Theories of Intentions in the framework of Situation Calculus

Pilar Pozos Parra[1], Abhaya Nayak[1], and Robert Demolombe[2]

[1] Division of ICS, Macquarie University
NSW 2109, Australia
{pilar,abhaya}@ics.mq.edu.au
[2] ONERA-Toulouse
2 Avenue E. Belin BP 4025, 31055 Toulouse, France
Robert.Demolombe@cert.fr

Abstract. We propose an extension of action theories to intention theories in the framework of situation calculus. Moreover the method for implementing action theories is adapted to consider the new components. The intention theories take account of the BDI (Belief-Desire-Intention) architecture. In order to avoid the computational complexity of theorem proving in modal logic, we explore an alternative approach that introduces the notions of belief, goal and intention fluents together with their associated successor state axioms. Hence, under certain conditions, reasoning about the BDI change is computationally similar to reasoning about ordinary fluent change. The approach can be implemented using declarative programming.

1 Introduction

Various authors have attempted to logically formulate the behaviour of rational agents. Most of them use modal logics to formalize cognitive concepts, such as beliefs, desires and intentions [1–6]. A weakness of the modal approaches is that they overestimate the reasoning capabilities of agents; consequently problems such as logical omniscience arise in such frameworks. Work on implementing modal systems is still scarce, perhaps due to the high computational complexity of theorem-proving or model-checking in such systems [7–9]. A proposal [10] based on the situation calculus not only allows representation of the BDI notions and their evolution, but also attends to finding a trade-off between the expressive power of the formalism and the design of a realistic implementation.

A theory of intentions requires a well-defined theory of actions, such as one provided in the situation calculus. In this paper, we propose to enhance Reiter's action theories [11] with BDI representation [10] to build intention theories. The notion of *knowledge-producing* actions is generalized to *propositional attitude-producing* actions whose effects modify the agent's beliefs, goals and intentions. We show how the proposed framework can be implemented using the method for implementing Reiter's action theories. The scenario presented in this paper has been implemented in Prolog.

The paper is organised as follows. We start with a brief review of the situation calculus and its use in the representation issues involving the evolution of the world and mental states. In Section 3, we define the basic theories of intentions and the method used to implement such theories. In Section 4, we present an example. In conclusion we discuss some of the issues.

2 Situation Calculus

The situation calculus allows modelling dynamic worlds [12]. It involves three types of terms, among which *situation* and *action* play an important role. In the following, s represents an arbitrary situation, and a an action. The result $do(a,s)$ of performing a in s is taken to be a situation. The world's properties (in general relations) that are susceptible to change, are represented by predicates whose last argument is of type *situation* called "fluents". For any fluent p and situation s, the expression $p(s)$ denotes the truth value of p in s. It is assumed that every change is caused by an action. The evolution of fluents is represented by "successor state axioms". These axioms were introduced to solve the infamous frame problem.

In order to distinguish between what relations are true in a situation and what relations are believed to be true or false in a situation, the notion of "belief fluents" together with the "successor belief state axioms" are introduced in [13].[3] This model of dynamic beliefs has been extended in order to consider dynamic generalised beliefs, dynamic goals and dynamic intentions in [10]. In short, the dynamic mental states are represented by suitable new fluents (such as *belief, goal* and *intention fluents*) and their appropriate successor state axioms. These axioms are a proposal to solve the corresponding frame problem in mental states. This approach has been compared with other formalisations of BDI architecture, in particular with the Cohen and Levesque's approach, in [10].

2.1 Dynamic Worlds

For a fluent p, the successor state axiom $\mathbf{S_p}$ is of the form:[4]

$$(\mathbf{S_p})\quad p(do(a,s)) \leftrightarrow \Upsilon_p^+(a,s) \vee (p(s) \wedge \neg\Upsilon_p^-(a,s))$$

where $\Upsilon_p^+(a,s)$ captures exactly the conditions under which p turns from false to true when a is performed in s, and similarly $\Upsilon_p^-(a,s)$ captures exactly the conditions under which p turns from true to false when a is performed in s. It is assumed that no action can turn p to be both true and false in a situation. These axioms define the truth values of the atomic formulas in any circumstances, and indirectly the truth value of every formula. Furthermore, in order to solve the qualification problem, a special fluent $Poss(a,s)$, meaning it is pos-

[3] A comparison with Scherl and Levesque's approach has been presented in [14].

[4] In what follows, it is assumed that all the free variables are universally quantified.

sible to execute the action a in situation s, was introduced, as well as the action preconditions axioms of the form:

$(\mathbf{P_A})$ $Poss(A,s) \leftrightarrow \Pi_A(s)$

where A is an action symbol and $\Pi_A(s)$ a formula that defines the preconditions for the executability of the action A in s. Note that Reiter's notation [11] shows explicitly all the fluent arguments $(p(x_1,\ldots,x_n,do(a,s)),\ \Upsilon_p^+(x_1,\ldots,x_n,a,s))$ and action arguments $(Poss(A(x_1,\ldots,x_n),s)$ or $\Pi_A(x_1,\ldots,x_n,s))$. For the sake of readability we show solely the action and situation arguments.

2.2 Dynamic Beliefs

A belief fluent is a syntactic combination of a modal operator and fluent or its negation. We say that the "modalised" fluent B_ip holds in situation s iff agent i believes that p holds in situation s and represent it as $B_ip(s)$. Similarly $B_i\neg p(s)$[5] represents the fact that the fluent $B_i\neg p$ holds in situation s: the agent i believes that p does not hold in situation s.

In this case, the evolution needs to be represented by two axioms, each allowing the representation of two attitudes out of four i's attitudes concerning her belief about the fluent p, namely $B_ip(s)$, $\neg B_ip(s)$, $B_i\neg p(s)$ and $\neg B_i\neg p(s)$. The successor belief state axioms for an agent i and a fluent p are of the form:

$(\mathbf{S_{B_ip}})$ $B_ip(do(a,s)) \leftrightarrow \Upsilon_{B_ip}^+(a,s) \lor (B_ip(s) \land \neg\Upsilon_{B_ip}^-(a,s))$

$(\mathbf{S_{B_i\neg p}})$ $B_i\neg p(do(a,s)) \leftrightarrow \Upsilon_{B_i\neg p}^+(a,s) \lor (B_i\neg p(s) \land \neg\Upsilon_{B_i\neg p}^-(a,s))$

where $\Upsilon_{B_ip}^+(a,s)$ are the precise conditions under which the state of i (with regards to the fact that p holds) changes from one of disbelief to belief when a is performed in s, and similarly $\Upsilon_{B_ip}^-(a,s)$ are the precise conditions under which the state of i changes from one of belief to disbelief. The conditions $\Upsilon_{B_i\neg p}^+(a,s)$ and $\Upsilon_{B_i\neg p}^-(a,s)$ have a similar interpretation. In these axioms as well as in the goals and intentions axioms, p is restricted to be a fluent representing a property of the real world. Some constraints must be imposed to prevent the derivation of inconsistent beliefs (see Section 3.1).

To address the qualification problem in the belief context, we have the belief fluent $B_iPoss(a,s)$, which represents the belief of agent i in s about the possible execution of the action a in s.

2.3 Dynamic Generalised Beliefs

The statements of the form $B_ip(s)$ represent i's beliefs about the present. In order to represent the agent's beliefs about the past and the future, the notation $B_ip(s',s)$ has been introduced, which means that in situation s, the agent i

[5] We abuse of notation B_ip and $B_i\neg p$ in order to have an easy identification of the agent and proposition. An "adequate" notation could be *Bip* and *Binotp*. A similar notation is used to represent goals and intentions.

believes that p holds in situation s'. Depending on whether $s' = s$, $s' \sqsubset s$ or $s \sqsubset s'$,[6] it represents belief about the present, past or future respectively.

The successor belief state axioms $\mathbf{S_{B_ip}}$ and $\mathbf{S_{B_i\neg p}}$ are further generalized to successor generalised belief state axioms as follows:

$(\mathbf{S_{B_ip(s')}})$ $B_ip(s', do(a,s)) \leftrightarrow \Upsilon^+_{B_ip(s')}(a,s) \vee (B_ip(s',s) \wedge \neg\Upsilon^-_{B_ip(s')}(a,s))$

$(\mathbf{S_{B_i\neg p(s')}})$ $B_i\neg p(s', do(a,s)) \leftrightarrow \Upsilon^+_{B_i\neg p(s')}(a,s) \vee (B_i\neg p(s',s) \wedge \neg\Upsilon^-_{B_i\neg p(s')}(a,s))$

where $\Upsilon^+_{B_ip(s')}(a,s)$ captures exactly the conditions under which, when a is performed in s, i comes believing that p holds in s'. Similarly $\Upsilon^-_{B_ip(s')}(a,s)$ captures exactly the conditions under which, when a is performed in s, i stops believing that p holds in s'. The conditions $\Upsilon^+_{B_i\neg p(s')}(a,s)$ and $\Upsilon^-_{B_i\neg p(s')}(a,s)$ are similarly interpreted. These conditions may contain communication actions or sensing actions which are examples of belief-producing actions. Communication actions allow the agent to gain information about the world in the past, present or future. For instance, if the agent receives one the following messages: "it was raining yesterday", "it is raining" or "it will rain tomorrow", then her beliefs about the existence of a precipitation, in the past, present and future respectively, can be revised. Sensing actions allow the agent to gain information solely in the present. For instance, if the agent observes raindrops, her belief about a current precipitation can be revised.

$B_iPoss(a,s',s)$ was introduced in order to solve the qualification problem about i's beliefs. The action precondition belief axioms are of the form:

$(\mathbf{P'_{A_i}})$ $B_iPoss(A,s',s) \leftrightarrow \Pi'_{A_i}(s',s)$.

where A is an action symbol and $\Pi'_{A_i}(s',s)$ a formula that defines the preconditions for i's belief in s concerning the executability of the action A in s'. A general setting can consider also the axioms of the form: $B_i\neg Poss(A,s',s) \leftrightarrow \Pi''_{A_i}(s',s)$ where $B_i\neg Poss(A,s',s)$ means that in s the agent i believes that it is not possible to execute the action A in s'.

Notice that s' may be non-comparable with $do(a,s)$ under $\sqsubset$. However, this can be used to represent hypothetical reasoning: although situation s' is not reachable from $do(a,s)$ by a sequence of actions, yet, $B_ip(s', do(a,s))$ means that i, in $do(a,s)$, believes that p would have held if the actions of s' had happened. We are mainly interested in beliefs about the future. Since to make plans, the agent must project her beliefs to the future to "discover" a situation s' in which her goal p holds. In other words, in the current situation s (present) the agent must find a sequence of actions to reach s' (hypothetical future), and she expects that her goal p will hold in s'. Therefore, we adopt the notation: $Bf_ip(s',s) \stackrel{\text{def}}{=} s \sqsubset s' \wedge B_ip(s',s)$ to denote future projections. Similarly, to represent the expectations of executability of actions in future situations, we have: $Bf_iPoss(a,s',s) \stackrel{\text{def}}{=} s \sqsubset$

[6] The predicate $s' \sqsubset s$ represents the fact that the situation s is obtained from s' after performance of one or several actions.

$s' \wedge B_i Poss(a, s', s)$ that represents the belief of i in s about the possible execution of a in the future situation s'.

2.4 Dynamic Goals

The goal fluent $G_i p(s)$ (respectively $G_i \neg p(s)$) means that in situation s, the agent i has the goal that p be true (respectively false). As in the case of beliefs, an agent may have four different goal attitudes concerning the fluent p. The evolution of goals is affected by actions of the sort "adopt a goal" or "admit defeat of a goal" called goal-producing actions. For each agent i and fluent p, we have two successor goal state axioms of the form:

$(\mathbf{S_{G_i p}})$ $G_i p(do(a, s)) \leftrightarrow \Upsilon^{+}_{G_i p}(a, s) \vee (G_i p(s) \wedge \neg \Upsilon^{-}_{G_i p}(a, s))$

$(\mathbf{S_{G_i \neg p}})$ $G_i \neg p(do(a, s)) \leftrightarrow \Upsilon^{+}_{G_i \neg p}(a, s) \vee (G_i \neg p(s) \wedge \neg \Upsilon^{-}_{G_i \neg p}(a, s))$

As in the case of beliefs, $\Upsilon^{+}_{G_i p}$ represents the exact conditions under which, when the action a is performed in s, the agent i comes to acquire as a goal 'p holds'. The other conditions Υ's can be analogously understood. The indifferent attitude about p can be represented by $\neg G_i p(s) \wedge \neg G_i \neg p(s)$. Some constraints must be imposed on the conditions Υ's in order to prevent the agent having inconsistent goals such as $G_i p(s) \wedge G_i \neg p(s)$, meaning the agent wants p to both hold and not hold simultaneously (see Section 3.1). A related example of inconsistency is when the agent wants to be at the same time divorced and not divorced.

2.5 Dynamic Intentions

Let T be the sequence of actions $[a_1, a_2, \ldots, a_n]$. The fact that an agent has the intention to perform T in the situation s to satisfy her goal p (respectively $\neg p$) is represented by the intention fluent $I_i p(T, s)$ (respectively $I_i \neg p(T, s)$). In the following, the notation $do(T, s)$ is used to represent $do(a_n, \ldots, do(a_2, do(a_1, s)) \ldots)$ when $n > 0$ and s when $n = 0$. For each agent i and fluent p, the successor intention state axioms are of the form:

$(\mathbf{S_{I_i p}})$ $I_i p(T, do(a, s)) \leftrightarrow G_i p(do(a, s)) \wedge [$
$(a = commit(T) \wedge Bf_i Poss(do(T, s), s) \wedge Bf_i p(do(T, s), s)) \vee$
$I_i p([a|T], s) \vee$
$\Upsilon'^{+}_{I_i p}(a, s) \vee$
$(I_i p(T, s) \wedge \neg \Upsilon'^{-}_{I_i p}(a, s))]$

$(\mathbf{S_{I_i \neg p}})$ $I_i \neg p(T, do(a, s)) \leftrightarrow G_i \neg p(do(a, s)) \wedge [$
$(a = commit(T) \wedge Bf_i Poss(do(T, s), s) \wedge Bf_i \neg p(do(T, s), s)) \vee$
$I_i \neg p([a|T], s) \vee$

$\Upsilon'^{+}_{I_i \neg p}(a, s) \vee$
$(I_i \neg p(T, s) \wedge \neg \Upsilon'^{-}_{I_i \neg p}(a, s))]$

where Υ''s capture some conditions under which i's intention attitude (concerning T and goal p) change when a is performed in s. Intuitively, $\mathbf{S_{I_i p}}$ means that in the situation $do(a, s)$, agent i intends to perform T to achieve goal p iff

(a) In $do(a, s)$ the agent has goal p; and
(b) either
  (1) the agent has just committed to execute the sequence of actions T (a plan): the action $commit(T)$ is executed in s, the agent believes that the execution of such a plan is possible $Bf_i Poss(do(T, s), s)$, and she expects that her goal will be satisfied after the execution of the plan $Bf_i p(do(T, s), s))$; or
  (2) in the previous situation, the agent had the intention to perform the sequence $[a|T]$ and the action a has just happened; or
  (3) a condition $\Upsilon'^{+}_{I_i p}(a, s)$ is satisfied; or
  (4) in the previous situation s, the agent had the same intention $I_i p(T, s)$ and the condition $\Upsilon'^{-}_{I_i p}(a, s)$ does not hold; this condition has the effect to abandon her intention.

This definition of intention, as Cohen and Levesque say, allows relating goals with beliefs and commitments. The action $commit(T)$ is an example of intention-producing actions that affect the evolution of intentions. An advantage of this approach is that we can distinguish between a rational intention trigger by condition 1 after analysis of present and future situations, and an impulsive intention trigger by condition 3 after satisfaction of $\Upsilon'^{+}_{I_i p}(a, s)$ that may not concern any analysis process (for example, running intention after seeing a lion, the agent runs by reflex and not having reasoned about it).

We have considered a "credulous" agent who makes plan only when she commits to follow her plan: she is convinced that there are not exogenous actions. However, other kinds of agents may be considered. For instance, if the projection to the future is placed at the goal level, we can define a "prudent" agent that replans after every action that "fails" to reach her goal. Discussion of prudent agents is beyond the scope of this paper.

Intuitively, $Bf_i Poss(do(T, s), s)$ means that in s, i believes that all the actions occurring in T can be executed one after the other.

$$Bf_i Poss(do(T, s), s) \stackrel{\text{def}}{=} \bigwedge_{i=1}^{n} Bf_i Poss(a_i, do([a_1, a_2, \ldots, a_{i-1}], s), s).$$

Notice the similarity of $Bf_i Poss(do(T, s), s)$ with an executable situation defined in [11] as follows:

$$executable(do(T, S_0)) \stackrel{\text{def}}{=} \bigwedge_{i=1}^{n} Poss(a_i, do([a_1, a_2, \ldots, a_{i-1}], S_0))$$

$executable(do(T, S_0))$ means that all the actions occurring in the action sequence T can be executed one after the other. However, there are differences to consider. In $executable(do(T, S_0))$, T is executable if the preconditions for every action in the sequence hold in the corresponding situation. On the other hand in $Bf_i Poss(do(T, s), s)$, T is believed to be executable in s if the agent believes that the preconditions for every action in T hold in the corresponding situation.

3 Intention Theories

Now we extend the language presented in [11] with cognitive fluents and we introduce the BDI notions to the action theories to build the intention theories. We adapt regression [11] appropriately to this more general setting. The extension of results about implementation of intention theories is immediate.

Let's assume $\mathcal{L}_{sitcalc}$, a language formally defined in [11]. This language has a countable number of predicate symbols whose the last argument is of type *situation*. These predicate symbols are called relational fluents and denote situation dependent relations such as $position(x, s)$, $student(Billy, S_0)$ and $Poss(advance, s)$. We extend this language to $\mathcal{L}_{sitcalc_{BDI}}$ with the following symbols: belief predicate symbols $B_i p$ and $B_i \neg p$, goal predicate symbols $G_i p$ and $G_i \neg p$, and intention predicate symbols $I_i p$ and $I_i \neg p$, for each relational fluent p and agent i. These predicate symbols are called belief, goal and intention fluents respectively and denote situation dependent mental state of agent i such as $B_{robot} position(1, S_0)$, $G_{robot} position(3, S_0)$, $I_{robot} position(3, [advance, advance], S_0)$: in the initial situation S_0, the robot believes to be in 1, wants to be in 3 and has the intention of advancing twice to fulfill her goal.

As a matter of simplification we consider only the languages without functional fluents (see [11] for extra axioms that deal with function fluents).

Definition 1. A *basic intention theory* $\mathcal{D}$ has the following form:
$\mathcal{D} = \Sigma \cup \mathcal{D}_{S_0} \cup \mathcal{D}_{una} \cup \mathcal{D}_{ap} \cup \mathcal{D}_{ss} \cup \mathcal{D}_{apB} \cup \mathcal{D}_{ssB} \cup \mathcal{D}_{ssD} \cup \mathcal{D}_{ssI}$
where,

1. Σ is the set of the foundational axioms of situation.
2. $\mathcal{D}_{S_0}$ is a set of axioms that defines the initial situation.
3. $\mathcal{D}_{una}$ is the set of unique names axioms for actions.
4. $\mathcal{D}_{ap}$ is the set of action precondition axioms. For each action symbol A, there is an axiom of the form $\mathbf{P_A}$ (See Section 2.1).
5. $\mathcal{D}_{ss}$ is the set of successor state axioms. For each relational fluent p, there is an axiom of the form $\mathbf{S_p}$ (See Section 2.1).
6. $\mathcal{D}_{apB}$ is the set of action precondition belief axioms. For each action symbol A and agent i, there is an axiom of the form $\mathbf{P'_{A_i}}$ (See Section 2.3).
7. $\mathcal{D}_{ssgB}$ is the set of successor generalised beliefs state axioms. For each relational fluent p and agent i, there are two axioms of the form $\mathbf{S_{B_i p(s')}}$ and $\mathbf{S_{B_i \neg p(s')}}$ (See Section 2.3).
8. $\mathcal{D}_{ssG}$ is the set of successor goal state axioms. For each relational fluent p and agent i, there are two axioms of the form $\mathbf{S_{G_i p}}$ and $\mathbf{S_{G_i \neg p}}$ (See Section 2.4).
9. $\mathcal{D}_{ssI}$ is the set of successor intention state axioms. For each relational fluent p and agent i, there are two axioms of the form $\mathbf{S_{I_i p}}$ and $\mathbf{S_{I_i \neg p}}$ (See Section 2.5).

The basic action theories defined in [11] consider only the first five sets of axioms. The right hand side in $\mathbf{P_A}$, $\mathbf{P'_{A_i}}$ and in the different successor state axioms must be a uniform formula in s in $\mathcal{L}_{sitcalc_{BDI}}$.[7]

3.1 Consistency Properties

For maintaining consistency in the representation of real world and mental states, the theory must satisfy the following properties:[8]
If ϕ is a relational or cognitive fluent, then

- $\mathcal{D} \models \forall\neg(\Upsilon^+_\phi \wedge \Upsilon^-_\phi)$.

If p is a relational fluent, i an agent and $\mathcal{M} \in \{B, G, I\}$, then

- $\mathcal{D} \models \forall\neg(\Upsilon^+_{\mathcal{M}_i p} \wedge \Upsilon^+_{\mathcal{M}_i \neg p})$
- $\mathcal{D} \models \forall(\mathcal{M}_i p(s) \wedge \Upsilon^+_{\mathcal{M}_i \neg p} \rightarrow \Upsilon^-_{\mathcal{M}_i p})$
- $\mathcal{D} \models \forall(\mathcal{M}_i \neg p(s) \wedge \Upsilon^+_{\mathcal{M}_i p} \rightarrow \Upsilon^-_{\mathcal{M}_i \neg p})$.

Other properties can be imposed in order to represent some definitions found in the literature. For example, the following properties:

- $\mathcal{D} \models \forall(B_i p(s) \vee \forall s'(s \sqsubset s' \rightarrow Bf_i \neg p(s', s)) \leftrightarrow \Upsilon^-_{G_i p})$
- $\mathcal{D} \models \forall(B_i \neg p(s) \vee \forall s'(s \sqsubset s' \rightarrow Bf_i p(s', s)) \leftrightarrow \Upsilon^-_{G_i \neg p})$:

characterize the notion of *fanatical commitment*: the agent maintains her goal until she believes either the goal is achieved or it is unachievable [6]. The following properties:

- $\mathcal{D} \models \forall(\Upsilon^+_{G_i p} \rightarrow \exists s' \; Bf_i p(s', s))$
- $\mathcal{D} \models \forall(\Upsilon^+_{G_i \neg p} \rightarrow \exists s' \; Bf_i \neg p(s', s))$

characterize the notion of *realism*: the agent adopts a goal that she believes to be achievable [6]. A deeper analysis of the properties that must be imposed in order to represent divers types of agents will be carried out in our future investigations.

3.2 Automated Reasoning

As a matter of simplification we assume that there are no communication actions. This assumption allows the representation of the generalised beliefs in terms of present beliefs as follows: $B_i p(s', s) \leftrightarrow B_i p(s')$.

Automated reasoning in the situation calculus is based on a regression mechanism that takes advantage of a regression operator. The operator is applied to a regressable formula. In particular, when the operator is applied to a regressable sentence, the regression operator produces a logically equivalent sentence whose only situation term is S_0.

---

[7] Intuitively, a formula is uniform in s iff it does not refer to the predicates $Poss$, $B_i Poss$ or $\sqsubset$, it does not quantify over variables of sort *situation*, it does not mention equality on situations, the only term of sort *situation* in the last position of the fluents is s.

[8] Here, we use the symbol $\forall$ to denote the universal closure of all the free variables in the scope of $\forall$. Also we omit the arguments (a, s) of the Υ's to enhance readability.

Definition 2. A formula W is *regressable* iff

1. Each situation used as argument in the atoms of W has syntactic form $do([\alpha_1, \ldots, \alpha_n], S_0)$, where $\alpha_1, \ldots, \alpha_n$ are terms of type *action*, for some $n \geq 0$.
2. For each atom of the form $Poss(\alpha, \sigma)$ mentioned in W, α has the form $A(t_1, \ldots, t_n)$ for some n-ary action function symbol A of $\mathcal{L}_{sitcalc_{BDI}}$.
3. For each atom of the form $B_i Poss(\alpha, \sigma'\sigma)$ mentioned in W, α has the form $A(t_1, \ldots, t_n)$ for some n-ary action function symbol A of $\mathcal{L}_{sitcalc_{BDI}}$.
4. W does not quantify over situations.

We extend the *regression operator* $\mathcal{R}$ defined in [15] with the following settings.
Let W be a regressable formula.

1. When W is an atom of the form $B_i Poss(A, \sigma'\sigma)$, whose action precondition belief axiom in $\mathcal{D}_{apB}$ is (P'_{A_i}),

$$\mathcal{R}[W] = \mathcal{R}[\Pi'_{A_i}(\sigma)]$$

2. When W is a cognitive fluent of the form $\mathcal{M}_i p(do(\alpha, \sigma))$, where $\mathcal{M} \in \{B, G, I\}$. If $\mathcal{M}_i p(do(a, s)) \leftrightarrow \Upsilon^{+}_{\mathcal{M}_i p}(a, s) \vee (\mathcal{M}_i p(s) \wedge \neg \Upsilon^{-}_{\mathcal{M}_i p}(a, s))$ is the associated successor state axiom in $\mathcal{D}_{ssgB} \cup \mathcal{D}_{ssG} \cup \mathcal{D}_{ssI}$,

$$\mathcal{R}[W] = \mathcal{R}[\Upsilon^{+}_{\mathcal{M}_i p}(\alpha, \sigma) \vee (\mathcal{M}_i p(\sigma) \wedge \neg \Upsilon^{-}_{\mathcal{M}_i p}(\alpha, \sigma))]$$

3. When W is a cognitive fluent of the form $\mathcal{M}_i \neg p(do(\alpha, \sigma))$, where $\mathcal{M} \in \{B, G, I\}$. If $\mathcal{M}_i \neg p(do(a, s)) \leftrightarrow \Upsilon^{+}_{\mathcal{M}_i \neg p}(a, s) \vee (\mathcal{M}_i \neg p(s) \wedge \neg \Upsilon^{-}_{\mathcal{M}_i \neg p}(a, s))$ is the associated successor state axiom in $\mathcal{D}_{ssgB} \cup \mathcal{D}_{ssG} \cup \mathcal{D}_{ssI}$,

$$\mathcal{R}[W] = \mathcal{R}[\Upsilon^{+}_{\mathcal{M}_i \neg p}(\alpha, \sigma) \vee (\mathcal{M}_i \neg p(\sigma) \wedge \neg \Upsilon^{-}_{\mathcal{M}_i \neg p}(\alpha, \sigma))]$$

Intuitively, these settings eliminates atoms involving $B_i Poss$ in favour of their definitions as given by action precondition belief axioms, and replaces cognitive fluent atoms about $do(\alpha, \sigma)$ by logically equivalent expressions about σ as given in their associated successor state axioms.

Note that $\mathbf{S_{I_i p}}$ is logically equivalent to $I_i p(T, do(a, s)) \leftrightarrow [(((a = commit(T) \wedge Bf_i Poss(do(T, s), s) \wedge Bf_i p(do(T, s), s)) \vee I_i p([a|T], s) \vee \Upsilon'^{+}_{I_i p}) \wedge G_i p(do(a, s))) \vee (I_i p(T, s) \wedge \neg \Upsilon'^{-}_{I_i p} \wedge G_i p(do(a, s)))]$, hence the successor intention state axioms, as well as every successor state axioms presented can be written in the standard format: $\phi(do(a, s)) \leftrightarrow \Upsilon^{+}_{\phi}(a, s) \vee (\phi(s) \wedge \neg \Upsilon^{-}_{\phi}(a, s))$.

For the purpose of proving W with background axioms $\mathcal{D}$, it is sufficient to prove $\mathcal{R}[W]$ with background axioms $\mathcal{D}_{S_0} \cup \mathcal{D}_{una}$. This result is justified by the following theorem:

Theorem 1. The Regression Theorem. *Let W be a regressable sentence of $\mathcal{L}_{sitcalc_{BDI}}$ that mentions no functional fluents, and let $\mathcal{D}$ be a basic intention theory, then*

$$\mathcal{D} \models W \quad \textit{iff} \quad \mathcal{D}_{S_0} \cup \mathcal{D}_{una} \models \mathcal{R}[W].$$

The proof is straightforward from the following theorems:

Theorem 2. The Relative Satisfiability Theorem. *A basic intention theory $\mathcal{D}$ is satisfiable iff $\mathcal{D}_{S_0} \cup \mathcal{D}_{una}$ is.*

The proof considers the construction of a model $\mathbb{M}$ of $\mathcal{D}$ from a model $\mathbb{M}_0$ of $\mathcal{D}_{S_0} \cup \mathcal{D}_{una}$. The proof is similar to the proof of Theorem 1 in [15].

Theorem 3. *Let W be a regressable formula of $\mathcal{L}_{sitcalc_{BDI}}$, and let $\mathcal{D}$ be a basic intention theory, then $\mathcal{R}[W]$ is a uniform formula in S_0. Moreover*

$$\mathcal{D} \models \forall (W \leftrightarrow \mathcal{R}[W]).$$

The proof is by induction based on the binary relation $\prec$ that has been defined in [15]. Since cognitive fluents can be viewed as ordinary situation calculus fluents, the proof is quite similar to the proof of Theorem 2 in [15].

The regression-based method introduced in [15] for computing whether a ground situation is executable can be employed to compute whether a ground situation is executable-believed. Moreover, the test is reduced to a theorem-proving task in the initial situation axioms together with action unique names axioms. Regression can also be used to consider the projection problem [11], i.e. answering queries of the form: Would G be true in the world resulting from the performance of a given sequence of actions T, $\mathcal{D} \models G(do(T, S_0))$? In our proposal, regression is used to consider projections of beliefs, i.e. answer queries of the form: Does i believe in s that p will hold in the world resulting from the performance of a given sequence of actions T, $\mathcal{D} \models Bf_i p(do(T, s), s)$?

As in [16], we make the assumption that the initial theory $\mathcal{D}_{S_0}$ is complete. The closed-world assumption about belief fluents characterizes the agent's lack of beliefs. For example, suppose there is only $B_r p(S_0)$ in $\mathcal{D}_{S_0}$ but we have two fluents $p(s)$ and $q(s)$, then under the closed-world assumption we have $\neg B_r q(S_0)$ and $\neg B_r \neg q(S_0)$, fact that represents the ignorance of r about q in S_0. Similarly, this assumption is used to represent the agent's lack of goals and intentions.

The notion of Knowledge-based programs [11] can be extend to BDI-based programs, i.e. Golog programs [16] that appeal to BDI notions as well as propositional attitude-producing actions. The evaluation of the programs is reducing to a task of theorem proving of sentence relative to a background intention theory. The Golog interpreter presented in [16] can be used to execute BDI-based programs due to the intention theories use the fluent representation to support beliefs,[9] goals and intentions.

[9] In Scherl and Levesque's approach [17], the notion that has been modelled is knowledge. Our interests to consider beliefs is motivated by our desire to avoid the logical omniscience problem.

4 A Planning Application

In this section we show the axiomatization for a simple robot. The goal of the robot is to attain a position x. To reach the goal, it can advance, reverse and remove obstacles. We consider two fluents: $p(x, s)$ meaning that the robot is in the position x in the situation s, and $o(x, s)$ meaning that there is an obstacle in the position x in the situation s. The successor state axiom of p is of the form:

$p(x, do(a, s)) \leftrightarrow [a = advance \wedge p(x-1, s)] \vee [a = reverse \wedge p(x+1, s)] \vee (p(x, s) \wedge \neg[a = advance \vee a = reverse])$.

Intuitively, the position of the robot is x in the situation that results from the performance of the action a from the situation s iff the robot was in $x - 1$ and a is advance or the robot was in $x + 1$ and a is reverse or the robot was in x and a is neither advance nor reverse.

Suppose that the robot's machinery updates its beliefs after the execution of *advance* and *reverse*, i.e. the assumption that the robot knows the law of evolution of p is made. So the successor belief state axioms are of the form:

$B_r p(x, do(a, s)) \leftrightarrow [a = advance \wedge B_r p(x-1, s) \vee a = reverse \wedge B_r p(x+1, s)] \vee B_r p(x, s) \wedge \neg[a = advance \vee a = reverse]$

$B_r \neg p(x, do(a, s)) \leftrightarrow [(a = advance \vee a = reverse) \wedge B_r p(x, s)] \vee B_r \neg p(x, s) \wedge \neg[a = advance \wedge B_r p(x-1, s) \vee a = reverse \wedge B_r p(x+1, s)]$.

The similarity between the successor state axiom of p and the successor belief state axiom of $B_r p$ represents formally this assumption. If initially the robot knows its position, we can show that the robot has true beliefs about its position in every situation $\forall s \forall x (B_r p(x, s) \rightarrow p(x, s))$.

Now if in addition we assume that there are no communication actions such as $communicate.p(x, s')$ which "sense" whether in s the position is/was/will be x in s', the successor generalised belief state axioms are of the form:

$B_r p(x, s', s) \leftrightarrow B_r p(x, s')$

$B_r \neg p(x, s', s) \leftrightarrow B_r \neg p(x, s')$

To represent the evolution of robot goals, we consider the two goal-producing actions: $adopt.p(x)$ and $adopt.not.p(x)$, whose effect is to adopt the goal to be in the position x and to adopt the goal to not be in the position x, respectively. Also we consider $abandon.p(x)$ and $abandon.not.p(x)$, whose effect is to give up the goal to be and not to be in the position x, respectively. The successor goal state axioms are of the form:

$$G_r p(x, do(a, s)) \leftrightarrow a = adopt.p(x) \vee G_r p(x, s) \wedge \neg(a = abandon.p(x))$$

$$G_r \neg p(x, do(a, s)) \leftrightarrow a = adopt.not.p(x) \vee G_r \neg p(x, s) \wedge \neg(a = abandon.not.p(x))$$

The successor intention state axioms are of the form:

$$I_r p(x, T, do(a, s)) \leftrightarrow G_r p(x, do(a, s)) \wedge [(a = commit(T) \wedge Bf_r Poss(do(T, s), s) \wedge Bf_r p(x, do(T, s), s)) \vee I_r p(x, [a|T], s) \vee I_r p(x, T, s) \wedge \neg(a = giveup(T))]$$

$$I_r \neg p(x, T, do(a, s)) \leftrightarrow G_r \neg p(x, do(a, s)) \wedge [(a = commit(T) \wedge Bf_r Poss(do(T, s), s) \wedge Bf_r \neg p(x, do(T, s), s)) \vee I_r \neg p(x, [a|T], s) \vee I_r \neg p(x, T, s) \wedge \neg(a = giveup(T))]$$

where the effect of action $giveup(T)$ is to give up the intention of carrying out T.

The successor state axiom of o is of the form:

$$o(x, do(a, s)) \leftrightarrow a = add\_obs \vee o(x, s) \wedge \neg(a = remove\_obs).$$

Intuitively, an obstacle is in x in the situation that results from the performance of the action a from the situation s iff a is *add_obs* or the obstacle was in x in s and a is not *remove_obs*. Suppose that the robot knows the law of evolution of o.

The plan generated by the robot can be obtained by answering queries of the form: What is the intention of the robot after it executes the action $commit(T)$ in order to satisfy its goal $I_r p(T, do(commit(T), S_0))$? For example, suppose that we have in the initial state the following information: $p(1, S_0)$, $o(3, S_0)$, $B_r p(1, S_0)$, $G_r p(4, S_0)$, i.e. the robot believes that its position is 1 and it wants attain 4 but it ignores that there is an obstacle in 3. The plan determined by it is [advance, advance, advance].

If the robot has incorrect information about the obstacle, for example $B_r o(2, S_0)$, the plan determined by it is [remove_obs, advance, advance, advance]. If the robot's beliefs corresponds to the real world, the robot determines a correct plan [advance, remove_obs, advance, advance].[10]

5 Conclusion

We have introduced intention theories in the framework of situation calculus. Moreover we have adapted the systematic, regression-based mechanism introduced by Reiter in order to consider formulas involving BDI. In the original approach, queries about hypothetical futures are answered by regressing them to equivalent queries solely about the initial situation. We used the mechanism to answer queries about the present beliefs of an agent about hypothetical futures by regressing them to equivalent queries solely about the initial situation. In the original approach, it is the designer (external observer, looking down on

[10] These plans have been automatically generated using SWI-Prolog.

the world) who knows the goal. In the proposal, it is the agent (internal element, interacting in the world) who knows the goal. Moreover, under certain conditions, the action sequence that represents a plan generated by the agent is obtained as a side-effect of successor intention state axioms.

The notions of belief-producing actions, goal-producing actions and intention-producing actions, namely propositional attitude-producing actions have been introduced just as Scherl and Levesque introduced knowledge-producing actions. The effect of propositional attitude-producing actions (such as sense, adopt, abandon, commit or give up) on mental state is similar in form to the effect of ordinary actions (such as advance or reverse) on relational fluents. Therefore, reasoning about this type of cognitive change is computationally no worse than reasoning about ordinary fluent change. Even if the framework presents strong restrictions on the expressive power of the cognitive part, the approach avoids complicating of the representation and updating of the world model. Diverse scenarios can be represented and implemented.

The notion of omniscience, the agent's beliefs correspond with the real world in every situation, can be represented under two assumptions: the agent knows the laws of evolution of the real world, and the agent knows the initial state of the world. In realistic situations, agents may have wrong beliefs about the evolution of world or initial state. Wrong beliefs can be represented by introducing successor belief axioms that do not correspond to successor state axioms, or by defining different initial settings between belief fluents and their corresponding fluents.

Acknowledgements

We are thankful to all the reviewers for their helpful observations. We are also grateful to Billy Duckworth, Mehmet Orgun and Robert Cambridge for their comments. This research is supported by a grant from the American Research Council.

References

1. Singh, M.P.: Multiagent Systems. A Theoretical Framework for Intentions, Know-How, and Communications. Springer LNAI 799 (1994)
2. Wooldridge, M.: Reasoning about Rational Agents. MIT Press (2000)
3. Singh, M.P., Rao, A., Georgeff, M.: Formal method in dai : Logic based representation and reasoning. In Weis, G., ed.: Introduction to Distributed Artificial Intelligence, New York, MIT Press (1998)
4. van Linder, B.: Modal Logics for Rational Agents. PhD thesis, University of Utrecht (1996)
5. Rao, A., Georgeff, M.: Modeling Rational Agents within a BDI Architecture. In: Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning, Morgan Kaufmann (1991)
6. Cohen, P.R., Levesque, H.J.: Intention is choice with commitment. Artificial Intelligence **42** (1990) 213–261

7. Rao, A.: Agentspeak(l): Bdi agents speak out in a logical computable language. In: Proceedings of the 7th European workshop on Modelling autonomous agents in a multi-agent world : agents breaking away, Springer-Verlag New York, Inc. (1996) 42–55
8. Dixon, C., Fisher, M., Bolotov, A.: Resolution in a logic of rational agency. In: Proceedings of the 14th European Conference on Artificial Intelligence (ECAI 2000), Berlin, Germany, IOS Press (2000)
9. Hustadt, U., Dixon, C., Schmidt, R., Fisher, M., Meyer, J.J., van der Hoek, W.: Verification within the KARO agent theory. Lecture Notes in Computer Science **1871** (2001)
10. Demolombe, R., Pozos Parra, P.: BDI architecture in the framework of Situation Calculus. In: Proc. of the Workshop on Cognitive Modeling of Agents and Multi-Agent Interactions at IJCAI, Acapulco, Mexico (2003)
11. Reiter, R. In: Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems. The MIT Press (2001)
12. Reiter, R.: The frame problem in the situation calculus: a simple solution (sometimes) and a completeness result for goal regression. In Lifschitz, V., ed.: Artificial Intelligence and Mathematical Theory of Computation: Papers in Honor of John McCarthy, Academic Press (1991) 359–380
13. Demolombe, R., Pozos Parra, P.: A simple and tractable extension of situation calculus to epistemic logic. In Ras, Z.W., Ohsuga, S., eds.: Proc. of 12th International Symposium ISMIS 2000, Springer. LNAI 1932 (2000)
14. Petrick, R., Levesque, H.: Knowledge equivalence in combined action theories. In: Proceeding 8th International Conference on Knowledge Representation and Reasoning. (2002)
15. Pirri, F., Reiter, R.: Some contributions to the metatheory of the situation calculus. Journal of the ACM **46** (1999) 325–361
16. Levesque, H., Reiter, R., Lespérance, Y., Lin, F., Scherl, R.: GOLOG: A Logic Programming Language for Dynamic Domains. Journal of Logic Programming **31** (1997) 59–84
17. Scherl, R., Levesque, H.: The Frame Problem and Knowledge Producing Actions. In: Proc. of the National Conference of Artificial Intelligence, AAAI Press (1993)