Integrating State Constraints and Obligations in Situation Calculus

Robert Demolombe
ONERA-Toulouse
2, Avenue Edouard Belin BP 4025,
31055 Toulouse Cedex 4, France.
Robert.Demolombe@cert.fr

Pilar Pozos Parra
Universidad Tecnológica de la Mixteca
K.M. 2.5 Carretera Huajuapan,
Oaxaca, C.P.69000. Mexico.
pilar@mixteco.utm.mx

Abstract

The ramification problem concerns the characterisation of indirect effects of actions. This problem arises when a theory of action is integrated with a set of state constraints. So integrating state constraints to a solution of the frame problem must deal with the ramification problem. In the situation calculus a general solution to both the frame and ramification problems has been proposed. This solution includes the indirect effects of actions in the successor state axioms. On the other hand, in the situation calculus, the notion of belief fluents has been introduced in order to distinguish between facts that hold in a situation and facts that are believed to hold in a situation. So apart from the traditional frame and ramification problems, a belief counterpart of these problems is considered. The successor belief state axioms were proposed to address the belief frame problem. Inspired in the mentioned approaches, we propose a general solution to the belief frame and ramification problems. We consider two sorts of constraints: the believed state constraints relating to physical laws and the believed mental constraints relating to social laws. Constraints imposed by social laws are well know in literature as obligations. Automated reasoning based on the proposal could easily be implemented in Prolog.

1. Introduction

The frame and ramification problems are classical problems that arise in reasoning about action using formal logic. The frame problem concerns the characterisation of facts that do not change when an action is performed. Reiter proposed a solution to the frame problem in [6]. He introduced the successor state axioms, which are obtained from (positive and negative) effect axioms. This solution based in the situation calculus can be very efficiently computed. However, in his solution, he does not consider the ramification problem, i.e. the axioms are obtained assuming that there are no state constraints. These constraints are a source of deep theoretical and practical difficulties in modeling dynamical systems [8].

Similarly to [3], McIlraith in [4] proposed an extension to successor state axioms that not only considers direct (positive and negative) effects of actions on fluents, but also indirect effects implicitly defined by a set of state constraints. Even when we have to accept limitations of expressiveness in the state constraints, this approach solve both the frame and ramification problems.

Another extension to the successor state axioms has been proposed in [2], the authors introduced the notion of belief fluents in order to distinguish between facts that hold in a situation and facts that are believed to hold in a situation. Performance of actions not only results in fluent changes, but contributes toward change in belief fluents as well. So we distinguish between the frame and ramification problems related with facts and the belief counterpart of these problems related with facts that are believed. In order to address the belief frame problem, the successor belief state axioms have been introduced. The expressive power of this solution is restricted to belief literals, however for a large class of applications, these restrictions are not real limitations. Automated reasoning could easily be implemented too.

In this paper, we propose a solution to the belief ramification problem. We consider two kind of constrains in the belief context: the believed state constraints relating to physical laws and the believed mental constraints relating to social laws. Constraints imposed by social laws (or normative systems) are well know in literature as obligations. The proposed solution is similar to McIlraith's approach in the sense that we extend the solution of the belief frame problem to include believed constraints. However, our approach is not restrained to stratified theories as McIlraith's approach. Moreover, we propose to include obligations in a way similar to that in which believed state constraints are included. Thus the successor belief state axioms are extended in order to include both state and mental constraints.

Automated reasoning could easily be implemented.

The rest of the paper is organized as follows: in the next section intuitive ideas concerning the representation of the world are presented through a simple example. Then, we give the logical solution to the problems concerning the representation of the world. Then, we give an example concerning belief representation, followed by the general logical solution for belief problems and finally the conclusions.

2 World representation

Situation calculus is a sort of classical first order logic that allows representation of dynamical worlds [8]. Situations represent sequences of actions which have been performed from the initial state to a current state. A situation is syntactically represented by a term of the form $do(a, s)$ where a denotes an action, and s denotes a situation. The initial situation is denoted by S_0. Predicates whose value may change from situation to situation are called *fluents*. The last argument of a fluent is a situation. For instance, $position(x, s)$ represents the fact that a given object is at the position x in the situation s. Action variables and situation variables can be quantified. For instance, $\neg\exists s(position(2, s))$ represents the fact that in no situation is a given object at position 2. Action quantification is an essential feature in the solution to the frame problem proposed by Reiter.

To intuitively present how the solution to the frame problem can be extended to include state constraints, we use the following scenario.

Let's consider a simple robot that can move forward (action adv) or backward (action rev) along a railtrack. Performance of actions adv or rev changes his position by one distance unit.

Indeed, the evolution of the fluents is defined by the successor state axioms. This axioms represent a solution to the frame problem as well. For each fluent F we have an axiom of the following form:

$$F(do(a, s)) \leftrightarrow \Upsilon_F^+(a, s) \vee F(s) \wedge \neg\Upsilon_F^-(a, s)$$

where Υ_F^+ characterises all the conditions that have positive effects on the fluent F, and Υ_F^- characterises all the conditions that have negative effects on the fluent F.

For example, for the fluent $position$, the positive effect of performing the action adv (respectively rev) when the robot is at the position $x - 1$ (respectively $x + 1$) in the situation s, is that it is at the position x in the situation $do(a, s)$. The positive effects on the position of a robot are represented by[1]:

$$\begin{aligned}&[a = adv \wedge position(x - 1, s) \vee \\ &a = rev \wedge position(x + 1, s)] \\ &\rightarrow position(x, do(a, s))\end{aligned} \tag{1}$$

The negative effect axiom expresses that if the robot is at the position x in the situation s and he performs either the action adv or the action rev, then in the situation $do(a, s)$ he is no longer at the position x:

$$\begin{aligned}&(a = adv \vee a = rev) \wedge position(x, s) \\ &\rightarrow \neg position(x, do(a, s))\end{aligned} \tag{2}$$

and the successor state axiom is:

$$\begin{aligned}&position(x, do(a, s)) \leftrightarrow [a = adv \wedge position(x-1, s) \\ &\vee a = rev \wedge position(x+1, s)] \vee position(x, s) \wedge \\ &\neg[(a = adv \vee a = rev) \wedge position(x, s)]\end{aligned}$$

This axiom represents the evolution of $position$ in terms of direct (positive and negative) effects of actions but the axiom does not consider the indirect effects of actions

State constraints represent invariant properties, that is properties that remain unchanged in every situation. For instance, the fact that the robot's position is unique is represented by the constraint:

$$\forall s, x, x'(position(x, s) \wedge position(x', s) \rightarrow x = x') \tag{3}$$

Let T be a theory and ψ a state constraint. To check if the constraint is satisfied in the theory, we have to prove that ψ is a consequence of T. That is we have to prove: $T \vdash \psi$. The formula ψ is assumed to be a standard formula of the situation calculus language.

McIlraith proposed considering state constraints as special cases of the conditions that make a fluent have the value "true" or "false". A practical advantage of this approach is that we can use the results proposed in [8] for automated reasoning. In what follows, we present a syntactic manipulation for including constraint (3) to the evolution of fluent $position$. From (3) we have:

$$position(x', do(a, s)) \wedge \neg(x = x') \rightarrow \neg position(x, do(a, s)) \tag{4}$$

Now, by applying properties of classical logic we obtain **all** the conditions that cause negative effects on the fluent from (2) and (4), and these conditions include the state constraint.

$$\begin{aligned}&[(a = adv \vee a = rev) \wedge position(x, s) \vee \\ &position(x', do(a, s)) \wedge \neg(x = x')] \\ &\rightarrow \neg position(x, do(a, s))\end{aligned} \tag{5}$$

If it is assumed that the conditions (1) and (5) are a complete representation of the conditions that may change the

[1] We adopt the convention that all the variables are universally quantified.

value of the fluent, we get the following successor state axiom (see [6] for details):

$$\begin{aligned}&position(x, do(a, s)) \leftrightarrow [a = adv \wedge position(x-1, s)\\&\vee a = rev \wedge position(x + 1, s)] \vee position(x, s) \wedge\\&\neg[(a = adv \vee a = rev) \wedge position(x, s) \vee\\&position(x', do(a, s)) \wedge \neg(x = x')]\end{aligned}$$

McIlraith called the result Intermediate Successor State Axioms. In order to obtain the Final Successor State Axioms, she proposed a stratified replacement of atoms expressed in terms of $do(a, s)$ for the corresponding right-hand side of the intermediate successor state axioms. The method is restrained to stratified theories in order to guarantee the termination of the process. The RHS of the final axioms are always expressed in terms of s. See [4] for a detailed description.

As we can see the example does not consider a stratified theory. The constraint (4) is not a stratified formula. The intent of replacing the atom $position(x', do(a, s))$ in and by the RHS of the intermediate successor state axiom of $position$ force the method to draw in a infinite cycle.

3 Extending successor state axioms

We propose a similar syntactic manipulation in order to extend Reiter's successor state axioms to include the state constrains. Let F be a fluent, $\Upsilon_F^+(a, s)$ all the conditions that have direct positive effect on the fluent F, and $\Upsilon_F^-(a, s)$ all the conditions that have direct negative effect on the fluent F. This Υ's are called the primary causes. Moreover, we assume that $\Upsilon_F^+(a, s)$ and $\Upsilon_F^-(a, s)$ are disjoint.

With a limited loss of generality, it is assumed that all the state constraints where F occurs can be represented as follows:

$$\begin{aligned}\psi(do(a, s)) &\rightarrow F(do(a, s))\\ \theta(do(a, s)) &\rightarrow \neg F(do(a, s))\end{aligned}$$

The initial theory must satisfy $\psi(S_0) \rightarrow F(S_0)$ and $\theta(S_0) \rightarrow \neg F(S_0)$. $\psi(do(a, s))$ and $\theta(do(a, s))$ are called the secondary causes and their effect over F depend of direct effects of the fluents appearing in them.

The first step is the following replacement: every fluent $\alpha(do(a, s))$ appearing in $\psi(do(a, s))$ or $\theta(do(a, s))$ is replaced with the condition $\Upsilon_\alpha^+(a, s)$, and the condition $\Upsilon_\alpha^-(a, s)$ substitutes any negation of fluent $\alpha(do(a, s))$ appearing in $\psi(do(a, s))$ or $\theta(do(a, s))$. So the resulting formulas $\psi'(a, s) = replacement(\psi(do(a, s)))$ and $\theta'(a, s) = replacement(\theta(do(a, s)))$ are expressed solely in terms of s. We have:

$$\begin{aligned}\psi'(a, s) &\rightarrow F(do(a, s))\\ \theta'(a, s) &\rightarrow \neg F(do(a, s))\end{aligned}$$

The substitution means intuitively that if the direct effects of actions trigger the condition $\psi'(a, s)$ or $\theta'(a, s)$ then the fluent changes due to an indirect effect. In this way the priority is given to the direct effects.

The next step considers the following causal completeness assumption: All the conditions under which an action a can lead, directly or indirectly, to fluent F becoming true or false in the successor state are characterised for $\Upsilon_F^+(a, s) \vee \psi'(a, s)$ and $\Upsilon_F^-(a, s) \vee \theta'(a, s)$, respectively. Then the new form of the successor state axiom that includes the state constraints is as follows:

$$\begin{gathered}F(do(a, s)) \leftrightarrow [\Upsilon_F^+(a, s) \vee \psi'(a, s)] \vee\\ F(s) \wedge \neg[\Upsilon_F^-(a, s) \vee \theta'(a, s)]\end{gathered}$$

This axiom may be understood as follows: F holds in $do(a, s)$ if and only if an action made it true or an indirect effect made it true or F was true in s and neither an action nor an indirect effect made it false.

It is assumed that for each fluent F, the positive and negative conditions are disjoint[2]:

$$\vdash T \rightarrow \forall\neg([\Upsilon_F^+(a, s) \vee \psi'(a, s)] \wedge [\Upsilon_F^-(a, s) \vee \theta'(a, s)])$$

4 Belief representation

To define the subjective representation of the evolution of the world, the language is extended with belief fluents of the form B_iF. We say that the "modalised" fluent B_iF holds in situation s if agent i believes that F holds in situation s and represent it as $B_iF(s)$. Similarly $B_i\neg F(s)$ represents the fact that the fluent $B_i\neg F$ holds in situation s: the agent i believes that F does not hold in situation s. Let's consider the robot's scenario. To represent the evolution of robot's beliefs, we have to consider four effect axioms for each fluent. For example, for the fluent $position(x, s)$, there are four distinct possible attitudes of the robot which are formally represented by: $B_rposition(x, s)$, $B_r\neg position(x, s)$, $\neg B_rposition(x, s)$ and $\neg B_r\neg position(x, s)$. Their corresponding axioms (6), (7), (8) and (9) are given below.

The effect of performing action adv (respectively rev) when the robot believes that he is at the position $x - 1$ (respectively $x + 1$) in the situation s is that he believes that he is at the position x in the situation $do(a, s)$. The same belief is produced by the following condition: he senses the position (action $sense$) in s and the real position is x:

$$\begin{aligned}&[a = adv \wedge B_rposition(x - 1, s) \vee\\&a = rev \wedge B_rposition(x + 1, s) \vee\\&a = sense \wedge position(x, s)]\\&\rightarrow B_rposition(x, do(a, s)) \qquad (6)\end{aligned}$$

[2] Here, we use the symbol $\forall$ to denote the universal closure of all the free variables in the scope of $\forall$.

Note that the two first conditions are examples of updates and the last one is a revision. The revision does not depend of a earlier epistemic state of the agent. Revisions make the connections with the real properties.

The effect of performing either action adv or rev when the robot believes that he is at the position x in the situation s is that in the situation $do(a, s)$ he does not longer believe that he is at the position x:

$$(a = adv \lor a = rev) \land B_r position(x, s) \\ \rightarrow \neg B_r position(x, do(a, s)) \tag{7}$$

We have two similar axioms defining, in the situation $do(a, s)$, the belief and unbelief attitudes of the robot with respect to the fact that he is not at the position x:

$$(a = adv \lor a = rev) \land B_r position(x, s) \\ \rightarrow B_r \neg position(x, do(a, s)) \tag{8}$$

$$[a = adv \land B_r position(x - 1, s) \lor \\ a = rev \land B_r position(x + 1, s) \lor \\ a = sense \land position(x, s)] \\ \rightarrow \neg B_r \neg position(x, do(a, s)) \tag{9}$$

If we extend the causal completeness assumptions to the robot's beliefs, we get, after some simplifications, the successor belief state axioms for the belief of robot concerning his position (see axioms (12) and (13) for the general form):

$$B_r position(x, do(a, s)) \leftrightarrow [a = adv \land \\ B_r position(x-1, s) \lor a = rev \land B_r position(x+1, s) \\ \lor a = sense \land position(x, s)] \lor B_r position(x, s) \land \\ \neg[(a = adv \lor a = rev) \land B_r position(x, s)] \tag{10}$$

$$B_r \neg position(x, do(a, s)) \leftrightarrow [(a = adv \lor a = rev) \land \\ B_r position(x, s)] \lor B_r \neg position(x, s) \land \neg[a = adv \land \\ B_r position(x-1, s) \lor a = rev \land B_r position(x+1, s) \\ \lor a = sense \land position(x, s)] \tag{11}$$

Notice that the definitions assume implicitly that if the robot performs the actions adv, rev or $sense$, he believes that he has performed these actions. However, the beliefs of an agent who does not know the performance of these actions can evolve in a different way. It is worth to notice that the approach allows to represent different evolutions of beliefs concerning the same fluent. Consider a pilot, if the only way for the pilot to be informed about the position is by checking it on a control panel (action $obs.position(x)$), then we have the following successor belief axioms for the pilot.

$$B_p position(x, do(a, s)) \leftrightarrow \\ a = obs.position(x) \lor B_p position(x, s)$$

$$B_p \neg position(x, do(a, s)) \leftrightarrow \\ B_p \neg position(x, s) \land \neg(a = obs.position(x))$$

In general for each agent i and each fluent F there are two axioms of the form:

$$B_i F(do(a, s)) \leftrightarrow \\ \gamma^+_{i_1,F}(a, s) \lor B_i F(s) \land \neg\gamma^-_{i_1,F}(a, s) \tag{12}$$

$$B_i \neg F(do(a, s)) \leftrightarrow \\ \gamma^+_{i_2,F}(a, s) \lor B_i \neg F(s) \land \neg\gamma^-_{i_2,F}(a, s) \tag{13}$$

5 Extending successor belief state axioms

We consider now the integration of invariant beliefs, that is integration of beliefs that remain unchanged in every situation. We include first the believed state constraints. These are invariant beliefs concerning physical laws. For instance, the fact that the robot believes that the position is unique is represented by the following believed state constraint:

$$\forall s, x, x' (B_r position(x, s) \land B_r position(x', s) \rightarrow x = x') \tag{14}$$

The procedure for including state constraints to the successor state axiom can be adapted to include believed state constraints to the successor belief state axioms.

Let $B_i F$ and $B_i \neg F$ be the belief fluents associated to the agent i and the fluent F, $\gamma^+_{i_1,F}(a, s)$ all the conditions that have direct positive effect on the fluent $B_i F$, $\gamma^-_{i_1,F}(a, s)$ all the conditions that have direct negative effect on the fluent $B_i F$, $\gamma^+_{i_2,F}(a, s)$ all the conditions that have direct positive effect on the fluent $B_i \neg F$, and $\gamma^-_{i_2,F}(a, s)$ all the conditions that have direct negative effect on the fluent $B_i \neg F$.

With a limited loss of generality, it is assumed that all the believed state constraints where $B_i F$ and $B_i \neg F$ occur can be represented as follows:

$$\begin{array}{rcl} \psi_1(do(a, s)) & \rightarrow & B_i F(do(a, s)) \\ \theta_1(do(a, s)) & \rightarrow & \neg B_i F(do(a, s)) \\ \psi_2(do(a, s)) & \rightarrow & B_i \neg F(do(a, s)) \\ \theta_2(do(a, s)) & \rightarrow & \neg B_i \neg F(do(a, s)) \end{array}$$

The initial theory must satisfy $\psi_1(S_0) \rightarrow B_i F(S_0)$, $\theta_1(S_0) \rightarrow \neg B_i F(S_0)$, $\psi_2(S_0) \rightarrow B_i \neg F(S_0)$, and $\theta_2(S_0) \rightarrow \neg B_i \neg F(S_0)$. The first step is to transform the formulas $\psi_1(do(a, s))$, $\theta_1(do(a, s))$, $\psi_2(do(a, s))$, and $\theta_2(do(a, s))$ in terms of direct effects over belief fluents. Every belief $B_i \alpha$ ($B_i \neg\alpha$, resp.) appearing in these formulas is replaced with $\gamma^+_{i_1,\alpha}(a, s)$ ($\gamma^+_{i_2,\alpha}(a, s)$, resp.) and every unbelief $\neg B_i \alpha$ ($\neg B_i \neg\alpha$, resp.) is replaced with $\gamma^-_{i_1,\alpha}(a, s)$ ($\gamma^-_{i_2,\alpha}(a, s)$, resp.) If $\psi'_1(a, s)$, $\theta'_1(a, s)$, $\psi'_2(a, s)$ and $\theta'_2(a, s)$ represent the result of the replacement in

$\psi_1(do(a,s))$, $\theta_1(do(a,s))$, $\psi_2(do(a,s))$ and $\theta_2(do(a,s))$, respectively, the believed state constrains are expressed as follows:

$$\begin{array}{rcl}\psi'_1(a,s) & \rightarrow & B_iF(do(a,s))\\ \theta'_1(a,s) & \rightarrow & \neg B_iF(do(a,s))\\ \psi'_2(a,s) & \rightarrow & B_i\neg F(do(a,s))\\ \theta'_2(a,s) & \rightarrow & \neg B_i\neg F(do(a,s))\end{array}$$

Next step considers the following causal completeness assumption: All the conditions under which an action a can lead, directly or indirectly, to belief fluent B_iF becoming true or false in the successor state are characterised for $\gamma^+_{i_1,F}(a,s) \vee \psi'_1(a,s)$ and $\gamma^-_{i_1,F}(a,s) \vee \theta'_1(a,s)$, respectively. And All the conditions under which an action a can lead, directly or indirectly, to belief fluent $B_i\neg F$ becoming true or false in the successor state are characterised for $\gamma^+_{i_2,F}(a,s) \vee \psi'_2(a,s)$ and $\gamma^-_{i_2,F}(a,s) \vee \theta'_2(a,s)$, respectively. Then the successor belief state axioms that include the believed state constraints take the following forms:

$$\begin{array}{c}B_iF(do(a,s)) \leftrightarrow [\gamma^+_{i_1,F}(a,s) \vee \psi'_1(a,s)] \vee\\ B_iF(s) \wedge \neg[\gamma^-_{i_1,F}(a,s) \vee \theta'_1(a,s)]\end{array}$$

$$\begin{array}{c}B_i\neg F(do(a,s)) \leftrightarrow [\gamma^+_{i_2,F}(a,s) \vee \psi'_2(a,s)] \vee\\ B_i\neg F(s) \wedge \neg[\gamma^-_{i_2,F}(a,s) \vee \theta'_2(a,s)]\end{array}$$

The first axiom may be understood as follows: i believes F holds in $do(a,s)$ if and only if an action made i believe F holds or an indirect effect made i believe F holds or i believes F was true in s and neither an action nor an indirect effect made disbelieve it. The second axiom has a similar interpretation.

In order to maintain consistent beliefs, it is assumed that the theory T meet the following conditions:

(P1) $\vdash T \rightarrow \forall\neg([\gamma^+_{i_1,F}(a,s) \vee \psi'_1(a,s)] \wedge [\gamma^-_{i_1,F}(a,s) \vee \theta'_1(a,s)])$

(P2) $\vdash T \rightarrow \forall\neg([\gamma^+_{i_2,F}(a,s) \vee \psi'_2(a,s)] \wedge [\gamma^-_{i_2,F}(a,s) \vee \theta'_2(a,s)])$

(P3) $\vdash T \rightarrow \forall\neg([\gamma^+_{i_1,F}(a,s) \vee \psi'_1(a,s)] \wedge [\gamma^+_{i_2,F}(a,s) \vee \psi'_2(a,s)])$

(P4) $\vdash T \rightarrow \forall(B_iF(s) \wedge [\gamma^+_{i_2,F}(a,s) \vee \psi'_2(a,s)] \rightarrow [\gamma^-_{i_1,F}(a,s) \vee \theta'_1(a,s)])$

(P5) $\vdash T \rightarrow \forall(B_i\neg F(s) \wedge [\gamma^+_{i_1,F}(a,s) \vee \psi'_1(a,s)] \rightarrow [\gamma^-_{i_2,F}(a,s) \vee \theta'_2(a,s)])$

We include now the believed mental constraints. These are invariant beliefs concerning social laws. Notice that state constraints cannot be physically violated. For instance, an object cannot be in two places at the same time (see, for example, (3)). However, in the case of mental constraints (social constraints or obligation), agents might violate them. Suppose the following formula: $\forall x, s(\mathit{trafficlight}(x, \mathit{red}, s) \rightarrow \neg \mathit{position}(x,s))$ represents a constraint, obviously the color red in a traffic light is not a physical barrier to reach the position of the corresponding crossroad. The fluent $\mathit{trafficlight}(x, \mathit{color}, s)$ means there is traffic light corresponding to a road intersection located at x in order to indicate when it is safe to advance in s using the following color code: if color is red, it is not safe to advance. If color is green, it is safe to advance. If color is yellow, it is a preventive sing saying that color is changing from green to red. So in order to keep save the advance, a normative system defines some mental barriers and suppose an ideal world where everybody obey the rules. The ideal world is defined through a set of obligations. For instance, we have the following obligation:

$$\forall x, s(\mathit{trafficlight}(x, \mathit{red}, s) \rightarrow O\neg \mathit{position}(x,s)) \quad (15)$$

which means if in the real world there is a crossroad with red traffic light then ideally the agent is not at the crossroad. Notice that the consequent does not represent a real fact neither a fact believed by an agent. $O\neg \mathit{position}(x,s)$ represents an ideal fact imposes by the normative system. Here, our interest does not concern the representation of ideal world evolution which must integrate a solution to the corresponding frame problem (see for example [1, 5]). Our concern is the design of obedient agents, i.e. agents respecting social constraints.

It is worth to notice that when a state constraints such as (3) is believed by an agent, the corresponding believed state constraint has the same schema (see (14)). However when an obligation as (15) is knew by an agent the mapping is not clear. Suppose a similar schema:

$$\forall x, s(B_r \mathit{trafficlight}(x, \mathit{red}, s) \rightarrow B_r\, O(\neg \mathit{position}(x,s))) \quad (16)$$

Where the belief fluent $B_r \mathit{trafficlight}(x, \mathit{color}, s)$ represents the robot's belief about the traffic light located at x. And $B_rO(\neg \mathit{position}(x,s)))$ represents the robot's belief about the obligation concerning the position. So the formula (16) represents the understanding of the obligation but this formula does not show an obedient attitude about a such obligation.

Our approach considers the fact that agents create habits by force of respect social rules. For example, some persons stop automatically when they see a red traffic light without thinking about the related social rule. Our approach propose to formalise this behaviour in order to assist a planner in the search of actions socially valid. So the believed mental constraint:

$$\forall x, a, s(B_r position(x, s) \wedge B_r trafficlight(x + 1, red, s) \\ \rightarrow B_r position(x, do(a, s)))$$

represents robot's custom in order to obey the rule (15). The intuitive meanings is that if the robot believes to be in front of a crossroad which has a red traffic light then wherever happens, he does not change his position (he stops). The interest in integration of obligations in this way concerns plan generation. In order to choose the adequate actions the agent must project his beliefs. If the agent want to avoid punishments, his projections must consider the satisfaction of social rules.

These constraints can be integrated to the successor belief state axioms in a similar manner to that in which believed state constraints were integrated.

Different behaviour can be defined through the believed mental constraints. For example, if we need a robot that tries to exit of a violation state when as soon as he is aware of such violation, we can define the following constraint:

$$\forall x, a, s(B_r position(x, s) \wedge B_r trafficlight(x, red, s) \\ \rightarrow B_r position(x + 1, do(a, s)))$$

meaning that if the robot believes itself to be at a crossroad with a red traffic light, he will try to move to the next position in order to pass the crossroad.

Thus we obtain the extended successor belief state axiom for $B_r position$ which include both believed state and mental constraints:

$$B_r position(x, do(a, s)) \leftrightarrow [a = adv \wedge \\ B_r position(x-1, s) \vee a = rev \wedge B_r position(x+1, s) \\ \vee a = sense \wedge position(x, s) \vee B_r position(x, s) \wedge \\ B_r trafficlight(x+1, red, s)] \vee B_r position(x, s) \wedge \\ \neg[(a = adv \vee a = rev) \wedge B_r position(x, s) \vee (a = adv \wedge \\ B_r position(y-1, s) \vee a = rev \wedge B_r position(y+1, s) \\ \vee a = sense \wedge position(y, s)) \wedge \neg y = x]$$

Notice that this axiom does not satisfy (P1) so the constraint is not captured. A solution to this problem is the introduction of preferences. For the sake of space we do not present the details.

6 Conclusion

We have presented a general solution to solve the frame and ramification problems in the belief context. A significant difference between this approach and that presented in [7] is that it can be easily implemented and the proof by induction is restricted to two steps (just as for mathematical induction): verification in the initial state and verification of the successor state. Constraints that are included in the successor belief state axioms could influence the agent to take a decision. For example, if a robot believes it cannot be stopped at a crossroad with a red traffic light, then it will reject the goal of having to wait for someone at the crossroad. The regression theorem presented in [8] justifies a simple Prolog implementation for reasoning with this proposal assuming a closed initial database.

References

[1] R. Demolombe and A. Herzig. Obligation change in dependence logic and situation calculus. In *Proc. of the 7th International Workshop on Deontic Logic in Computer Science*, 2004.

[2] R. Demolombe and P. Pozos Parra. A simple and tractable extension of situation calculus to epistemic logic. In Z. W. Ras and S. Ohsuga, editors, *Proc. of 12th International Symposium ISMIS 2000*. Springer. LNAI 1932, 2000.

[3] F. Lin and R. Reiter. State constraints revisited. *The journal of Logic and Computation. Special issue on action and processes*, 4:655–678, 1994.

[4] S. McIlraith. Integrating action and state constraints: A closed-form solution to the ramification problem (sometimes). *Artificial Intelligence*, 116, 2000.

[5] P. Pozos Parra, A. Nayak, and R. Demolombe. Theories of intentions in the framework of situation calculus. In P. Torroni P. Yolum J. Leite, A. Omicini, editor, *Declarative Agent Languages and Technologies*. Springer Verlag, 2005.

[6] R. Reiter. The frame problem in the situation calculus: a simple solution (sometimes) and a completeness result for goal regression. In V. Lifschitz, editor, *Artificial Intelligence and Mathematical Theory of Computation: Papers in Honor of John McCarthy*, pages 359–380. Academic Press, 1991.

[7] R. Reiter. Proving Properties of States in the Situation Calculus. *Artificial Intelligence*, 64:337–351, 1993.

[8] R. Reiter. *Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems*. The MIT Press, 2001.