Goal Revision for a Rational Agent

Célia da Costa Pereira[1] and **Andrea G. B. Tettamanzi**[1] and **Leila Amgoud**[2]

Abstract. We propose a general framework to represent changes in the mental state of a rational agent due to the acquisition of new information and/or to the arising of new desires; fundamental postulates and properties of the function which generates the goal set are also provided.

1 Introduction

While constructing plans is central in planning, the problem of how goals arise also needs to be addressed. That problem may be approached as in over-subscription planning [8],or as a revision problem, in a spirit akin to belief revision.

Although there has been much discussion on belief revision [6, 5, 10], goal revision has not received much attention. The few works on goal *change* [4, 9, 7] do not build on belief revision. Their main lack is that agents do not use their own knowledge for revising goals.

We propose an approach for dynamically constructing the goal set to be pursued by a rational agent, by considering changes in its mental state. The key to understanding goal change lies in the fundamental assumption, which is a consequence of agent rationality, that the set of desires an agent pursues must only depend on the agent's state, and not on the history of its previous states.

2 Preliminaries

Following [3], we distinguish two disjoint sets of formulas: the set $\mathcal{D}$ of all possible desires of agents and the set $\mathcal{K}$ of all possible knowledge items.

Definition 1 (Desire-generating Rule) *A desire-generating rule is an expression of the form* $b_1 \wedge \ldots \wedge b_n \wedge d_1 \wedge \ldots \wedge d_m \Rightarrow d$*, where* $b_i \in \mathcal{K}$ *and* $d_j, d \in \mathcal{D}$*, with* $d \neq d_j$ *for all* j*.*

Definition 2 (Planning Rule) *A planning rule is an expression of the form* $d_1 \wedge \ldots \wedge d_n \rightarrow d$*, where* $d_i, d \in \mathcal{D}$*,* $d \neq d_i$ *for all* i*.*

We shall denote $\mathrm{lhs}(R)$ the set of literals that make up the conjunction on the left-hand side of a rule R, and $\mathrm{rhs}(R)$ the literal on the right-hand side of R. Given a set S of rules, $\mathrm{rhs}(S) = \{\mathrm{rhs}(R) : R \in S\}$.

Definition 3 (Agent's bases) *An agent is equipped with four bases:*

- *belief base:* $\mathcal{B} \subseteq \mathcal{K}$*;*
- *desire base:* $\mathcal{J} \subseteq \mathcal{D}$ *;*
- *desire-generating rule base:* $\mathcal{R}_J$*;*
- *planning rules base:* $\mathcal{P}$*.*

The state of an agent is completely described by a 4-tuple $\mathcal{S} = \langle \mathcal{B}, \mathcal{R}_J, \mathcal{J}, \mathcal{P} \rangle$. For the sake of simplicity, we assume here that the planning rules are given and fixed.

Definition 4 (Active Desire-generating Rule) *A desire-generating rule* R *is active in* $\mathcal{S}$ *iff* $\mathcal{S} \models \mathrm{lhs}(R)$*.*

The activation of a desire-generating rule brings about changes in the desire base of an agent. These changes, in turn, may cause the activation/deactivation of other rules. Let $\mathrm{Act}_z^{\mathcal{S}}$ be the set of all desire-generating rules activated by belief or desire z in $\mathcal{S}$, and $\mathrm{Deact}_z^{\mathcal{S}}$ the set of all desire-generating rules deactivated by belief or desire z in $\mathcal{S}$.

We do not expect a rational agent to formulate desires out of whim, but based on some rational argument. To model that state of affairs, desire-generating rules play the role of rational arguments and we define a desire to be justified as follows.

Definition 5 (Justified Desire) *A desire* d *is justified in state* $\mathcal{S}$ *iff* d *is in* $\mathcal{J}$*, i.e, there is an active desire-generating rule* R *in* $\mathcal{R}_J$ *such that* $\mathrm{rhs}(R) = d$*.*

3 Changes in the State of an Agent

The state $\mathcal{S}$ of an agent may change either because of the acquisition of a new belief b, in which case the well known AGM operator $*$ for belief revision [1] is applied to $\mathcal{B}$, or because a new desire d arises. A desire d is retracted from the desire set $\mathcal{J}$ if and only if d becomes not justified, i.e., all active desire-generating rule such that $\mathrm{rhs}(R) = d$ become inactive. A desire d is added to a desire set $\mathcal{J}$ if and only if the new information activates a desire-generating rule R with $\mathrm{rhs}(R) = d$.

If $A_b^{\mathcal{S}}$ be the set of desires acquired because of the new belief b, $A_b^{\mathcal{S}} = \mathrm{rhs}(\mathrm{Act}_b^{\mathcal{S}})$; if $L_b^{\mathcal{S}}$ is the set of desires lost because of the acquisition of the new belief b, $L_b^{\mathcal{S}} = \{d : d \in \mathrm{rhs}(\mathrm{Deact}_b^{\mathcal{S}}) \wedge \neg\exists R\,(\mathcal{S} \models \mathrm{lhs}(R) \wedge R \notin \mathrm{Deact}_b^{\mathcal{S}} \wedge \mathrm{rhs}(R) = d)\}$.

Let us denote by $\oplus$ a desire revision operator. Then, $\mathcal{J} \oplus b = (\mathcal{J} \cup A_b^{\mathcal{S}}) \setminus L_b^{\mathcal{S}}$. It is easy to verify that $A_b^{\mathcal{S}} \cap L_b^{\mathcal{S}} = \emptyset$, for all state $\mathcal{S}$.

Proposition 1 *The order in which two different pieces of information* b_i *and* b_j *are acquired does not affect the content of the final set of desires:* $(\mathcal{J} \oplus b_i) \oplus b_j = (\mathcal{J} \oplus b_j) \oplus b_i$*.*

For the sake of simplicity, we consider that a new desire arises when a desire-generating rule with an empty left hand side is inserted into $\mathcal{R}_J$. Let us denote by $\otimes$ the operator for updating a desire-generating rule base. When a new desire d arises, the desire-generating rule $\Rightarrow d$ is added to $\mathcal{R}_J$, d is added to $\mathcal{J}$, all desire-generating rules R in $\mathrm{Act}_d^{\mathcal{S}}$ become activacted, and all desires in the right hand side of these rules are also added to $\mathcal{J}$. Therefore, $\mathcal{J} \oplus d = \mathcal{J} \cup \{d\} \cup \{\mathrm{rhs}(R), \forall R \in \mathrm{Act}_d^{\mathcal{S}}\}$, and $\mathcal{R}_J \otimes d = \mathcal{R}_J \cup \{\Rightarrow d\}$.

[1] University of Milan, Italy, email: pereira,tettamanzi@dti.unimi.it
[2] IRIT-CNRS, France, email: amgoud@irit.fr

Proposition 2 *The order in which two different desires d_1 and d_2 arise does not affect the final state $\mathcal{S}$ of an agent: $\mathcal{J} \oplus d_1 \oplus d_2 = \mathcal{J} \oplus d_2 \oplus d_1$, and $\mathcal{R}_J \otimes d_1 \otimes d_2 = \mathcal{R}_J \otimes d_2 \otimes d_1$.*

Proposition 3 *Let b be a new piece of information and let d be a new desire for an agent. The new set of desires obtaining from the old set of desires $\mathcal{J}$ by considering b and d does not depend of the order in which these two news components are known by the agent: $\mathcal{J} \oplus b \oplus d = \mathcal{J} \oplus d \oplus b$, and $\mathcal{R}_J \otimes b \otimes d = \mathcal{R}_J \otimes d \otimes b$.*

4 Comparing Desires and Sets of Desires

An agent may have many desires. However, it is essential to be able to represent the fact that not all desires have the same importance or urgence for a rational agent. One approach would be to define a function $u : \mathcal{D} \to \mathbb{R}$ which associates a real value, utility, to all desires. An alternative approach would be to establish a (partial or total) ordering among desires. In either case, we can define a reflexive and transitive preference relation $\succeq$ between desires, by saying that $d \succeq d'$ iff the agent desires d at least as much as it desires d'. If utilities are defined, $d \succeq d'$ iff $u(d) \geq u(d')$. The $\succeq$ relation can be easily extended from desires to sets of desires: we omit the details due to lack of space.

5 Revising Goal Sets

The main point about desires is that we expect a rational agent to try and manipulate its surrounding environment to fulfill them. In general, not all desires can be fulfilled at the same time, especially when they are conflicting [2]. Therefore, a rational agent will select a justified, coherent and feasible set of desires to realize.

Definition 6 (Coherence of a Desire) *A desire d is coherent, w.r.t. $\mathcal{S}$, with a set of desires $\mathcal{J}$ iff $\exists \mathcal{I} \models \mathcal{S}$, s.t. $\forall w \in \mathcal{J}, \mathcal{I} \models d \wedge w$.*

Definition 7 (Partial plan) *A partial plan for acheiving desire d is a pair $\langle H, d \rangle$ such that: $d \in \mathcal{J}$ and $H = \{d_1, \ldots, d_n\}$ if there exists a planning rule $P = d_1 \wedge \ldots \wedge d_n \to d$.*

Definition 8 (Complete Plan) *A complete plan [2] for achieving a desire d is a pair $\langle C, d \rangle$ with $d \in \mathcal{J}$ and C is a finite tree such that its root is a partial plan $\langle H, d \rangle$, a node $\langle \{d_1, \ldots, d_n\}, d' \rangle$ has exactly n children $\langle H_1, d_1 \rangle, \ldots \langle H_n, d_n \rangle$, where $\langle H_i, d_i \rangle$ is a partial plan for d_i, and the leaves are partial plans of the form $\langle \emptyset, d_l \rangle$. The function $\mathrm{Des}(C)$ returns the set of all desires in C.*

Definition 9 (Feasible Desire) *A desire d is feasible if there exists a complete plan $\langle C, d \rangle$ to achieve it.*

Definition 10 (Conflict between complete plans) *Two complete plans $\langle C, d \rangle$ and $\langle C', d' \rangle$ are conflicting in a state $\mathcal{S}$, denoted as $\langle C, d \rangle \bowtie_{\mathcal{S}} \langle C', d' \rangle$, iff $\forall \mathcal{I} \models \mathcal{S}, \mathcal{I} \not\models \mathrm{Des}(C) \cup \mathrm{Des}(C') \cup \mathcal{B}$.*

Definition 11 (Feasible set of Desires) *A set of feasible desires $\{d_1, \ldots, d_n\}$ with the respective complete plans $\langle C_1, d_1 \rangle, \ldots, \langle C_n, d_n \rangle$, is said to be feasible in state $\mathcal{S}$ iff the plans C_i are not conflicting, i.e., iff there is an interpretation $\mathcal{I}$ such that $\mathcal{I} \models \mathcal{S}$ and $\mathcal{I} \models \mathrm{Des}(C_1) \cup \ldots \cup \mathrm{Des}(C_n) \cup \mathcal{B}$.*

Definition 12 (Consistency of a desire with a set) *A feasible desire d is consistent, w.r.t $\mathcal{S}$, with a set $\mathcal{J}$ of desires, $\mathrm{cons}(d|\mathcal{J})$, iff d is coherent with $\mathcal{J}$, justified and $\{d\} \cup \mathcal{J}$ is a feasible set.*

Definition 13 (Goal Set) *A goal set is a set of desires $\mathcal{G} \subseteq \mathcal{J}$ such that, $\forall d \in \mathcal{G}$, $\mathrm{cons}(d|\mathcal{G})$.*

5.1 Postulates for Goal Revision

In general, given a set of desires $\mathcal{J}$, there are many possible goal sets $\mathcal{G} \subseteq \mathcal{J}$. However, a rational agent in state $\mathcal{S}$ will elect as the set of goals it is pursuing one precise goal set $\mathcal{G}^*$, which depends on $\mathcal{S}$.

Let us call G the function which maps a state $\mathcal{S}$ into the goal set elected by a rational agent in state $\mathcal{S}$: $\mathcal{G}^* = G(\mathcal{S})$. Let $\odot$ be the goal revision operator. The goal election function G must obey three fundamental postulates:

- $(\mathbf{G} \odot \mathbf{1})$ $\forall \mathcal{S}, G(\mathcal{S})$ is a goal set;
- $(\mathbf{G} \odot \mathbf{2})$ $\forall \mathcal{S}$, if $\mathcal{G} \subseteq \mathcal{J}$ is a goal set, then $G(\mathcal{S}) \succeq \mathcal{G}$, i.e., a rational agent always selects the most preferable goal set.

A new desire d may become a goal in state $\mathcal{S}_d = \langle \mathcal{B}, \mathcal{J} \oplus d, \mathcal{R}_J \otimes d, \mathcal{P} \rangle$ either if d is consistent with $G(\mathcal{S})$ or if d is not consistent with $G(\mathcal{S})$, but it is preferred to all goals whose complete plans are all conflicting with all its complete plans. Therefore, revising the goal set $G(\mathcal{S})$ must be equivalent to updating the state of the agent, in particular $\mathcal{J}$ and $\mathcal{R}_J$, then electing the new goal set by means of function G.

Proposition 4 $G(\mathcal{S}) \odot d = G(\mathcal{S}_d)$.

6 Conclusion

Goal revision is more complex than AGM belief revision, because the last desire does not always prevail on previous desires, and a completely different approach is required for a satisfactory treatment.

Many simplifying hypotheses have been made, but we plan on relaxing them in future work.

REFERENCES

[1] Carlos E. Alchourrón, Peter Gärdenfors, and David Makinson, 'On the logic of theory change: Partial meet contraction and revision functions.', *J. Symb. Log.*, **50**(2), 510–530, (1985).

[2] L. Amgoud, 'A formal framework for handling conflicting desires', in *Proc. of the ECSQARU-2003 Conference*, pp. 552–563, (2003).

[3] L. Amgoud and I. Rahwan, 'An argumentation-based approach for practical reasoning', in *To Appear in Proc. of the AAMAS-2006 Conference*, (2006).

[4] John Bell and Zhisheng Huang, 'Dynamic goal hierarchies', in *PRICAI '96: Proceedings from the Workshop on Intelligent Agent Systems, Theoretical and Practical Issues*, pp. 88–103, London, UK, (1997). Springer-Verlag.

[5] Jon Doyle, 'Rational belief revision', in *KR'91: Principles of Knowledge Representation and Reasoning*, eds., James F. Allen, Richard Fikes, and Erik Sandewall, pp. 163–174, San Mateo, California, (1991). Morgan Kaufmann.

[6] P. Gärdenfors, 'The dynamics of belief systems: Foundations vs. coherence', *Revue Internationale de Philosophie*, (1989).

[7] Steven Shapiro, Yves Lespérance, and Hector J. Levesque, 'Goal change', in *IJCAI-05, Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence*, eds., Leslie Pack Kaelbling and Alessandro Saffiotti, pp. 582–588, Edinburgh, Scotland, UK, (July 30–August 5 2005). Professional Book Center.

[8] David E. Smith, 'Choosing objectives in over-subscription planning.', in *Proceedings of ICAPS 2004*, eds., S. Zilberstein, J. Koehler, and S. Koenig, pp. 393–401 (June 3–7 2004). AAAI.

[9] John Thangarajah, Lin Padgham, and James Harland, 'Representation and reasoning for goals in bdi agents', in *CRPITS '02: Proceedings of the twenty-fifth Australasian conference on Computer science*, pp. 259–265, Darlinghurst, Australia, Australia, (2002). Australian Computer Society, Inc.

[10] M-A. Williams, 'Towards a practical approach to belief revision: Reason-based change', in *Proceedings of the Fifth International Conference on Principles of Knowledge Representation and Reasoning*, eds., L.C. Aiello and C. Shapiro, pp. 412–421. Morgan Kaufmann Publishers, (1996).