In 2000 Proceedings of PRICAI-2000 Workshop on Teams with Adjustable Autonomy, Australia.

A Cognitive Model of Situated Autonomy

Henry Hexmoor

Computer Science & Computer Engineering Department
Engineering Hall, Room 313
Fayetteville, AR 72701
hexmoor@uark.edu

Abstract. We introduce situated autonomy and present it as part of the process of action selection. We then discuss the cognitive ingredients of situated autonomy and derive a degree of situated autonomy.

1 Situated Autonomy and Action Selection

Autonomous agents have been defined to be agents with self-generated goals using the agent's motivations [4]. Such agents perform action selection, which is the function of selecting the most relevant and meaningful action [9], entirely for selfish reasons [7]. We believe circumstances of everyday agents provide opportunities for reasoning about relative levels of autonomy. Instead of being autonomous in the general sense, we will focus on the notion of autonomy in the context of a situation and in a team with other agents. We will consider agents able to perform autonomy considerations very fast while they are in the situation. Imagine in a game of basketball, the agent who is in the midst of running anticipates a block and reflects about whether to pass the ball or to run with it. Here, autonomy is a split-second situated assessment.[1] The player considers choosing each of the behaviors "pass the ball" and "run with the ball." The agent's considerations of autonomy involve the higher-level goals of scoring or driving the ball into the opponent zone. The agent decides between its orientation to "pass the ball" which means sharing its autonomy toward scoring/driving with another player or its orientation to "run with the ball" which means self-autonomy. Situatedness is to consider situations in the environment as integral component of the agent's process of deliberation or reactive response generation. Situation is a state of the world as it pertains to a problem. **We define situated autonomy as an agent's stance, as well as the cognitive function of forming the stance, toward assignment of the performer of a goal at a particular moment when facing a particular situation.** Assumption of individual versus social rationality affects the cognitive function. At a coarse level the agent's orientation toward 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 consider the goal a collaborative effort, or to consider an inclination for the goal that is less than totally self-directed. Here, we are not concerned about responsibility for a goal, which is the amount of effort or commitment 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 situated autonomy. In this paper, the degree of

[1] Assessment of autonomy is either a deliberative process or an automatic association of a stance that might be a historic stance or based on the agent's personality.

autonomy will be a measure of the agent's deliberateness over its autonomy decision. Two useful measures of autonomy beyond the scope of this paper are (1) degree of relative dependence on environmental factors such as other agents, and (2) degree of control (or influence) an agent has over a goal. Generally determining degree of autonomy is more time-consuming than determining an overall position. In our discussion of situated autonomy we will not care whether the goals are internally generated or externally imposed.

Action selection generates an action in response to a new situation. An important step in action selection is choosing among possible plans or possible primitive actions. We propose that situated autonomy can be used in this decision. Given that an agent may have several alternative plans and actions to achieve a goal with each alternative appropriate at different commitment level, an agent's situated autonomy can be used as the context for arbitration.

For a cognitive agent, action selection is affected by the frequency of encountering new situations. We describe an agent's assessment of its situated autonomy that is also affected at varying situation frequencies.

At the highest frequency, the agent may produce reflex-like actions. Such agents have no time to account for their situated autonomy. At a relatively fast frequency, the agent may produce reactive actions with minimal time to account for situated autonomy. Such situated autonomy assessment will consider pre-disposition toward the goal. Pre-disposition here is taken as "an evaluative tendency that is expressed by evaluating a particular entity with some degree of favor or disfavor" [3, p. 693]. An agent's pre-disposition toward a goal is based on semi-permanent beliefs and goals about enabling factors for the goal. Our understanding of pre-disposition is a cognitive function that operates on the agent's weak beliefs and unmotivated goals. **Enabling factors for a goal are subset of the elements of the situation that are either necessary or facilitating conditions for attempting the goal.** We consider individual, social, and physical enablers with different origins: (a) entirely endogenous, (b) exogenous and social, and (c) exogenous and physical in nature. The individual enablers are the agent's competencies. The social enablers are the social influences and motivators. The physical enablers are the physical conditions, physical resources, and physical tools. We will further discuss these types of enablers in a later part of this paper. An agent may have a model of other agents as well as its own. We will use prefixes "Other-" and "Self-" to distinguish between an agent's model of other agent's enablers and its own. For instance, Other-Social-Enabler will denote an agent's model of another agent's social enablers and Self-Physical-Enabler will denote an agent's model of its own physical enablers.

At slower frequencies, the agent will have more time to assess the quality and quantity of the enabling factors. Situated autonomy at that level will be based on dispositions toward the goal. Perkins, Jay and Tishman [8], define dispositions as "people's tendencies to put their capabilities into action" (p. 75). At yet slower frequencies, the agent will have time to consider going beyond dispositions derived from enabling factors and include motivations. Human motivations are a collection of

psychogenic needs, which guides behavior [6]. At the slowest frequency, the agent may consider long-term ramifications of its options in order to produce optimal actions. In this paper we consider goal-oriented social agents in domains with relatively fast situation frequency. Such agents may have limited time to consider situated autonomy. They may just have enough time for assessing an overall position.

Consider action selection as a linear process where somehow the agent's action selection has settled on a goal. The next step and the focus of this paper are the agent's reflections on its autonomy with respect to the goal at hand. Finally, the agent uses the results of its introspection and renders a decision about action(s).

Figure 1 summarizes the types of action generated by action selection that is at different frequencies of Situations. Reflex actions are generated without much deliberation for situated autonomy. Other than reflex actions, situated autonomy consideration for actions generated to the right are coarser than for action to the left.

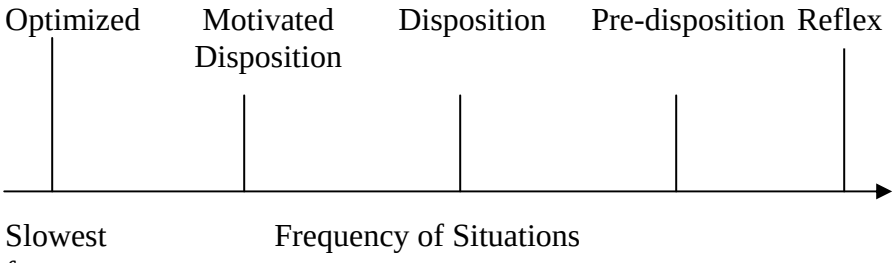

Figure 1 Action selection at different frequencies of Situations

Given a goal, the agent's assessment of situated autonomy is a cognitive process that is comprised of several stages, Figure 2. The process begins by the agent's consideration of pre-dispositions only. If the agent has a habit of disfavoring the goal, it will decide to ignore it for no other reason other than its habit and considers itself Not-autonomous with respect to that goal. The agent who habitually favors the goal and favors itself to be the executor of the goal will consider itself to be Self-autonomous. The agent who habitually favors the goal and favors itself not to be the executor of the goal will lean toward delegation and will consider itself to be Del-autonomous.

The agent with some more time may consider the goal further and form dispositions toward it. If the agent perceives the goal to be impossible, the agent forms a mental state of Not-autonomous. If the agent perceives that the goal is doable either exclusively by the agent alone or by delegation, it will stop further considerations of situated autonomy. If such an agent solely using its dispositions considers itself to be the executor of the goal, it will consider itself to be Self-autonomous. When we say an agent is autonomous with respect to a goal, we may mean one of two things about its disposition toward the goal. We may mean the agent is self-reliant in the sense that it is not affected by any exogenous sources such as social or physical. Alternatively, we may mean the agent can bring about the desired effect given its access to its

exogenous sources such as other agents or resources or it can do it itself. If it considers other agents to be executors of the goal, it will consider itself to be Del-autonomous. If the goal is deemed clearly appropriate for delegation due to the agent's inability to perform the goal itself, the agent is considered Del-autonomous and subsequently a communicative act will be generated.

An agent who has formed a disposition toward its goal that has not resulted in determination of either the agent or others being the exclusive executors may further use its motivations. Moreover, motivations can modify a decision that is previously determined based on disposition. We will consider motives to be captured by a policy that produces a preference to favor/disfavor the goal as well as the executor of the goal. If a goal is deemed inappropriate due to the agent's motivation policy, the initial commitment is revised and the agent is considered to be Not Autonomous with respect to that goal. If a goal is deemed feasible for self-performance due to the agent's disposition and additionally appropriate due to the agent's motivation, the agent is considered to Self-autonomous and the goal might be sent to the motoric system.

If the agent has not determined exclusive execution, the agent is slated to perform the goal with other agents and its autonomy is classed as Semi-autonomous or Shared-autonomous. Shared-autonomous implies getting assistance from another agent or reliance on some environmental elements such as tools or resources, or offering help to some other agent who will be the primary executioner of the goal. Shared autonomy implies more demand for the agent than semi-autonomy. With semi-autonomy, the agent knows it is dependent on outside sources to perform the goal. With shared autonomy the agent knows furthermore that there are one or more agents that complement its autonomy. An example of shared autonomy is shown between a human air traffic controller and a collision-detection agent [5].

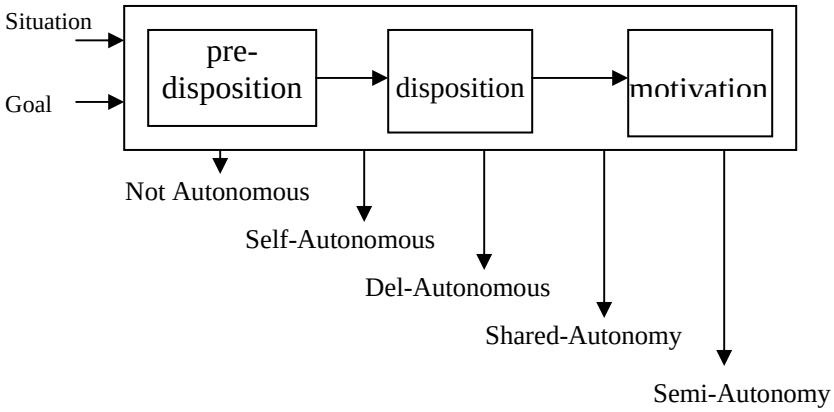

Figure 2 Situated Autonomy as part of the process of action selection

An agent may perceive the goal to be shared by a team. Such an agent will be motivated to help the team and may override its determination of shared-autonomy or semi-autonomy based on its dispositions in favor of self-autonomy.

2 A BDI Model of Autonomy: Necessary Cognitive Components for Situated Autonomy

In this section we will provide a predicate calculus account of autonomy. We do not claim our characterization to be complete but we believe it covers the major human-intuitive aspects of situated autonomy.

We refine our earlier use of pre-disposition with a combination of a weak belief in the enabling factors, and an unmotivated goal. An agent's belief is weak B^w when it is not carefully formed and it is premature. An agent may have little time to form its belief or the agent senses uncertainties in the object of the belief. An agent may form sketchy beliefs solely based on prior experience and otherwise not have substantiated support for holding the belief. In contrast to B^w, we define B^s as a strong belief when the agent has inductive or deductive reasons for holding the belief. This style of treating belief differs from explicit/declarative evaluation. Instead they are closer to that in [1].

A goal is unmotivated G^u when the process of adopting the goal has suffered from weaknesses in related prior intentions or beliefs. G^m is defined as a motivated goal when the agent has (a) inductive or deductive reasons for adopting the goal including goals that the agent shares with a team [2], and (b) a wish for expecting it. Otherwise, an agent's goal is unmotivated. An unmotivated goal is adopted out of habit but fails to have justification and the agent's wish.

We list a few factors that contribute to forming a motivated self-directed goal. We will not use any notation for brevity. In the following, a goal is owned by the agent if another agent did not suggest the goal. I.e., the agent has endogenous reasons for the goal.

- x *owns the goal* to bring about that goal.
- x is *not the owner of the goal* but the goal is owned by the team and x perceives a great deal of personal *commitment and responsibility* for that goal.
- x perceives itself to be the *only agent who can do the goal*.

One or more of the following mental states may support motivated goal that is other - directed:

- x *owns the goal* to bring about that goal
- x is *not the owner of the goal* but is owned by the team and x perceives a great deal of *commitment and responsibility* for the delegation of that goal.
- x *believes* it does not have individual enablers.

We are now ready to define Self-autonomy in terms of pre-disposition.

Definition 2.1 Self-autonomous^p

Agent x is self-autonomous based on its pre-dispositions with respect to goal g in situation s iff x can perform an act to bring about g with permission but however, it has a weak belief in situation s about Self-Enablers and it has g as an unmotivated goal in situation s.
(Self-autonomous^{p} x s g) $\equiv \exists \alpha$ (Agts α x) $\wedge$ (Achvs α g) $\wedge$ (Per x α) $\wedge$ (B^{w} x s Self-Enablers) $\wedge$ (G^{u} x s g)

The predicate "Per" stands for the deontic notion of permission for individuals (See [10] for typical usage of permissions) and "G-Per" (used later in this paper) stands for group permission. Note that individuals and groups typically do have the same permissions and that group permissions cannot be reduced to the permissions of the individuals forming the group. The notations "Agts" and "Achvs" stands for "agent of" and " Achieves" respectively and are similar to their usage in [11].

We consider disposition to be a combination of an unmotivated goal G^{u} and a strong belief B^{s} in the enabling factors. Using disposition, we define Self-autonomy again, Self-autonomous$^{D.}$

Definition 2.2 Self-autonomous^{D}
Agent x is self-autonomous based on its dispositions with respect to goal g in situation s iff x can perform an act to bring about g with permission, it has a strong belief in situation s about Self-Enablers but however, it has g as an unmotivated goal in situation s.
(Self-autonomous^{D} x s g) $\equiv \exists \alpha$ (Agts α x) $\wedge$ (Achvs α g) $\wedge$ (Per x α) $\wedge$ (B^{S} x s Self-Enablers) $\wedge$ (G^{u} x s g)

We define motivated disposition as a combination of a goal that is motivated G^{m} with motivation, and a strong belief B^{s} in the enabling factors. We define Self-autonomy this time based on motivated disposition, Self-autonomous$^{MD.}$

Definition 2.3 Self-autonomous^{MD}
Agent x is self-autonomous based on its motivated dispositions with respect to goal g in situation s iff x can perform an act to bring about g with permission, it has a strong belief in situation s about Self-Enablers, and it has g as a motivated goal in situation s.
(Self-autonomous^{MD} x s g) $\equiv \exists \alpha$ (Agts α x) $\wedge$ (Achvs α g) $\wedge$ (Per x α) $\wedge$ (B^{S} x s Self-Enablers) $\wedge$ (G^{m} x s g)

Next, we define Del-autonomy in terms of pre-disposition.

Definition 2.4 Del-autonomous^{p}
Agent x is del-autonomous based on its pre-dispositions with respect to goal g in situation s iff there is an agent y (other than x) that can perform an act to bring about g with permission but however, x has a weak belief in situation s about Other-Enablers and it has g as an unmotivated goal in situation s.
(Del-autonomous^{p} x s g) $\equiv \exists \alpha y$ (Agts α y) $\wedge$ (Achvs α g) $\wedge$ (Per y α) $\wedge$ (B^{w} x s Other-Enablers) $\wedge$ (G^{u} x s g)

Definition 2.5 Del-autonomous^{D}
Agent x is del-autonomous based on its dispositions with respect to goal g in situation s iff there is an agent y (other than x) that can perform an act to bring about g with permission, x has a strong belief in situation s about Other-Enablers but however, it has g as an unmotivated goal in situation s.
(Del-autonomous^{D} x s g) $\equiv \exists \alpha y$ (Agts α y) $\wedge$ (Achvs α g) $\wedge$ (Per y α) $\wedge$ (B^{S} x s Other-Enablers) $\wedge$ (G^{u} x s g)

Definition 2.6 Del-autonomous^{MD}
Agent x is del-autonomous based on its motivated dispositions with respect to goal g in situation s iff there is an agent y (other than x) that can perform an act to bring about g with permission, x has a strong belief in situation s about Other-Enablers, and it has g as a motivated goal in situation s.
(Del-autonomous^{MD} x s g) $\equiv \exists \alpha y$ (Agts α y) $\wedge$ (Achvs α g) $\wedge$ (Per y α) $\wedge$ (B^{S} x s Other-Enablers) $\wedge$ (G^{m} x s g)

We believe Shared_autonomy requires more than a pre-disposition.

Definition 2.7 Shared-autonomous^{D}
Agent x is shared-autonomous based on its dispositions with respect to goal g in situation s iff x is a part of a group of agents t, where t can perform an act to bring about g with permission, x has a strong belief in situation s about Group-Enablers but however, the group has g as an unmotivated goal in situation s.
(Shared-autonomous^{D} x s g) $\equiv \exists \alpha t$ (x $\in$ t) $\wedge$ (Agts α t) $\wedge$ (Achvs α g) $\wedge$ (G-Per t α) $\wedge$ (B^{S} x s Group-Enablers) $\wedge$ (G^{u} t s g)

Definition 2.8 Shared-autonomous^{MD}
Agent x is shared-autonomous based on its motivated dispositions with respect to goal g in situation s iff x is a part of a group of agents t, where t can perform an act to bring about g with permission, x has a strong belief in situation s about Group-Enablers, and the group has g as a motivated goal in situation s.
(Shared-autonomous^{D} x s g) $\equiv \exists \alpha t$ (x $\in$ t) $\wedge$ (Agts α t) $\wedge$ (Achvs α g) $\wedge$ (G-Per t α) $\wedge$ (B^{S} x s Group-Enablers) $\wedge$ (G^{u} t s g)

We believe Semi_autonomy requires more than a pre-disposition but typically does not change with motivation.

Definition 2.8 Semi-autonomous^{D}
Agent x is shared-autonomous based on its dispositions with respect to goal g in situation s iff x is a part of a group of agents t, where t can perform an act to bring about g with permission, However, x has a weak belief in situation s about the Group-Enablers, and the group has g as an unmotivated goal in situation s.
(Semi-autonomous^{D} x s g) $\equiv \exists \alpha t$ (x $\in$ t) $\wedge$ (Agts α t) $\wedge$ (Achvs α g) $\wedge$ (G-Per t α) $\wedge$ (B^{w} x s Group-Enablers) $\wedge$ (G^{u} t s g)

Finally, an agent is not autonomous if no other form of Autonomy holds.

Definition 2.8 Not-autonomous

Agent x is not-autonomous with respect to goal g in situation s iff there are no acts that x has both permission to perform and perform to achieve g.

$(\text{Semi-autonomous}^{D}\ x\ s\ g) \equiv \neg\exists\ \alpha\ (\text{Agts}\ \alpha\ t) \wedge (\text{Achvs}\ \alpha\ g) \wedge (\text{Per}\ x\ \alpha)$

Not-autonomous necessarily presupposes that x does not have any Self-, Del-, Shared-, or Semi-autonomy.

Our notations so far can only help with a coarse reasoning about situated autonomy for the agent. We have identified four categories of situated autonomy. Self-Autonomous gives the agent the most choices to complete its action selection. An agent's action selection must decide on method of delegation with Del-Autonomous. Semi-autonomous is the least clear and the agent's action selection may use other consideration for action selection. With Shared-autonomous, the action selection must consider the other agents sharing the autonomy over the goal for an action. If the agent is Not_autonomous, its action selection can terminate and the agent will not rationally perform that goal.

3 Degree of Situated Autonomy

In the previous section we presented several categories of situated autonomy. Stances based on pre-disposition are weaker than the ones based on disposition. Stances based on disposition are weaker than the ones that include motivation. We propose that the strength of an agent's stance is the basis for its degree. For example, an agent's stance Del-autonomous^{MD} is stronger and has a higher degree than the agent's stance Self-autonomous^{D}. The degree of situated autonomy within each stance is a function of (a) the agent's subjective certainty in the agent's beliefs about enabler ingredients, (b) the perception of the absolute quantity of the enabler components that gives the agent a sense of liberty, and (c) the strength of positive mental states due to situation at-hand and lack of strength in the negative mental states due to the situation at-hand. First, we will briefly discuss the liberties an agent senses with regard to three main enabling quantities. This will be used in defining degrees of Self-, and Del-Autonomy.

The physical enablers. For example, the road condition and mechanical integrity of a vehicle can be considered environmental elements for a driving goal. An agent's perception of the quantitative condition of the road and the car combined with its certainty about its own beliefs forms the overall physical enabler component of its autonomy. Colloquially, we may think of the agent's Faith or Trust in the physical enablers or Faith/Trust in its perception or its beliefs. To the extent the agent experiences freedom from concerns about physical enablers, the agent has physical liberties.

The social enablers. Other agents may provide positive or negative social influences. Other agents may facilitate achievement of the goal or detract from it. The perception of the quantity of social elements affecting the goal as well as the agent's belief in

such elements makes up the contribution of the social enablers in determining a degree of situated autonomy. For example, a friend who is familiar with the driving directions for a driving goal can make the goal easier. The agent's perception of the degree of its dependence on such a friend for help contributes to the agent's degree of social enablers. To the extent the agent is invested in the favorable social elements and is ready to guard against unfavorable elements, the agent has social liberties.

The individual enablers. The agent's perception of the degree of its competencies as well as the certainty of beliefs in them makes up the contribution of individual enablers in determining a degree of situated autonomy. To the extent the agent is self-confident about the goal the agent has individual liberties.

The necessary conditions for Self-Autonomy require an agent to sense high levels of Individual, Social, and Physical liberties with respect to the goal. Whether such an agent has a high degree of Self-autonomy depends on the strength of its mental states due to the situation at-hand.

The necessary condition for Del-Autonomy requires an agent to sense a high level of Social liberty of other agents with respect to the goal. Whether such an agent has a high degree of Del-autonomy further depends on (a) the strength in one or more of the following positive mental states:

- x *owns the goal* to bring about the goal

and (b) lack of strength in the following negative mental state:

- x *believes* it does not have individual enablers.

We treat the degree of an agent's Shared-Autonomy to be its relative autonomy with respect to other agents for a given goal in a given situation. Since other agent's autonomy and the situation affect Shared-Autonomy, it is more dynamic than Self-, and Del-Autonomy. The necessary condition for Shared-Autonomy requires an agent to sense that there are other agents who share the goal and complement its autonomy with respect to the goal. Whether such an agent has a high degree of Shared-autonomy depends on (a) the strength in the following positive mental state:

- x *owns the goal* to bring about the goal.

and (b) lack of strength in one or more of the following negative mental states:

- x *believes* it does not have individual enablers,

and (c) the extent to which,

- relative to other agents with whom x shares the goal, x perceives a relatively large level of personal power and control over the goal.

For an agent to sense a high degree when it is Semi-Autonomous depends on the intensity of the agent's mental state that "x *owns the goal* to bring about that goal" combined with the agent's mental state that "x *believes* it does not have adequate individual enablers."

Finally, an agent that is Not Autonomous has the least degree of situated autonomy.

4 Conclusion

We have discussed the idea that situated autonomy is at the core of an agent's action selection. It is responsible for the agent's mental state about how to relate to a goal in the context of relevant enabling factors in an environment that includes other agents. We have presented central cognitive ingredients that constitute notions of autonomy of self, delegation, and sharing. Degree of autonomy as a measure of the agent's deliberativeness of its decision is then presented. It is argued to be dependent on the qualities of enabling factors and the strength of the agent's beliefs in them.

We plan to refine these notions and to implement agents that exhibit dynamic changes in their autonomy. Previously we presented some quantified results of different levels of autonomy [5]. Our future plans include extending our empirical studies with implementations of situated autonomy.

References

1. Castelfranchi C., Falcone R., (2000), Trust and Control: A Dialectic Link, **Applied Artificial Intelligence journal,** Special Issue on "Trust in Agents" Part1, Castelfranchi C., Falcone R., Firozabadi B., Tan Y. (Editors), Taylor and Francis 14 (8)

2. Cohen, P. R., and Levesque, H. J. (1990). Persistence, intention, and commitment, Cambridge, MA: MIT Press,1990, Cambridge, MA: MIT Press,.

3. Eagly, A.H. (1992). Uneven Progress: Social Psychology and the Study of Attitudes. Journal of Personality and Social Psychology 63(5): 693-710.

4. d'Inverno M. and Luck M., (1996). Understanding Autonomous Interaction, in ECAI '96: Proceedings of the 12th European Conference on Artificial Intelligence, W. Wahlster (ed.), 529-533, John Wiley and Sons.

5. Hexmoor, H.. (2000). Case Studies of Autonomy, In proceedings of **FLAIRS 2000**, J. Etherege and B. Manaris (eds), p. 246- 249, Orlando, FL.

6. McAdams. D.P. (1989). The Development of a Narrative Identity. In D.M. Buss and N. Cantor (Eds.). Personality Psychology: Recent Trends and Emerging Directions. New York: Springer-Verlag. pp. 160-176.

7. Luck M. and d'Inverno M., (1995). A Formal Framework for Agency and Autonomy, in Proceedings of the First International Conference on Multi-Agent Systems, 254-260, AAAI Press / MIT Press,

8. Perkins, D., Jay, E., and Tishman, S. (1993). New Conceptions of Thinking: From Ontology to Education. Educational Psychologist 28(1): 67-85.

In 2000 Proceedings of PRICAI-2000 Workshop on Teams with Adjustable Autonomy, Australia.

9. Pirjanian, P., (2000). Multiple Objective Behavior-Based Control, **Robotics and Autonomous Systems**, Volume 31, Issue 1-2, 30 April 2000, pp. 53-60.

10. Wieringa R. J. and Ch. Meyer, J.-J. (1993). Applications of Deontic Logic in Computer Science: A concise view, In **Deontic Logic in Computer Science**, 17 – 40, John Wiley & Sons, Chichester, England.

11. Wooldridge, M. (2000). **Reasoning about Rational Agents**, The MIT Press, Boston.