

Deriving Goals and Intentions

Graça Gaspar

Faculdade de Ciências de Lisboa
Bloco C5 Piso 1, Campo Grande
1700 Lisboa, Portugal
gg@di.fc.ul.pt

Helder Coelho

INESC
9, Rua Alves Redol, 7^o
1000 Lisboa, Portugal
hcoelho@inesc.pt

Extended Abstract

Defining intentions, as well as relating intentions to other components of the mental state of an agent, has been a topic of active research. Nevertheless, when one looks for foundations on which to base implementations of societies of agents, all the models of intention that have been presented so far have serious drawbacks: they adopt idealistic assumptions, like for instance, that agents believe all logic consequences of their beliefs; they do not give a clue on where do goals or intentions come from; and, they do not address the problem of revising intentions or goals during a communication process.

In this paper, we defend a model of goals and intentions, based on deductive structures, and an integrated framework for revising them, that can and has been adopted as a specification for designing and building up artificial agents.

The goals of an agent are chosen from among a set of formulas that represent desires of the agent, or goal hypotheses, that may be contradictory. The agent will choose a subset of preferred non-contradictory desires, that must also be non-contradictory with its beliefs. The goal model makes explicit the preferences among desires as well as the special inference rules the agent uses to derive goals, besides the rules it uses for beliefs. A set of goal formulas is said to be contradictory if the agent is able to derive false from that set.

Agents also manipulate belief formulas about the possibility of decomposing a goal into a set of subgoals. Such formulas may be included initially in their set of belief hypotheses or agents may come to believe them as a result of applying inference rules to beliefs about actions.

The considerations that led us to our definition of intention were based on the arguments of Bratman over the functionality of intentions and their relation with the goals and beliefs of the agent, but were also based on the assumption that agents have limited reasoning capabilities

and should therefore avoid planning in detail before committing themselves to a goal.

Intentions are basic goals and, therefore, non-contradictory (not necessarily consistent with the agent's beliefs, but only believed to be compatible with them). An agent must also be able to establish ways to satisfy intentions. We impose a restriction on intentions that, on one side, forces the agent to sketch a plan to satisfy an intention, and on the other side, does not force it to detail that plan nor commits it to follow that plan in the future under all conditions. With that aim, we define satisfiable goals and restrict intentions to being satisfiable goals.

The goals and intentions of an agent may change as a consequence of the execution of plans and actions, as well as of communication with other agents. Such a change is specified in our model by a function called assimilation that defines the revision of goals of the agent, in terms of changes in the corresponding deductive structures, as a result of the reception of a message or execution of an action. This function is not defined directly in an explicit form, in an agent model, but it is indirectly defined through the specification of general revision principles, that all the agents in a society must conform to, and through the indication of an ordered list of heuristic criteria, that may be different for different agents and that determine the attitude of an agent toward the others. The list of adopted heuristic criteria can be considered as defining the agent's "personality".

The development of our model has guided the parallel development of a workbench, and experimentation with earlier versions of the workbench has in turn provided guidance for the refinement and extension of the conceptual model. We argue that research for theoretical foundations of multiagent systems must open the way for such walks, forth towards implementations and back, in order to lead up to realistic models..