal goal changes are useful to model for the purposes of describing autonomy. In any system where agent $a_i$ has authority over agent $a_j$ (e.g. leader of a coalition, military command structure, employer/employee, etc.), agent $a_i$ need not convince agent $a_j$ that some goal needs to be done. Agent $a_i$ simply assigns the goal to $a_j$. Much future work is required to develop classification algorithms for agent interactions, which may ultimately depend on knowledge of the internal design of the particular agents under study. Nevertheless, these suggested categories are useful at this stage to frame discussions of agent autonomy. Because autonomy relates directly to intervention, it is important to be able to identify the nature and impact of these interventions.

This discussion suggests that freedom from intervention of the type “goal/task determination” is the primary dimension of agent autonomy (Barber & Martin, 2000). Goal/task determination is modeled as the process of deciding and assigning which subgoals or subtasks an agent should perform in order to carry out its higher-level goal or inherent purpose. Since any actionable “oversight” or “control” would require such intervention, those terms can be removed from the proposed definition. Therefore, the primary dimension of autonomy is 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. This statement presents autonomy as an absolute value (i.e. either an agent is autonomous or it is not). However, it is more useful to model agents as able to possess different degrees of autonomy, allowing the representation of stronger or weaker intervention.

In addition, it is important to recognize that agents often have multiple goals, some of which may be implicit. This discussion considers an agent’s degree of autonomy on a goal-by-goal basis, rather than attempt to discuss an agent’s overall autonomy as an indivisible top-level concept. This view recognizes that an agent’s autonomy may be different for each goal. For example, some would argue that a thermostat is autonomous and others would argue that it is not. This argument actually hinges on which goal is most important in the assessment of the thermostat’s overall autonomy. It should be quite easy to agree that the thermostat does autonomously carry out the goal to maintain a particular temperature range but that it does not autonomously determine its own set point. Once an agent’s level of autonomy has been specified for each of its goals, the argument can focus (properly) on determining how important each goal is in the assessment of the agent’s overall autonomy. The final proposed definition of autonomy follows: 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.

Agents in a multi-agent system must coordinate to achieve their goals, in general. Establishing an organizational structure (coalition) that specifies how agents in the system should work together helps multi-agent systems achieve effective coordination. Among other things, an organizational structure specifies agent decision-making frameworks. A decision-making framework identifies the locus of decision-making control for a given goal and the authority of decision-makers to assign subtasks in order to achieve that goal. Agents may participate in different decision-making frameworks for each goal they pursue. Agents who implement the capability of Adaptive Decision-Making Frameworks (ADMF) are able to dynamically modify their decision-making frameworks at runtime to best meet the needs of their current situation. Through ADMF, agents are able to reorganize decision-making coalitions by dynamically changing (1) who makes the decisions for a particular goal and (2) who must carry out these decisions. Discussions regarding computational representations of Decision-Making Frameworks (DMFs) can be found in (Barber et. al., 2000) and experiments demonstrating the utility of ADMF are documented in (Barber & Martin, 2001).

References


