

Rational Agents : Prioritized Goals, Goal Dynamics, and Agent Programming Languages with Declarative Goals

(Extended Abstract)

Shakil M. Khan

Department of Computer Science and Engineering
York University, Toronto, ON, Canada
skhan@cse.yorku.ca

Categories and Subject Descriptors

I.2.11 [Artificial Intelligence]: Distributed Artificial Intelligence—*Intelligent agents, Multiagent systems*; I.2.4 [Artificial Intelligence]: Knowledge Representation Formalisms and Methods—*Modal Logic*

General Terms

Theory, Languages

Keywords

Prioritized goals, goal change, subgoals, intention, logic of agency, agent programming languages with declarative goals

1. INTRODUCTION

Central to the agent concept is the notion of goals. In the past two decades, much research has been devoted to formalizing various motivational attitudes of agents, such as goals and intentions. Nevertheless, the existing formal models of goals and intentions suffer from a host of problems. For instance, most of them assume that all goals are equally important, and many only deal with achievement goals. Moreover, they do not guarantee that an agent's goals will properly evolve when an action/event occurs, when the agent's beliefs/knowledge changes, or when a goal is adopted or dropped. Also, most of these frameworks do not model the dependencies between goals and the subgoals and plans adopted to achieve these goals; subgoals adopted to bring about a goal should be dropped when the parent goal becomes impossible, is achieved, or is dropped. Dealing with these issues is important for developing effective models of rational agency and BDI agent programming languages.

There has also been much work on agent programming languages with declarative goals where the dynamics of goals and intentions and the dependencies between goals and subgoals are modeled (e.g. see the references in [9]). It has been argued that these declarative goals play an essential role for monitoring goal achievement and performing recovery when plans fail. Unfortunately, most of these accounts do not provide a formal semantics for declarative goals, goal dynamics, and subgoals, and do not handle goals with different priorities. Generally, there are no requirements for intended plans to be consistent with intended declarative goals. Also, most

Cite as: Rational Agents : Prioritized Goals, Goal Dynamics, and Agent Programming Languages with Declarative Goals (Extended Abstract), Shakil M. Khan, *Proc. of 9th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, van der Hoek, Kaminka, Lespérance, Luck and Sen (eds.), May, 10–14, 2010, Toronto, Canada, pp. XXX-XXX.

Copyright © 2010, International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

of these frameworks do not deal with temporally extended goals, and as a result they often need to accommodate inconsistent goal-bases to allow the agent to achieve conflicting goals at different time points.

In my dissertation, I propose a first order (with some higher order features) logical account of prioritized goals and their dynamics in the situation calculus. My model of prioritized goals supports the specification of general temporally extended goals, not just achievement goals, and handles subgoals and their dynamics. As an application of my theory, I also develop a Simple Rational Agent Programming Language (SR-APL, henceforth) with declarative goals that is based on a version of this rich theory, combining elements from BDI agent programming languages and from the situation calculus based ConGolog agent programming language [2]. While doing this, I address some of the aforementioned problems of current agent programming languages. In particular, SR-APL supports prioritized goals and is grounded on a formal theory of goal change. It ensures that the agent's declarative goals and procedural plans are consistent with each other. I then show that agents programmed in SR-APL satisfy some key rationality requirements.

2. RESULTS ALREADY OBTAINED

In [5, 7], I present a formalization of a rational agent that always tries to optimize her prioritized goals. To support modeling temporally extended goals, I introduce a new sort of *paths* in the situation calculus, and propose an axiomatization of infinite paths. By incorporating infinite paths into my framework, I can evaluate goals over these and handle arbitrary temporally extended goals. Thus, my framework is more expressive than those where goal formulae are evaluated w.r.t. finite paths, and unlike them I can handle for example unbounded maintenance goals.

In my agent theory, an agent can have multiple goals/desires at different priority levels, possibly inconsistent with each other. I assume that goals are totally ordered w.r.t. the priority ordering. I define intentions/chosen goals, i.e. the goals that the agent is actively pursuing, as the maximal set of highest priority goals that is consistent with each-other and with the agent's knowledge. My model of intentions supports the specification of general temporally extended goals. I also specify how these goals evolve when actions/events occur, or when the agent adopts or drops a goal. My formalization of prioritized goal dynamics ensures that the agent always tries to maximize her "utility". She will abandon a chosen goal ϕ if an opportunity to commit to a higher priority but inconsistent with ϕ goal arises (cf. Section 5 of [5] for a concrete example). As such my model displays an idealized form of rationality. This is in contrast to Bratman's [1] model of practical rationality that takes

into consideration the resource-boundedness of real world agents. According to Bratman, intentions limit the agent's reasoning as they serve as a filter for adopting new intentions. However, the agent is allowed to override this filter if necessary, e.g. when adopting ϕ increases her utility considerably. My framework in [5] can be viewed as a theory of intention where the filter override mechanism is always triggered. Note that, this is costly and requires the agent to constantly deliberate about her intentions, and thus may not be suitable as a foundation for an agent programming language.

I show that my agent theory has many desirable properties such as the consistency of chosen goals, that adopt and drop have the expected effects, etc. I also prove some properties that specify the conditions under which an agent's achievement prioritized goals and achievement chosen goals persist.

In [6], I propose an account of subgoals and their dynamics in an early version of the account discussed in [5]. In that paper, I give a definition of subgoals and discuss how subgoals change when an agent's knowledge changes as a result of the execution of an action, or when she adopts a subgoal with respect to a parent goal. I prove some properties to show that my formalization of subgoal dynamics ensures that a subgoal is dropped when its parent goal becomes impossible or is dropped, but not necessarily vice-versa. However, in that paper I do not handle early achievement of a parent goal, in which case too, the subgoal should be dropped.

As mentioned above, my logic of prioritized goals for optimizing agents [5] is not quite suitable for resource-bounded agents. In [4], I develop a modified version where agents are more committed to their chosen goals. This is achieved by eliminating the filter override mechanism altogether. I think that this framework is more suitable as a background theory for an agent programming language, since it accounts for the resource-boundedness of the agent by limiting her deliberation.

3. WORK THAT REMAINS

Building on this rich theory of prioritized goals and subgoals, I am currently working on a specification for SR-APL (see [4] for a preliminary version). I am focusing on developing an expressive and robust agent programming logic that captures rationality without worrying about tractability. In [4], I discuss how the proposed APL compares to existing APLs with declarative goals. In particular, I show that when effects of actions are not reversible or when time specific goals are considered, agents specified in other APLs may behave irrationally in the sense that they can adopt and execute plans that makes some of their other goals/plans impossible to achieve/execute. I then show that in the absence of external interferences, an agent specified in SR-APL behaves rationally in the sense that the actions performed and the subgoals/plans adopted by the agent are consistent with her (declarative and procedural) intentions. I also prove that the consistency between her adopted declarative goals and procedural plans are maintained.

I also plan to prove additional properties of the frameworks in [5, 4], such as introspection of goals. Finally, in light of the above discussion, it would be nice to modify my proposed subgoal dynamics to handle early achievement of goals, i.e. automatically drop subgoals whose parent goal have been achieved.

4. DISCUSSION AND FUTURE WORK

In recent years, there has been a limited amount of work that deals with these issues. Shapiro and Brewka's situation calculus based model of prioritized goals [10] has some unintuitive properties: the agent's chosen goals in $do(a, s)$ may

be quite different from her goals in s , although a did not make any of her goals in s impossible or inconsistent with higher priority goals, because inconsistencies between goals at the same priority level are resolved differently. In their framework, this can happen because goals are only partially ordered. Also, I provide a more expressive formalization of prioritized goals – I model goals using infinite paths, and thus can model many types of goals that they cannot. Finally my possible world semantics has some advantages over their account: it clearly defines when goals are consistent with each other and with what is known. One can easily specify how goals change when an action occurs.

In [3], the authors present an extension of the GOAL APL to incorporate temporally extended goals. Their formalization of goal dynamics is syntactic, in contrast to my semantic approach. Also, like the GOAL APL, their APL does not appeal to a BDI-style plan selection and hierarchical decomposition of plans, and rather uses simple (primitive) action selection rules.

In [5] and [4], I propose two frameworks that to some extent lies at the two extremes of the “resource-boundedness/tractability vs. ideal rationality” spectrum – [5] formalizes ideally rational agents that always reconsider their intentions while [4] formalizes over-committed agents that never give up their intentions when opportunities to commit to higher priority goals arise, thus effectively minimizing their reasoning costs w.r.t. intention reconsideration. Hence in the future, it would be interesting to develop a hybrid account of intention reconsideration where the agent is strongly committed to her chosen goals but where she reconsiders some of her prioritized goals under specific conditions. Moreover, I would like to identify a set of postulates for goal change and examine how they differ from belief change postulates. Furthermore, I would like to investigate on restricted versions of SR-APL to improve its efficiency/tractability, and more generally, to make it a practical APL. Finally, while my agent theory supports arbitrary temporally extended goals, in SR-APL I consider achievement goals only, which I would like to relax in the future.

5. ACKNOWLEDGMENTS

My work is conducted under the helpful guidance of Prof. Yves Lespérance.

6. REFERENCES

- [1] Bratman, M.E.: *Intentions, Plans, and Practical Reason*. Harvard Univ. Press, Cambridge, MA (1987)
- [2] De Giacomo, G., Lespérance, Y., Levesque, H.J.: ConGolog, a Concurrent Programming Language Based on the Situation Calculus. *Artificial Intelligence* **121** (2000) 109–169
- [3] Hindriks, K. V.; van der Hoek, W.; and van Riemsdijk, M. B.: 2009. Agent Programming with Temporally Extended Goals. In *AAMAS-09* (2009) 137–144
- [4] Khan, S.M. and Lespérance, Y.: Towards a Rational Agent Programming Language with Prioritized Goals. To appear in *DALT-10* (2010)
- [5] Khan, S.M. and Lespérance, Y.: A Logical Framework for Prioritized Goal Change. To appear in *AAMAS-10* (2010)
- [6] Khan, S.M. and Lespérance, Y.: Prioritized Goals and Subgoals in a Logical Account of Goal Change – A Preliminary Report. In *DALT-09*, LNAI Vol. 5948, Springer (2010) 119–136
- [7] Khan, S.M. and Lespérance, Y.: A Logical Account of Prioritized Goals and their Dynamics. In *Commonsense-09*, Outstanding student paper award (2009) 85–90
- [8] Khan, S.M. and Lespérance, Y.: Handling Prioritized Goals and Subgoals in a Logical Account of Goal Change (Extended Abstract). In *AAMAS-09* (2009) 1155–1156
- [9] Khan, S.M.: Agent Programming Languages with Declarative Goals : A Survey. *Technical Report CSE-2007-07*, York University, Toronto, ON, Canada, <http://www.cse.yorku.ca/techreports/2007/> (2007)
- [10] Shapiro, S., Brewka, G.: Dynamic Interactions Between Goals and Beliefs. In *IJCAI-07* (2007) 2625–2630