Towards a Framework for Goal Revision

Célia da Costa Pereira   Andrea Tettamanzi

Università degli Studi di Milano
Dipartimento di Tecnologie dell'Informazione
Via Bramante 65, I-26013 Crema (CR), Italy
`pereira@dti.unimi.it`, `andrea.tettamanzi@unimi.it`

Abstract

A rational agent revises its goals if something changes in its mental state. In this paper, we propose (i) a general framework based on classical propositional logic, to represent changes in the mental state of the agent after the acquisition of new information and/or after the arising of new desires; (ii) fundamental postulates that the function which generates the goal set must obey; and (iii) properties that this function must have to guarantee both the agent's maximal satisfaction and the consistency of the goal set.

1 Introduction

Although there has been much discussion on belief revision, goal revision has not received much attention. To the best of our knowledge, the few works on goal *change* found in the literature do not build on results on belief revision. That is the case of [4], in which the authors propose a formal representation for goals as rational desires and introduce and formalize dynamic goal hierarchies, but do not formalize explicitly beliefs and plans; or of [8], in which the authors propose an explicit representation of goals suited for conflict resolution based on a preference ordering of sets of goals. A more recent approach is [6], which models a multi-agent system in which an agent adopts a goal if requested to do so and the new goal is not conflicting with existing goals. This approach is based on goal persistence, i.e., an agent maintains its goals unless explicitly requested to drop them by the originating agent. The main lack of these approaches is that agents do not use their own knowledge for revising goals.

A static approach to the problem of how goals arise has been proposed within planning is over-subscription planning [7, 5]. In over-subscription planning, the problem of identifying the best subset of goals, given resource constraints, is addressed. Our work is very much in that line, except that it attempts to provide a model of rationality, which is a slightly different focus.

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 formalism is kept very simple because it is intended as a first step. In particular, we use an unsophisticated definition of planning rules. A more sophisticated — and realistic — formalism for expressing plans and planning rules could be used; however, planning *per se* is not the focus of this work and we decided to keep things simple for the sake of clarity.

2 Example

At the beginning of 2006, Nick has $93,000. Nick wants to invest that amount and have a return of at least 40% at the end of the year with a probability of at least $\frac{2}{3}$ ($d_{40\%}$). Nick also wants to buy a house (d_H) at the beginning of 2007, and he knows he will need $100,000 with at least 90% probability ($d_{\$100,000}$) for it. Nick believes that X, a stock currently trading at $56, will reach $80 ($b_{80}$), and *a fortiori* also $60.24 ($b_{60}$), by the end of the year. Nick might face two alternative possibilities for investing in X: buy 1660 shares at $56 ($d_X$), then

A wait until X reaches \$80 to sell them (d_{80}), realizing a return of more than 40% with probability 68%, thus fulfilling $d_{40\%}$;

B wait until X reaches \$60.24 to sell them (d_{60}), realizing a wealth of \$100,000 with probability 95%, thus fulfilling d_H but not $d_{40\%}$.

Failing to sell at the desired level, Nick will sell all of his X stock at market price at the end of the year. Under possibility A, this will yield enough money to buy a new house with a probability of 12%, which gives a total probability of 80% for the ability to buy a new house, not enough to fulfill d_H.

3 Preliminaries

We present the logical language, inspired by [3], which will be used throughout this paper, as well as the four bases representing the mental state of an agent.

Desires are necessary, not sufficient, conditions for action. When a desire is met by other conditions that make it possible for an agent to act, that desire becomes a *goal*. Therefore, given this technical definition of a desire, all goals are desires, but not all desires are goals.

We distinguish two sets of literals: the set $\mathcal{D}$ of all possible desires of agents and the set $\mathcal{K}$ of all possible knowledge items. For the sake of simplicity we make the assumption that desires and knowledge items are on completely different levels: a desire is not a knowledge and vice versa. An interpretation $\mathcal{I}$ is an assignment of truth values to all propositions of $\mathcal{D}$ and $\mathcal{K}$.

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

The meaning of the rule is "if the agent believes $b_1, \ldots, b_n$ and desires $d_1, \ldots, d_m$, then the agent must desire d as well".

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.

The rule states that if $d_1, \ldots, d_n$ are fulfilled then d is fulfilled as well [1].

Given a desire-generating or a planning rule R, we shall denote $\mathrm{lhs}(R)$ the set of literals that make up the conjunction on the left-hand side of R, and $\mathrm{rhs}(R)$ the literal on the right-hand side of R. Furthermore, if S is a set of rules, we define $\mathrm{rhs}(S) = \{\mathrm{rhs}(R) : R \in S\}$.

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

- *belief base:* $\mathcal{B} \subseteq \mathcal{K}$[2]*;*
- *desire set:* $\mathcal{J} \subseteq \mathcal{D}$ *;*
- *desire-generating rule base:* $\mathcal{R}_J = \{R : R = b_1 \wedge \ldots \wedge b_n \wedge d_1 \wedge \ldots \wedge d_m \rightarrowtail d\}$ *with* $b_j \in \mathcal{K}$ *and* $d, d_k \in \mathcal{D}$*;*
- *planning rules base:* $\mathcal{P} = \{P : P = d_1 \wedge \ldots \wedge d_n \rightarrow d\}$ *with* $d, d_j \in \mathcal{D}$*;*

The state of an agent is completely described by a 4-tuple $\mathcal{S} = \langle \mathcal{B}, \mathcal{R}_J, \mathcal{J}, \mathcal{P} \rangle$. The set $\mathcal{B}$ represents the agent's knowledge about the world, $\mathcal{R}_J$ contains the rules which generate desires from beliefs and other (more basic) desires, $\mathcal{J}$ contains all desires which may be deduced from the agent's knowledge and the agents's desire-generating rule base, and $\mathcal{P}$ contains all available planning rules for achieving agent desires. For the sake of simplicity, we assume here that the planning rules are given and fixed.

In the example, Nick's state at the beginning of the year, $\mathcal{S}_0$, consists of $\mathcal{B} = \{b_{80}, b_{60}\}$, $\mathcal{R}_J = \{\rightarrowtail d_H, \rightarrowtail d_{40\%}, b_{80} \rightarrowtail d_{80}, b_{60} \wedge \neg b_{80} \rightarrowtail d_{60}\}$, $\mathcal{J} = \{d_H, d_{40\%}, d_{80}\}$, and $\mathcal{P} = \{d_{\$100{,}000} \rightarrow d_H, d_X \wedge d_{80} \rightarrow d_{40\%}, d_X \wedge d_{60} \rightarrow d_{\$100{,}000}, \neg d_{80} \rightarrow d_{60}, \neg d_{60} \rightarrow d_{80}\}$.

Definition 4 (Active Desire-generating Rule) *A desire-generating rule* R *is active in* $\mathcal{S}$ *iff* $\mathcal{S} \models \mathrm{lhs}(R)$[3].

[1] Note that the implications defined in desire-generating rules and planning rules are not material. So for example, from $\neg b$ and $a \rightarrow b$ or $a \rightarrowtail b$, we can not deduce $\neg a$.

[2] Unlike $\mathcal{J}$, $\mathcal{B}$ must be consistent.

[3] $\mathcal{S} \models \mathrm{lhs}(R)$ iff $\forall \mathcal{I}, \mathcal{I} \models \mathcal{S} \Rightarrow \mathcal{I} \models \mathrm{lhs}(R)$.

The next three definitions introduce a notation to refer to the set of rules that become active, resp. inactive, in a given state $\mathcal{S}$, if a new belief or a new desire are introduced in the respective bases. 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.

Definition 5 (Rule activated by a Belief/Desire) *Let z be a new belief or desire in state $\mathcal{S}$. We define a sequence $\{A_i\}_{i=0,\ldots}$ of subsets of $\mathcal{R}_\mathcal{J}$ as follows:*

$$\begin{aligned} A_0 &= \{R : z \in \text{lhs}(R) \wedge \forall x \in \text{lhs}(R), x \neq z \Rightarrow \mathcal{S} \models x\}, \\ A_i &= \{R : \text{lhs}(R) \cap \text{rhs}(A_{i-1}) \neq \emptyset \wedge \forall x \in \text{lhs}(R), x \notin \text{rhs}(A_{i-1}) \Rightarrow \mathcal{S} \models x\}. \end{aligned}$$

$\text{Act}_z^\mathcal{S} \equiv \bigcup_{i=0}^{\infty} A_i$ *is the set of all desire-generating rules activated by belief or desire z in $\mathcal{S}$. A_0 is the set of the desire-generating rules directly activated by z, i.e. with z in their left hand side. A_i is the set of desire-generating rules indirectly activated by z by way of rules in A_{i-1}.*

Definition 6 (Rule deactivated by a Belief/Desire) *Let z be a new belief or desire in the state $\mathcal{S}$. We define a sequence $\{D_i\}_{i=0,\ldots}$ of subsets of $\mathcal{R}_\mathcal{J}$ as follows:*

$$\begin{aligned} D_0 &= \{R : \mathcal{S} \models \text{lhs}(R) \wedge \neg z \in \text{lhs}(R)\}, \\ D_i &= \{R : \mathcal{S} \models \text{lhs}(R) \wedge \text{lhs}(R) \cap \text{rhs}(D_{i-1}) \neq \emptyset\}. \end{aligned}$$

$\text{Deact}_z^\mathcal{S} \equiv \bigcup_{i=0}^{\infty} D_i$ *is the set of all desire-generating rules deactivated by belief or desire z in $\mathcal{S}$. D_0 is the set of desire-generating rules directly deactivated by z, i.e. with $\neg z$ in their left hand side. D_i is the set of desire-generating rules indirectly deactivated by z by way of rules in D_{i-1}.*

In all four cases, when speaking of rule activation (resp. deactivation), we will say a rule R is *downstream* of another rule R' if $R' \in A_0$ (resp. D_0) and $R \in A_i$ (resp. D_i), with $i > 0$.

Finally, 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 7 (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* $\text{rhs}(R) = d$.

4 Changes in the State of an Agent

The acquisition of a new belief in state $\mathcal{S}$, may cause a change in the belief base $\mathcal{B}$ and this may also cause a change in the desire set $\mathcal{J}$ with the retraction of existing desires and/or the addition of new desires. 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 $\text{rhs}(R) = d$ 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 $\text{rhs}(R) = d$.

4.1 Changes caused by a New Belief

Let $\mathcal{S}$ be the state of an agent, b a new piece of information, $*$ the well known AGM operator for belief revision [1], $\oplus$ our operator for desire updating, and $\mathcal{J}$ the base of agent's desires. Two considerations must be taken into account:

1. By definition of the revision operator $*$, $b \in \mathcal{B} * b$, thus all desire-generating rules $R \in \text{Act}_b^\mathcal{S}$ become active and all new desires $d = \text{rhs}(R)$ are added to the desire set $\mathcal{J}$.

2. If, before the arrival of b, $\neg b \in \mathcal{B}$ then all active desire-generating rules R, such that $\neg b \in \text{lhs}(R)$, become inactive and, if there is not an active desire-generating rule R', such that $\text{rhs}(R') = \text{rhs}(R)$, then the desire $d = \text{rhs}(R)$ is retracted from the desire set $\mathcal{J}$.

We can summarize the above considerations into one desire-updating formula which tells how the desire set $\mathcal{J}$ of a rational agent in state $\mathcal{S}$ should change in response to the acquisition of a new belief b. Let $A_b^\mathcal{S}$ be the set of desires acquired because of the new belief b:

$$A_b^\mathcal{S} = \text{rhs}(\text{Act}_b^\mathcal{S}). \tag{1}$$

Let $L_b^{\mathcal{S}}$ be the set of desires lost because of the acquisition of the new belief b:

$$L_b^{\mathcal{S}} = \{d \quad : \quad d \in \text{rhs}(\text{Deact}_b^{\mathcal{S}}) \wedge \\ \neg\exists R\,(\mathcal{S} \models \text{lhs}(R) \wedge R \notin \text{Deact}_b^{\mathcal{S}} \wedge \text{rhs}(R) = d)\}. \tag{2}$$

According to the above considerations, we have:

$$\mathcal{J} \oplus b = (\mathcal{J} \cup A_b^{\mathcal{S}}) \setminus L_b^{\mathcal{S}}. \tag{3}$$

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:*

$$(\mathcal{J} \oplus b_i) \oplus b_j = (\mathcal{J} \oplus b_j) \oplus b_i. \tag{4}$$

The proof is based on the fact that (i) after acquiring b_i, all inactive desire-generating rules with $b_j \wedge \neg b_i$ in their left hand side, as well as their downstream rules will not be activated; (ii) $A_{b_k}^{\mathcal{S}} \cap L_{b_k}^{\mathcal{S}} = \emptyset$ for all k; and (iii) rules deactivated with the acquisition of b_j have either $\neg b_j$ or $b_i \wedge \neg b_j$ in their left hand side, or are downstream of such rules.

Going back to the example, if Nick learns that earnings of X dropped 10% with respect to his estimates, he has to revise his belief that X will reach \$80. Therefore, his belief base and desire set become, respectively, $\mathcal{B} = \{\neg b_{80}, b_{60}\}$ and $\mathcal{J} = \{d_H, d_{40\%}, d_{60}\}$.

4.2 Changes caused by a New Desire

In this work, 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 $\otimes$ be the operator for updating a desire-generating rule base, $\mathcal{S}$ be the state of the agent whose desire base is $\mathcal{J}$, and $A_d^{\mathcal{S}} = \{\text{rhs}(R) : R \in \text{Act}_d^{\mathcal{S}}\}$ be the set of desires acquired with the arising of d in the state $\mathcal{S}$. How does $\mathcal{S}$ change with the arising of the new desire d?

1. the desire-generating rule $\rightarrowtail d$ is added to $\mathcal{R}_J$,
2. d is added to $\mathcal{J}$,
3. all desire-generating rules R in $\text{Act}_d^{\mathcal{S}}$ become activacted, and all desires appearing in the right hand side of these rules are also added to $\mathcal{J}$.

Therefore,

$$\mathcal{J} \oplus d = \mathcal{J} \cup \{d\} \cup \{\text{rhs}(R), \forall R \in \text{Act}_d^{\mathcal{S}}\}. \tag{5}$$
$$\mathcal{R}_J \otimes d = \mathcal{R}_J \cup \{\rightarrowtail d\}. \tag{6}$$

If, before learning about the earnings drop, Nick decides he also wants to buy a new car (d_C), then $\mathcal{R}_J = \{\rightarrowtail d_H, \rightarrowtail d_{40\%}, b_{80} \rightarrowtail d_{80}, b_{60} \wedge \neg b_{80} \rightarrowtail d_{60}, \rightarrowtail d_C\}$, $\mathcal{J} = \{d_H, d_{40\%}, d_{80}, d_C\}$. $\mathcal{B}$ does not change.

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:*

$$\mathcal{J} \oplus d_1 \oplus d_2 = \mathcal{J} \oplus d_2 \oplus d_1. \tag{7}$$
$$\mathcal{R}_J \otimes d_1 \otimes d_2 = \mathcal{R}_J \otimes d_2 \otimes d_1. \tag{8}$$

The proof of these propositons is trivial.

Finally, we are interested in characterizing how the state $\mathcal{S}$ of an agent evolves when acquiring new beliefs after the arising of new desires and vice-versa. In particular, we have to check that the order of acquisition of new beliefs and desires does not matter in determining the agent's final state:

Proposition 3 *Let $\mathcal{J}$ be a set of desires, 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:*

$$\mathcal{J} \oplus b \oplus d = \mathcal{J} \oplus d \oplus b. \quad (9)$$

$$\mathcal{R}_J \otimes b \otimes d = \mathcal{R}_J \otimes d \otimes b. \quad (10)$$

The proof of Equation 9 in this proposition is based on an argument similar to the one used for proving Proposition 1; the proof of the other part is trivial.

5 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. A natural choice for representing the importance of desires would be to adopt the same concept of utility.

A utility function for desires is a function $u : \mathcal{D} \to \mathbb{R}$ which associates a real value, utility, to all desires. An important assumption we have to make is that if $d \in \mathcal{J}$, i.e., d is desired by a rational agent, it must be $u(d) > 0$. In other words, a rational agent cannot desire to incur a cost and cannot waste time having desires which, if realized, would not bring any benefit.

In the example, we may assume the utilities Nick attaches to his desires are: $u(d_H) = 100$, $u(d_{40\%}) = 40$, $u(d_{80}) = 4$, $u(d_{60}) = 3$.

One problem with utilities is that, in general, we are not able to attach a precise numerical value to desires. However, in some contexts, "desires", i.e., given sets of states of the world, might lend themselves naturally to an economic valuation. For instance, if the agent is an investor, its plans are the purchase or sale of financial instruments, and its desires are levels of wealth or return on investment, attaching a dollar value to desires would be the most natural way of representing their importance. An alternative approach would be to establish a (partial or total) ordering among desires. In either case, we can define preference between desires as follows.

Definition 8 (Preference between Desires) *A desire d is preferred to a desire d', denoted $d \succeq d'$ iff the agent desires d at least as much as it desires d'.*

Of course, if utilities are defined, $d \succeq d'$ iff $u(d) \geq u(d')$. Therefore, in the example $d_H \succeq d_{40\%} \succeq d_{80} \succeq d_{60}$.

The $\succeq$ relation, which is reflexive and transitive, can be extended from desires to sets of desires.

Definition 9 (Preference between Sets of Desires) *A set of desires D_1 is preferred to a set of desires D_2, denoted $D_1 \succeq D_2$:*

- *if utilities are defined, iff*

$$\sum_{d \in D_1} u(d) \geq \sum_{d \in D_2} u(d);$$

- *otherwise, iff one of the following three conditions is satisfied:*

  1. *$D_2 \subseteq D_1$;*
  2. *$\forall d \in D_2$, $\exists d' \in D_1$, s.t. $d' \succeq d$;*
  3. *neither 1 nor 2 are satisfied, and $\forall d \in D_1$, $\exists d' \in D_2$, s.t. $d \succeq d'$.*

For example, $\{d_H, d_{60}\} \succeq \{d_H\} \succeq \{d_{40\%}, d_{80}\} \succeq \{d_{40\%}, d_{60}\}$.

6 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 10 (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. $\forall w \in \mathcal{J}$, $\mathcal{I} \models d \wedge w$.*

Besides being coherent with other desires, a desire must be feasible. Nick might desire to get a 100% return with 99% probability from his investment: while that desire is coherent with Nick's other desires, unfortunately for him, it is not feasible. Given that some desires are not feasible, a rational agent should concentrate on feasible desires. To define what "feasible" means, we need a number of other definitions.

Definition 11 (Partial plan) *A partial plan for acheiving desire d in the state $\mathcal{S}$ is a pair $\langle H, d\rangle$ such that: (i) $d \in \mathcal{J}$ and (ii) $H = \{d_1, \ldots, d_n\}$ if there exists a planning rule $P = d_1 \wedge \ldots \wedge d_n \rightarrow d$.*

Note that a desire may have several partial plans.

The complete way to fulfill a given desire d is called in [2] a *complete plan*. A *complete plan* for a given desire d is an *AND* tree. Its nodes are partial plans and its edges represent the relationship between desires necessary for the justification of d. It is an *AND* tree because all these desires must be considered.

Definition 12 (Complete Plan) *A complete plan for achieving a desire d in the state $\mathcal{S}$ is a pair $\langle C, d\rangle$ with $d \in \mathcal{J}$ and C is a finite tree such that: (i) The root of the tree is a partial plan $\langle H, d\rangle$; (ii) 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 (iii) The leaves of the tree are partial plans $\langle H_l, d_l\rangle$ with $H_l = \emptyset$.*
The function $\mathrm{Des}(C)$ returns the set of all desires in C.

If Nick's mental state is $\mathcal{S}_0$, then the unique complete plan for desire d_H has root $\langle\{d_{\$100{,}000}\}, d_H\rangle$ which has one child $\langle\{d_X, d_{60}\}, d_{\$100{,}000}\rangle$, which in turn has two children $\langle\emptyset, d_X\rangle$ and $\langle\{\neg d_{80}\}, d_{60}\rangle$. The latter has one child $\langle\emptyset, \neg d_{80}\rangle$.

The unique complete plan for desire $d_{40\%}$ has root $\langle\{d_X, d_{80}\}, d_{40\%}\rangle$ which in turn has two children $\langle\emptyset, d_X\rangle$ and $\langle\{\neg d_{60}\}, d_{80}\rangle$. The latter has one child $\langle\emptyset, \neg d_{60}\rangle$

Definition 13 (Feasible Desire) *A desire d is feasible if there exists a complete plan $\langle C, d\rangle$ to achieve it.*

Definition 14 (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 $\mathrm{Des}(C) \cup \mathrm{Des}(C') \cup \mathcal{B} \vdash \bot$.*

Definition 15 (Feasible set of Desires) *A set of feasible desires $\{d_1, \ldots, d_n\}$ is said to be feasible in state $\mathcal{S}$ iff $\forall \langle C_1, d_1\rangle, \ldots, \langle C_n, d_n\rangle$, $\mathrm{Des}(C_1) \cup \ldots \cup \mathrm{Des}(C_n) \cup \mathcal{B} \vdash \top$.*

It is clear that the set of Nick's desires in state $\mathcal{S}_0$, $\{d_H, d_{40\%}, d_{80}\}$ is not feasible.

Definition 16 (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.*

The set of the goals a rational agent is pursuing must be consistent.

Definition 17 (Goal Set) *A goal set in the state $\mathcal{S}$ is a set of desires $\mathcal{G} \subseteq \mathcal{J}$ such that, $\forall d \in \mathcal{G}$, $\mathrm{cons}(d|\mathcal{G})$.*

6.1 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 $\mathcal{S} = \langle\mathcal{B}, \mathcal{J}, \mathcal{R}_J, \mathcal{P}\rangle$ 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 $\mathcal{S}$: $\mathcal{G}^* = G(\mathcal{S})$. This goal election function G must obey two fundamental postulates:

- (**G** $\odot$ **1**) $\forall \mathcal{S}, G(\mathcal{S})$ is a goal set;
- (**G** $\odot$ **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.

Let d be a new desire arising in state $\mathcal{S}$, let $\mathcal{S} = \langle \mathcal{B}, \mathcal{J}, \mathcal{R}_J, \mathcal{P} \rangle$ be the state of the agent before desire d, let $\mathcal{S}_d = \langle \mathcal{B}, \mathcal{J} \oplus d, \mathcal{R}_J \otimes d, \mathcal{P} \rangle$ be the state resulting from the arising of the new desire d in the state $\mathcal{S}$. The function G must respect the following properties:

- (**P1**) If $\text{cons}(d|G(\mathcal{S}))$ then $d \in G(\mathcal{S}_d)$;
- (**P2**) If $\neg\text{cons}(d|G(\mathcal{S}))$, and there is a goal g in $G(\mathcal{S})$ whose complete plans are all conflicting with all possible complete plans for d and $g \succeq d$ then $G(\mathcal{S}_d) = G(\mathcal{S})$. Thus, in case it is impossible to maintain both the new desire and the old desire, a rational agent maintains its old desire.
- (**P3**) If $\neg\text{cons}(d|G(\mathcal{S}))$ and for all goals g in $G(\mathcal{S})$ whose complete plans are all conflicting with all possible complete plans for d we have $d \succeq g$ then $d \in G(\mathcal{S}_d)$ and $g \notin G(\mathcal{S}_d)$.

Summarizing, a new desire d may become a goal in state $\mathcal{S}_d$ in two cases:

- d is consistent with $G(\mathcal{S})$.
- 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.

Let $\odot$ be the goal revision operator. We have $G(\mathcal{S}) \odot d = G(\mathcal{S}_d)$.

6.2 Defining the Goal Set Election Function

We now propose three definitions of G, namely G_u, $G_{\succeq}$, and $G_{\subseteq}$ which are applicable, respectively, to the case whereby utilities are defined, to the weaker case in which the total ordering $\succeq$ of desires is available, and to the weakest case in which only a partial ordering or no ordering at all of desires is available.

Function G_u is computed by solving the following combinatorial optimization problem: given state $\mathcal{S}$,

$$\begin{array}{ll} \text{maximize} & u(\mathcal{G}) = \sum_{g \in \mathcal{G}} u(g), \\ \text{subject to} & \mathcal{G} \subseteq \mathcal{J}, \\ & \mathcal{G} \text{ is a goal set.} \end{array} \tag{11}$$

Going back to the example, $G_u(\mathcal{S}_0) = \{d_H, d_{80}\}$.

Function $G_{\succeq}$ can be computed in $O(\|\mathcal{J}\|^2)$ time by means of the following algorithm. Given $\mathcal{S}$, consider any ordering $d_1, d_2, \ldots, d_{\|\mathcal{J}\|}$ of the elements of $\mathcal{J}$.

1. $\mathcal{G} \leftarrow \emptyset$;
2. $i \leftarrow 0$;
3. if $i = \|\mathcal{J}\|$, end: $G(\mathcal{S}) = \mathcal{G}$;
4. $i \leftarrow i + 1$;
5. if $\text{cons}(d_i|\mathcal{G})$, then $\mathcal{G} \leftarrow \mathcal{G} \cup \{d_i\}$, go to 3;
6. let $D = \{g \in \mathcal{G}$: the plans for g and the plans for d_i are conflicting, or $\forall \mathcal{I}, \mathcal{I} \not\models g \wedge d_i\}$; if $\{d_i\} \succeq D$, then $\mathcal{G} \leftarrow \mathcal{G} \cup \{d_i\} \setminus D$, go to 2;
7. otherwise, go to 3.

In the worst case, i.e., when the desires are ordered in such a way that, each time a new desire is considered, it is inconsistent with $\mathcal{G}$ and the condition in Step 6 is satisfied, the total number of iterations is $\|\mathcal{J}\|(\|\mathcal{J}\| - 1)/2$.

The above algorithm always terminates with the most preferable goal set, since, at each iteration t, $\mathcal{G}_t$ is a goal set, $\mathcal{G}_{t+1} \succeq \mathcal{G}_t$, and the algorithm attempts adding all d_i's to $\mathcal{G}$; whenever a desire already in $\mathcal{G}$ must be dropped, the algorithm attempts again adding all d_i's to $\mathcal{G}$.

We observe that, in the case of the example, $G_{\succeq}(\mathcal{S}_0) = G_u(\mathcal{S}_0)$. This happens because we dispose of utilities and the preference relation derives from utilities when the latter are available.

Function $G_{\subseteq}$ returns the maximal (for set inclusion) subset of consistent desires. The complexity of the corresponding problem appears to be NP-hard. We are still working on a proof of this claim.

In the example, $G_{\subseteq}(\mathcal{S}_0)$ may return two consistent sets of desires $\{d_H, d_{80}\}$ and $\{d_{40\%}, d_{80}\}$.

7 Conclusion

We believe the key points of this work are the following:

(i) the proposition of a framework for goal revision;

(ii) the idea of computing the goal set from the mental state of an agent, instead of a previous goal set;

(iii) the provision of three alternative methods to find the most preferable goal set under different hypotheses about desire preference information.

Many simplifying hypotheses have been made, but we plan on relaxing them in future work. The cost of actions and, as a consequence, of plans, has not been considered, but its role is not negligible in real applications and should be taken into account, for example by discounting the utility of a desire with the cost of the optimal plan to achieve it; interactions and trade-offs between desires could be in order and should be properly accounted for. This direction of investigation calls for an integration of decision theoretic concepts and results in our work. The use of a classical logic framework should not be seen as a limitation, as the same, or very similar, results can be proved for different logical frameworks, e.g., argumentation.

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*, pages 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*, pages 88–103, London, UK, 1997. Springer-Verlag.

[5] J. Benton, Minh Binh Do, and Subbarao Kambhampati. Over-subscription planning with numeric goals. In Leslie Pack Kaelbling and Alessandro Saffiotti, editors, *IJCAI-05, Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence*, pages 1207–1213, Edinburgh, Scotland, UK, July 30–August 5 2005. Professional Book Center.

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

[7] David E. Smith. Choosing objectives in over-subscription planning. In Shlomo Zilberstein, Jana Koehler, and Sven Koenig, editors, *Proceedings of the Fourteenth International Conference on Automated Planning and Scheduling (ICAPS 2004)*, pages 393–401, Whistler, British Columbia, Canada, June 3–7 2004. AAAI.

[8] 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*, pages 259–265, Darlinghurst, Australia, Australia, 2002. Australian Computer Society, Inc.