

**Situated Autonomy
(Autonomy Panel Position Statement)**

Henry Hexmoor

October 3, 2000

Our interest in autonomy is grounded in the context of an agent, a situation, and a goal. We limit our view of autonomy to an agent's moment-to-moment action selection instead of a long-lived agent characteristic. Autonomy of an agent maps an agent, a situation, and a goal to a stance towards the goal such that the stance will be used to generate the most appropriate or the most relevant action for the agent [Hexmoor 2000]. At a coarse level the agent's stance towards the goal will be whether to abandon it or to decide its overall position toward the goal: to make it an entirely personal goal, to make a goal for another agent, to collaborate with other agents with some level of responsibility, or to have some responsibility about it that is less than total responsibility. Responsibility for a goal is the amount of effort an agent is willing to spend on seeing to its accomplishment. At a finer level the agent's stance will go beyond an overall position to include a degree of autonomy.

Time affects the autonomy of an agent. Consider an agent that must have fast action selection decisions. This agent does not have time for deliberation and will use instinctual means to produce its stance. An agent with more leisurely action selection requirements can use reasoning and perhaps examines its Beliefs, Desires, and Intentions before producing a stance. An agent's stance may have a temporal dimension that includes external cues. External cues might be observation of a second agent's actions or autonomies it exhibits. For instance, an agent's stance to pay for dinner for two might be "If the dinner bill is less than \$20, I will offer to pay unless the other guy offers to pay".

Teamwork provides context for an agent's autonomy. Once an agent is motivated to form or to participate in a team toward a common goal, its stance will be to share the goal. Such a stance will require the agent to dynamically monitor and adjust its stance relative to teammates. Situation in autonomy consideration must not only include the state of the world as it pertains to the agent's goal but also the agent's model of teammate situations and autonomies.

Next, I make brief remarks about concepts that are discussed alongside autonomy and point out the differences and misunderstandings.

Control and autonomy are often conflated. A system with ability to operate at several different control regimes is said to have adjustable autonomy. Whereas, autonomy produces a stance, control is a skill. An agent who ponders autonomy with respect to a goal has not fully committed to the goal. An agent who is concerned with control of a goal, has largely finished its consideration of autonomy. Variable or changing control is desirable but that is not the same as the ability for adjustable autonomy. An agent's capacity or power for choice and commitment to a control level is autonomy. This view of "capacity or power for choice and commitment" can be taken too far as in the idea of *free will*, which requires an independence from natural, social, or divine restraints. From the *free will* perspective, a program or a robot would never have autonomy. However, we consider "capacity or power for choice and commitment" as a relational notion and in the context of a goal and a situation.

Experiencing independence or self-sufficiency seems to be proportionally related to the experience of autonomy. Autonomy is proactive in that it is an agent stance whereas dependence or reliance is a perception. "I feel independent, so I choose to be autonomous." However, if I decide on being autonomous, it does not make me feel independent. Castelfranchi points out that Social Dependence is a sub-set of Non-Autonomy [Castelfranchi 2000]. We agree with his position that autonomy is related to the computational and representational powers of an agent's architecture.

Unlike the purported paradox that autonomy is lowered with cooperation [Castelfranchi 1995], we believe that an agent's cooperative attitude is not at odds with its autonomy. Autonomy and cooperative attitude are independent. Consider the example of race-drivers in a "draft" who mutually benefit from the aerodynamics

ATAL-2000 proceedings

of the drafting. In this example, drivers keep their autonomy fairly high and their cooperation is not with any specific driver but a driver that is at the right position for “drafting”.

References

C. Castelfranchi (1995). Guarantees for Autonomy in a Cognitive Agent Architecture, In Agent Theories, Languages, and Languages (**ATAL**), p. 56-70, (LNAI volume 890), Springer-Verlag..

C. Castelfranchi (2000). Founding Agent's 'Autonomy' On Dependence Theory, European Conference in Artificial Intelligence, ECAI 2000, Berlin.

H. Hexmoor, (2000). A Cognitive Model of Situated Autonomy, In Proceedings of PRICAI-2000 Workshop on Teams with Adjustable Autonomy, Australia.