Action, Belief Change and the Frame Problem: A Fluent Calculus Approach

Richard B. Scherl
Computer Science Department
Monmouth University
West Long Brach, NJ 07764
U.S.A.
email: rscherl@monmouth.edu

Abstract

This paper develops a solution to the analogue of the frame problem that arises when the belief state of an agent is axiomatized in the presence of belief changing actions. It follows the work of Scherl and Levesque which adapted the approach to the frame problem of Reiter to the case of the analogue of the frame problem that arises when knowledge and knowledge producing actions are added to the situation calculus. For the case of belief, it is necessary to use the somewhat more expressive relative of the situation calculus, the fluent calculus which is a formalism that allows quantification over states and fluents.

1 Introduction

The frame problem was noted early on in the study of formalizing actions and their effects on the world [MH69]. In work in this area, axioms are used to specify the prerequisites of actions as well as their effects, that is, the fluents that they change. As noded in [MH69], it is in general also necessary to provide frame axioms to specify which fluents remain unchanged by the actions. Reiter [Rei91; Rei01] has given a set of conditions under which the explicit specification of frame axioms can be avoided.

In [SL03], this solution to the frame problem was extended to cover *knowledge-producing actions*, that is, actions whose effects are to change a state of knowledge. To incorporate knowledge-producing actions like these into the situation calculus, it is necessary to treat knowledge as a fluent that can be affected by actions. This is precisely the approach taken by Moore [Moo80]. With the presence of knowledge, there emerges a new analogue to the frame problem. It is necessary to ensure that after an action has taken place (whether it be a sensing or a non-sensing action), there are no unwanted losses or gains in know.edge.

In [SL03], knowledge and knowledge-producing actions are handled in a way that avoids this extended frame problem: they are able to prove as a consequence of their specification that knowledge-producing actions do not affect fluents other than the knowledge fluent, and that actions that are not knowledge-producing only affect the knowledge fluent as appropriate. In addition, they show that *memory* emerges as a side-effect: if something is known in a certain situation, it remains known at successor situations, unless something relevant has changed.

But this work only considers knowledge and knowledge producing actions. That is, it is assumed that the agent's beliefs and sensor results are all correct. The approach simply fails when the agent being modeled acquires information that contradicts its knowledge. The agent then knows all sentences of the language since there will be no accessible possible worlds/situations as all accessible worlds/situations in which the new piece of information is false are eliminated. It is clearly unrealistic given the goals of *cognitive robots* to limit attention to agents who begin with only correct beliefs about the world.

What is needed is the incorporation of some sort of belief revision into the framework. In [SPLL00] the model of [SL93; SL03] is extended to include a process of belief revision. Additionally, [JT04] extend the closely related fluent calculus to incorporate belief revision. But they do not address the solution to the analogue of the frame problem that arises with belief.

This paper addresses the problem of dealing with these unwanted changes in belief when a new fact comes to be believed. In order to do this, we need greater expressivity than is allowed in the situation calculus. We need the ability to quantify over fluents and states. As this is allowed by the fluent calculus [Thi98; Thi00], the fluent calculus is utilized in this paper. In this paper, we limit our attention to knowledge of sentences in a propositional language.

Like [SPLL00] and [JT04], the approach developed here is able to incorporate both revision and update [KM91a] into a uniform framework. Changes in belief due to the incorporation of new information through sensing respect the revision postulates. Changes in belief due to changes in the world respect the update postulates. But the main contribution in this paper is an investigation of the analogue of the frame problem when belief is incorporated into situation/fluent calculus action theories.

2 The Fluent Calculus: A Language for Specifying Dynamics

The fluent calculus[Thi98; Thi00] is a many-sorted language with the sorts *action*, *sit*, *fluent*, and *state*. Fluents are reified.

In other words, they are terms. States are also terms which are constructed out of fluents with the binary function symbol $\circ$. In the following, the letter *f* is used for fluent variables, the letter *z* for state variables, *s* for situation variables, and *a* for action variables. These letters may have superscripts or subscripts. Unless otherwise indicated, all variables in formulas are universally quantified.

Unlike the situation calculus, the fluent calculus separates the notion of state and situation. In the fluent calculus, situations contain a history of actions that have been performed and states contain the fluents that hold in that state. Each situation has an associated state.

The function symbol $\circ$, used to construct the state terms, is axiomatized to be associative, commutative, and has a unit element $\emptyset$. The following set of axioms (AC1) ensures these properties:

$$(z_1 \circ z_2) \circ z_2 = z_1 \circ (z_2 \circ z_3)$$
$$z_1 \circ z_2 = z_2 \circ z_1$$
$$z \circ \emptyset = z$$

Additionally, unique name axioms for state terms are needed. These (EUNA) are given below:

$$z = f \rightarrow z \neq \emptyset \wedge [z = z' \circ z'' \rightarrow z' = \emptyset \vee z'' = \emptyset]$$
$$z_1 \circ z_2 = z_3 \circ z_4 \rightarrow (\exists z_a, z_b, z_c, z_d)[z_1 = z_a \circ z_b \wedge z_2 = z_c \circ z_d \wedge z_3 = z_a \circ z_c \wedge z_4 = z_b \circ z_d]$$

We also have the foundational axiom

$$\text{STATE}(s) \neq f \circ f \circ z$$

which prohibits double occurrences of fluents in states. Additionally, we have the following abbreviations:

$$\mathbf{Holds}(f, s) \stackrel{\text{def}}{=} \mathbf{Holds}(f, \text{STATE}(s))$$
$$\mathbf{Holds}(f, z) \stackrel{\text{def}}{=} (\exists z')\, z = f \circ z'$$
$$\mathbf{Holds}(\neg\varphi, z) \stackrel{\text{def}}{=} \neg\mathbf{Holds}(\varphi, z)$$
$$\mathbf{Holds}(\varphi \wedge \phi, z) \stackrel{\text{def}}{=} \mathbf{Holds}(\varphi, z) \wedge \mathbf{Holds}(\phi, z)$$

We require that for each action $\text{A}(\vec{x})$, there is a precondition axiom of the form,

$$\text{POSS}(\text{A}(\vec{x}), s) \equiv \pi(\vec{x}, s) \qquad (1)$$

Additionally, state update axioms are needed to specify the relationship between states at two consecutive situations. Below is the general form :

$$\text{POSS}(A(\vec{x}), s) \rightarrow \text{STATE}(\text{DO}(\text{A}(\vec{x}), s)) \circ \text{v}^- = \text{STATE}(s) \circ \text{v}^+ \qquad (2)$$

Here v^- are the negative effects and v^+ are the positive effects of action A. An example is:

$$\text{POSS}(\text{OPEN}(\text{DOOR}_1), s) \rightarrow \text{STATE}(\text{DO}(\text{OPEN}(\text{DOOR}_1), s)) = \text{STATE}(s) \circ \text{CLOSED}(\text{DOOR}_1) \qquad (3)$$

After the execution of an open action, the door is no longer closed. It has been shown [Thi99] that a collection of state updates in this form constitute a solution to the frame problem.

3 Adding Belief to the Fluent Calculus

For belief we can adapt some of the machinery [Thi98; Thi00] developed for the case of knowledge. We have a predicate BSTATE of type $sit \times state$ indicating that the second argument is a possible state of the situation in the first argument. Intuitively, something is believed in a situation if it holds in each of the belief states associated with that situation. We need an axiom similar to the foundational axiom given earlier:

$$\text{BSTATE}(s, z) \rightarrow \forall f, z'\ z \neq f \circ f \circ z' \qquad (4)$$

$$\textbf{Believes}(\varphi, s) \stackrel{\text{def}}{=} (\forall z)\text{BSTATE}(s, z) \rightarrow \mathbf{Holds}(\varphi, z) \qquad (5)$$

where **Holds** is as defined previously.

Belief in the initial situation can easily be specified as follows:

$$\mathbf{Believes}(\text{P}, \text{S}_0) \qquad \mathbf{Believes}(\neg\text{Q}, \text{S}_0)$$

We want to model actions that provide the agent information about the state of the world. For example, we might imagine a SENSE_P action for a fluent P, such that after doing a SENSE_P, the truth value of P is believed. We introduce the notation **Bwhether**(P, s) as an abbreviation for a formula indicating that the truth of a fluent P is known (in the sense of belief) by the agent.

$$\textbf{Bwhether}(\text{P}, s) \stackrel{\text{def}}{=} \textbf{Believes}(\text{P}, s) \vee \textbf{Believes}(\neg\text{P}, s),$$

Certainly, the effect of a SENSE_P action is **Bwhether**(P, s).

The next step is to correctly axiomatize changes in the belief accessible states. The issue is what is the relationship between the states (*z*) for which $\text{BSTATE}(s, z)$ is true and the set z' for which $\text{BSTATE}(\text{DO}(a, s), z')$ is true. We might continue to follow [Thi98; Thi00] and develop belief update axioms of the form:

$$\textbf{Bstate}(\text{DO}(a, s), z) \equiv \exists z'\ (\mathbf{Bstate}(s, z') \wedge \Psi(z, z', s)) \qquad (6)$$

Here Ψ is a first-order formula expressing the relation between the two sets of belief states. The following is an example:

$$\text{POSS}(\text{SENSE}_P, s) \rightarrow \textbf{Bstate}(\text{DO}(\text{SENSE}_P, s), z) \equiv \mathbf{Bstate}(s, z) \wedge [\mathbf{Holds}(\text{P}, z) \equiv \mathbf{Holds}(\text{P}, s)] \qquad (7)$$

But the problem here is that the agent may already believe that $\neg$P holds and then there will not be a z' such that $\mathbf{Bstate}(\text{DO}(\text{SENSE}_P, s), z')$. Then the agent's beliefs will be in a state of contradiction as for any proposition Q, both

$$\mathbf{Believes}(\neg\text{Q}, \text{DO}(\text{SENSE}_P, s))$$

and

$$\mathbf{Believes}(\text{Q}, \text{DO}(\text{SENSE}_P, s))$$

will hold. Revision must occur to prevent the agent from believing falsity.

4 Axiomatizing Changes in Belief

Here a successor state axiom is developed for specifying the belief set (i.e., those z such that **Bstate**(DO(a,s),z) holds) in terms of the belief set at the previous situation (i.e., those z' such that **Bstate**(s, z') holds), the action a and the result of the sensing (if the action was a sensing action).

It is necessary to distinguish between 3 possible cases.

- The action was not a sensing action.
- The action was a sensing action and the result did not contradict the previous beliefs.
- The action was a sensing action and the result did contradict the previous beliefs.

To simplify matters, following [SL03], all actions will be either pure sensing actions that do not alter the world or ordinary actions that only alter the world and do not provide any information to the agent beyond the fact that the action has occurred. It is necessary to require that the axiomatization correctly distinguishes between sensing (information-producing actions) and ordinary actions that alter the state of the world. For every action a, the axiomatization must entail either that $\text{TYPE}(a) =$ "SENSE" or $\text{TYPE}(a) \neq$ "SENSE".

The successor state axiom for Bstate requires some additional machinery as well. In general, there may be many information-producing actions, as well as many ordinary actions. To characterize all of these, we have a predicate SR (for sensing result), and for each action α, a sensing-result axiom of the form:

$$\text{SR}(\alpha, s, z) \;\equiv\; \phi_\alpha(s, z) \tag{8}$$

The following is an SR axiomatization for an action that determines accurately whether or not P is true in the current state.

$$\text{SR}(\text{SENSE}_{\text{P}}, s, z) \;\equiv\; (\textbf{Holds}(\text{P}, z) \wedge \textbf{Holds}(\text{P}, s)) \tag{9}$$

Since the situation is also an argument to SR, it is possible to axiomatize functions for sensors that are not accurate, but rather give different results depending on the situation; regardless of the current state. For example:

$$\text{SR}(\text{SENSE}_{\text{P}}, s, z) \;\equiv\; (\exists a, b\; s = \text{DO}(a, \text{DO}(b, \text{S}_o)) \wedge \textbf{Holds}(\text{P}, z)) \tag{10}$$

The idea is that if SENSE_P is the third action to occur from the beginning of the history, the result of the sensing will be that P holds regardless of whether it actually does. There are many other possibilties. But this paper is primarily concerned with accurate sensors.

The SR axiom for ordinary (non-sensing) actions are all a default with true for the $\phi_\alpha(s, z)$. For example

$$\text{SR}(\text{PICKUP}, s, z) \;\equiv\; \text{TRUE} \tag{11}$$

For ordinary actions, we need to have a correctly axiomatized state update function SUF of the following form:

$$\text{SUF}(\text{PICKUP}(\text{obj}_1), z) = z' \;\equiv\; z' \circ z_2 \;=\; z \circ z_1 \tag{12}$$

Consider the following two examples:

$$\text{SUF}(\text{PICKUP}(\text{obj}_1), z) = z' \;\equiv\; z' \;=\; z \circ \text{HOLDING}(\text{obj}_1) \tag{13}$$

$$\text{SUF}(\text{PUTDOWN}(\text{obj}_1), z) = z' \;\equiv\; r \circ \text{HOLDING}(\text{obj}_1) \;=\; z \tag{14}$$

For sensing actions the SUF function needs to have no effect on the state and therefore the right hand side of the equivalence needs to be $z = z'$ indicating that sensing actions have no effect on the world. For example:

$$\text{SUF}(\text{SENSE}, z) = z' \;\equiv\; z = z' \tag{15}$$

If the result of sensing does not contradict the agent's previous beliefs, then it is necessary to perform update. In this case the result is similar to that of [SL03]. But the complicated case is when the sensing contradicts the agent's previously held beliefs. In this case revision must occur.

Here an ordering on states is needed. Peppas, Foo, and Nayak [PFN00] develop a domain-independent criterion for measuring the similarity between two alternative belief states called PMA (Possible Models Approach) since it is based upon the Possible Models Approach for reasoning about actions[Win88]. The criterion of similarity is based upon the literals which are true in each model or state. Given two states w and r, Diff(w, r) is the symmetric difference of the literals true in w and r. This criterion is essentially that for a given state w, a state r is more similar to w than r' if Diff$(w, r) \subset$ Diff(w, r'). See also [Dal88] and [KM91b].

Peppas, Foo, and Nayak [PFN00] follow Grove[Gro88] and imagine a system of spheres interpreted as a plausibility measure. Similarity is interpreted as differences in the truth of fluents. We imagine that there is a system of spheres centered around each possible world (state). Given a system of spheres (S) centered around w for any possible world r, the smaller Diff(w, r) is, the closer r is to the center, i.e., to w.

In other words, given any two models or worlds r and r', if Diff$(w, r) \subset$ Diff(w, r') then there is a sphere $U \in S$ that contains r and not r'. Following, Grove when we want to revise a theory by φ, the new theory is determined by the most plausible worlds satisfying φ. The new worlds are precisely those in the sphere closest to the center that has worlds in which φ is true.

The proposition that state z^* is more similar to z' than to z is to z' (Diff$(z, z^*) \subset$ Diff(z, z')) is expressed by the following formula:

$$\forall f\; \textbf{Holds}(f, z') \not\equiv \textbf{Holds}(f, z^*) \rightarrow \textbf{Holds}(f, z) \not\equiv \textbf{Holds}(f, z') \tag{16}$$

The formula states that every fluent in the symmetric difference of z' and z^* is also in the symmetric difference of z and z'.

All of these notions are then incorporated into the successor state axiom for BSTATE given below:

Successor State Axiom for Bstate

$$\begin{aligned}
&\forall z\ \textbf{Bstate}(\textsc{do}(a,s),z) \equiv \\
&\quad (\textsc{type}(a) \neq \text{“}\textsc{sense}\text{”} \wedge \\
&\qquad \exists z'\ \textbf{Bstate}(s,z')\ \wedge \textsc{suf}(a,z') = z) \\
&\qquad\qquad \vee \\
&\quad [\textsc{type}(a) = \text{“}\textsc{sense}\text{”} \wedge \\
&\qquad (\textsc{Poss}(a,z) \wedge\ \textsc{sr}(a,s,z)\ \wedge \textbf{Bstate}(s,z)) \\
&\qquad\qquad \vee \\
&\qquad (\neg(\exists z' \textbf{Bstate}(s,z')\ \wedge\ \textsc{Poss}(a,z') \wedge \\
&\qquad\qquad \textsc{sr}(a,s,z')) \wedge \\
&\qquad (\exists\, z'\ \textbf{Bstate}(s,z')\ \wedge\ \textsc{Poss}(a,z) \wedge\ \textsc{sr}(a,s,z) \wedge \\
&\qquad\quad \neg\exists z^*(\textsc{sr}(a,s,z^*)\ \wedge z^* \neq z\ \wedge\ \textsc{Poss}(a,z^*) \wedge \\
&\qquad\qquad \forall f \textbf{Holds}(f,z') \not\equiv \textbf{Holds}(f,z^*) \rightarrow \\
&\qquad\qquad \textbf{Holds}(f,z) \not\equiv \textbf{Holds}(f,z'))))]
\end{aligned} \tag{17}$$

The idea here is that either the action is a sensing type action (i.e., an action that alters the state of belief, but not the world) or not a sensing type action (i.e., an ordinary action, one that alters the state of the world). If the action is an ordinary action then each belief accessible state must be updated in exactly the same way that the state associated with the situation is updated. This case works exactly as [SL03; Thi00] handle the combination of knowledge and ordinary actions. It is the first disjunct of the successor state axiom.

If the action is a sensing action then either the result of the sensing action contradicts the current state of knowledge or it does not. If it does not, then there must be at least one belief accessible state consistent with the result of the sensing action. This is the second disjunct of the successor state axiom. The belief accessible states accessible after the sensing action are precisely those which were accessible prior to the action and are consistent with the result of the sensing action. In this case things work very much as in the case of knowledge, although there is no guarantee that the fluents true in all of the belief accessible states are in fact true in the actual state. This case works exactly as [SL03] handles the combination of knowledge and sensing actions. It differes from [Thi00] in that the agent does know all of the axiomatized effects of actions.

Now, it very well may be the case that the result of the sensing action does contradict the current state of knowledge. If it does, then there will not be any belief accessible states consistent with the result of the sensing action. Then the belief accessible states are those states which are both consistent with the result of the sensing action and are minimally close to a state which was belief accessible prior to the sensing action. The minimal closeness is handled by the the third disjunct. This ensures that if z is in the new **Bstate**, then there is no other belief state z^* in which $\textsc{sr}(a,s,z^*)$ holds, but which differs in fewer fluents than z from a z' in the initial belief state. Note that the z' in the initial belief state can not be in the new belief state because of the fact that

$$\neg(\exists z' \textbf{Bstate}(s,z')\ \wedge\ \textsc{Poss}(a,z') \wedge\ \textsc{sr}(a,s,z'))$$

holds. But it must be the case that for every z in the new belief state, there is a z' in the original belief state to which it is minimally close.

5 Example

An axiomatization of a domain needs the axioms AC1, EUNA, the foundational axiom for STATE, the foundational axiom for BSTATE, the successor state axiom for BSTATE (17), the abbreviations for **Holds** and **Believes**, the axiomatization of the initial situation, and for each action a precondition axiom, and a state update axiom, a SR axiom, and a SUF axiom. This set is called $\mathcal{A}$.

The following example is taken from [SPLL00]. There are two rooms R_1 and R_2. The agent has one sensor which detects whether or not the light is on in the room in which the agent is located. The other sensor indicates whether or not the agent is in R_1. The fluents $\textsc{Light}_1$, $\textsc{Light}_2$ indicating that the lights are on in rooms 1 and 2, and also $\textsc{InR}_1$ indicating that the agent is in room 1. If $\neg\textsc{InR}_1$ holds, then the agent is in room 2.

Initially, the lights in both rooms are on and the agent is in R_1. The agent believes that $\neg\textsc{Light}_1$ and that $\textsc{InR}_1$ both hold. We have:

$$\begin{aligned}
&\textbf{Holds}(\textsc{InR}_1, \mathrm{S}_0)\ \wedge\ \textbf{Holds}(\textsc{Light}_1, \mathrm{S}_0) \\
&\textbf{Holds}(\textsc{Light}_2, \mathrm{S}_0) \\
&\textbf{Believes}(\neg\textsc{Light}_1, \mathrm{S}_0)\ \wedge\ \textbf{Believes}(\textsc{InR}_1, \mathrm{S}_0)
\end{aligned}$$

We also need the SR axiomatization of the sense action.

$$\begin{aligned}
&\textbf{Holds}(\textsc{InR}_1, s)\ \rightarrow \\
&\quad \textsc{sr}(\textsc{sense}_{\mathrm{LIGHT}}, s, z)\ \equiv\ \textbf{Holds}(\textsc{Light}_1, z)
\end{aligned} \tag{18}$$

It follows that:

$$\textbf{Believes}(\textsc{Light}_1, \textsc{do}(\textsc{sense}_{\mathrm{LIGHT}}, \mathrm{S}_0))$$

This was a case of revision.

The agent also has the capability of moving from one room to another with the LEAVE action. The following is the state update axiom for this action.

$$\begin{aligned}
&\textsc{Poss}(\textsc{leave}, s) \rightarrow \\
&\quad \textbf{Holds}(\textsc{InR}_1, s) \wedge \\
&\qquad \textsc{State}(\textsc{do}(\textsc{leave}, s)) = \textsc{State}(s) - \textsc{InR}_1 \vee \\
&\quad \neg\textbf{Holds}(\textsc{InR}_1, s) \wedge \\
&\qquad \textsc{State}(\textsc{do}(\textsc{leave}, s)) = \textsc{State}(s) + \textsc{InR}_1
\end{aligned} \tag{19}$$

The following is the SUF axiom:

$$\begin{aligned}
&\textbf{Holds}(\textsc{InR}_1, z)\ \rightarrow \\
&\quad \textsc{suf}(\textsc{leave}, z) = z'\ \equiv z'\ =\ z \circ \textsc{InR}_1\ \wedge \\
&\textbf{Holds}(\neg\textsc{InR}_1, z)\ \rightarrow \\
&\quad \textsc{suf}(\textsc{leave}, z) = z'\ \equiv z'\ =\ z - \textsc{InR}_1
\end{aligned} \tag{20}$$

The same information is repeated as it is needed in this form for the successor state axiom for Bstate (17). It follows that:

$$\textbf{Believes}(\textsc{Light}_2, \textsc{do}(\textsc{sense}_{\mathrm{LIGHT}}, \textsc{do}(\textsc{leave}, \mathrm{S}_0)))$$

This was a case of update.

6 Properties of the Result

In general, when a sensing action takes place, the result respects the AGM [AGM85; Gar88] postulates for revision. Additionally, when a world changing action occurs, the change in belief respects the postulates of Katsuno and

Mendelzon (KM) for update [KM91a]. The notation $\mathcal{B}_s$ is used to represent the set of sentences believed by the agent at situation s.

$$\mathcal{B}_s = \{\varphi \mid \mathcal{A} \models \textbf{Believes}(\varphi, s)\} \tag{21}$$

To make the comparison with the revision/update postulates, it is necessary (following [SPLL00]) to equate a belief set (of the AGM theory) or a knowledge base (of KM) with $\mathcal{B}_s$. Katsuno and Mendelzon (KM)[KM91a] have distinguished between update and revision by stating the AGM postulates for revision as postulates R1–R6 and their postulates for update as U1–U8.

Theorem 1 (KM Postulates) *When $\mathcal{B}_s$ is viewed as a knowledge base, an axiomatization $\mathcal{A}$ conforms to postulates U1–U4 when update occurs and R1–R4 when revision occurs.*

Space does not permit a detailed discussion of all of the postulates for revision and update [KM91a] and a comparison of the properties of the approach described here with the approachs of Shapiro etal. [SPLL00] and the approach of Jin and Thielscher[JT04].

The most important results here are that changes in belief are minimal in the sense of the analogue of the frame problem for belief. There are no unnecessary increases in things believed and decreases in things believed.

First note that for each action, there must be a formula Π_α such that the axiomatization entails

$$\forall s \ \textsc{Poss}(\alpha, s) \rightarrow \textbf{Holds}(\Pi_\alpha, s)$$

We call the formula Π_α, the action precondition formula for action α.

For every sensing action α and situation s, there must be a formula Σ^s_α such that the axiomatization entails

$$\textsc{sr}(\alpha, s, \textsc{state}(s)) \rightarrow \textbf{Holds}(\Sigma^s_\alpha, s)$$

We call the formula Σ^s_α, the sensed formula for α. For non-sensing actions $\Sigma^s_\alpha \equiv T$.

The statement of the theorems to follow requires an additional definition based on $\mathcal{B}_s$:

$$\mathcal{B}^{-\varphi}_s = \mathcal{B}_s - \{\varphi_1 \ldots \varphi_n\} \tag{22}$$

where $\varphi_1 \ldots \varphi_n$ is a minimal set of formulas (not necessarily unique) such that

$$\varphi_1 \ldots \varphi_n \models \varphi$$

and

$$\mathcal{B}_s - \{\varphi_1 \ldots \varphi_n\} \not\models \varphi \tag{23}$$

In the simplest case $\mathcal{B}^{-\varphi}_s = \mathcal{B}_s - \{\varphi\}$ would hold as long as $\mathcal{B}^{-\varphi}_s \not\models \varphi$.

The notation $\mathcal{B}^{-\varphi}_s$ is needed to capture the circumstances under which a belief is irrelevant to the reasons that a revision needs to occur. If it is the case that $\textbf{Believes}(\neg(\Sigma^s_\alpha \wedge \Pi_\alpha), s)$ holds then revision needs to occur. In other words, it must be the case that

$$\mathcal{B}_s \cup \{\Sigma^s_\alpha\} \cup \{\Pi_\alpha\} \models \textsc{False}.$$

But if for every $\mathcal{B}^{-\mathrm{P}}_s$

$$\mathcal{B}^{-\mathrm{P}}_s \cup \{\Sigma^s_\alpha\} \cup \{\Pi_\alpha\} \models \textsc{False}$$

then the belief in P is not relevant to the causes of the revision and therefore should continue to be believed after revision has taken place.

The following two theorems correspond to the two theorems with the same name in [SL03]. The difference here is that the two cases of revision and update need to be distinguished. If it is a case of update, then both lack of belief and belief in a literal P persists as long as the effect of the action is not to change P. But if we have a case of revision, then not only must it be the case that the action have no effect on P, but it must also be the case that P must not be a cause of the contradiction.

Theorem 2 (Default Persistence of Ignorance) *For all literals* P*, an action α and a situation s, if $\neg\textbf{Believes}(\mathrm{P}, s)$ holds and the axiomatization entails*

$$\forall s \, \textbf{Holds}(\mathrm{P}, s) \equiv \textbf{Holds}(\mathrm{P}, \textsc{do}(\alpha, s))$$

then $\neg\textbf{Believes}(\mathrm{P}, \textsc{do}(\alpha, s))$ holds as well as long as one of the following two conditions applies.

1. $\textbf{Believes}(\neg(\Sigma^s_\alpha \wedge \Pi_\alpha), s)$ *does not hold and* $\neg\textbf{Believes}((\Sigma^s_\alpha \wedge \Pi_\alpha \rightarrow \mathrm{P}), s)$ *does hold.*
2. $\textbf{Believes}(\neg(\Sigma^s_\alpha \wedge \Pi_\alpha), s)$ *holds and it is not the case that for every minimal set* $\{\Psi_1, \ldots, \Psi_n\}$ *such that* $\mathcal{B}_s - \{\Psi_1, \ldots, \Psi_n\} \cup \{\Sigma^s_\alpha\} \cup \{\Pi_\alpha\} \not\models \textsc{False}$
$$\mathcal{B}_s - \{\Psi_1, \ldots, \Psi_n\} \cup \{\Sigma^s_\alpha\} \cup \{\Pi_\alpha\} \models \mathrm{P}$$

If the new information does not contradict the original beliefs of the agent (item 1), P is continued to be not believed as long as the agent did not originally believe that the new information imples P. In the case of a contradiction (item 2), the agent continues not to believe P as long as it is not the case that P is true in all possible minimal revisions.

Theorem 3 (Memory) *For all literals* P *and situations s, if $\textbf{Believes}(\mathrm{P}, s)$ holds then $\textbf{Believes}(\mathrm{P}, \textsc{do}(\alpha, s))$ holds as long as the axiomatization entails*

$$\forall s \, \textbf{Holds}(\mathrm{P}, s) \equiv \textbf{Holds}(\mathrm{P}, \textsc{do}(\alpha, s))$$

and one of the following two conditions applies:

1. $\textbf{Believes}(\neg(\Sigma^s_\alpha \wedge \Pi_\alpha), s)$ *does not hold.*
2. $\textbf{Believes}(\neg(\Sigma^s_\alpha \wedge \Pi_\alpha), s)$ *holds and for every* $\mathcal{B}^{-\mathrm{P}}_s$
$$\mathcal{B}^{-\mathrm{P}}_s \cup \{\Sigma^s_\alpha\} \cup \{\Pi_\alpha\} \models \textsc{False}$$

Again if there is no revision (item 1), the case works exactly as with knowledge. If revision has occurred (item 2), the fluent P is still believed as long as P is not relevant to the causes of the contradiction.

Note that if the agent begins believing R, P $\rightarrow$ Q, and $\neg$Q, and then senses P, revision will occur. The new beliefs will be R, P, and Q. This is completely intuitive and the agent continues to believe R since it is irrelevant to the reasons that revision must occur. Similarly, the agent begins without believing T or $\neg$T. After revision has occurred, this state of non-belief is unchanged.

But the approach does lead to some unintuitive results in the case of iterative revision with faulty sensors. For example, assume the agent starts out believing $P \rightarrow Q$, and $\neg P \rightarrow R$. But the agent does not believe P or $\neg P$. Then if P is sensed as being true, the agent will correctly believe both P and Q. But if the agent then senses P and this time P turns out to be false, the agent will believe $\neg P$ but will not believe R since it has lost a belief in $\neg P \rightarrow R$, when the belief in P was aquired.

7 Comparisons

Unlike [SL03; SPLL00], the approach presented here is unable to handle introspection. This feature is inherited from the framework of [Thi00]. On the other hand, unlike [SPLL00], the solution to the frame problem of [SL03; Thi00] is extended to the case of belief. Jin and Thielscher[JT04] extend the approach of [Thi00] to handle belief and belief revision, but without introspection. Both [JT04]and [SPLL00] utilize a numerical ranking of states/situations as a method of representing the relative believability of possible worlds. It may be the case that with the appropriate ranking, each of these methods could satisfy the theorems presented above. In this case, the work presented here can be seen as a general method of providing such a ranking. Additionally, it may be the case that the work described here can be augmented with a ranking of states to overcome the unintuitive results with regard to iterative revision with faulty sensors.

8 Summary and Future Work

This paper has presented a method for modeling agents with possibly false beliefs and belief producing actions in the situation calculus, while still preserving memory. Current and future work involves the extension of the work to consider knowledge of sentences in first-order logic, handling iterative revision with faulty sensors, the development of reasoning methods to work with this axiomatization, and the incorporation into an agent programming language such as GoLog or Flux.

Acknowledgements

This work was supported by the Knowledge Fusion Center of the Army Research Laboratory under contract number DAAD-03-2-0034 and by ARDA under the AQUAINT program. Thanks are due to the anonymous reviewers for NRAC who provided many useful suggestions. Additionally, helpful comments on an earlier draft were provided by Chitta Baral and Michael Gelfond.

References

[AGM85] C Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change: partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50:510–530, 1985.

[Dal88] M. Dalal. Investigations into a theory of knowledge base revision: Preliminary Report. In *Proc. of 7th national Conference on AI*, pages 475–479, 1988.

[Gar88] Peter Gardenfors. *Knowledge In Flux: Modeling the Dynamics of Epistemic States*. MIT Press, Cambridge, 1988.

[Gro88] Adam Grove. Two modellings for theory change. *Journal of Philosophical Logic*, 17:157–70, 1988.

[JT04] Yi Jin and Michael Thielscher. Representing beliefs in the fluent calculus. In *Proceedings of the European Conference on Artificial Intelligence*, pages 823–827, Valencia, Spain, 2004.

[KM91a] H. Katsuno and A. Mendelzon. On the difference between updating a knowledge base and revising it. In *Proc. of KR 91*, pages 387–394, 1991.

[KM91b] Hirofumi Katsuno and Alberto Mendelzon. Propositional knowledge base revision and minimal change. *Artificial Intelligence*, 52:263–294, 1991.

[MH69] J. McCarthy and P. Hayes. Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer and D. Michie, editors, *Machine Intelligence 4*, pages 463–502. Edinburgh University Press, Edinburgh, UK, 1969.

[Moo80] R.C. Moore. Reasoning about knowledge and action. Technical Note 191, SRI International, October 1980.

[PFN00] Pavlos Peppas, Norman Foo, and Abhaya Nayak. Measuring similarity in belief revision. *Journal of Logic and Computation*, 10(4), 2000.

[Rei91] Raymond Reiter. The frame problem in the situation calculus: A simple solution (sometimes) and a completeness result for goal regression. In Vladimir Lifschitz, editor, *Artificial Intelligence and Mathematical Theory of Computation: Papers in Honor of John McCarthy*, pages 359–380. Academic Press, San Diego, CA, 1991.

[Rei01] Raymond Reiter. *Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems*. The MIT Press, Cambridge, Massachusetts, 2001.

[SL93] Richard B. Scherl and Hector J. Levesque. The frame problem and knowledge producing actions. In *Proceedings, Eleventh National Conference on Artificial Intelligence*, pages 689–695, 1993.

[SL03] Richard Scherl and Hector Levesque. Knowledge, action, and the frame problem. *Artificial Intelligence*, 144:1–39, 2003.

[SPLL00] Steven Shapiro, Maurice Pagnucco, Yves Lespérance, and H.J. Levesque. Iterated belief change in the situation calculus. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Seventh International Conference (KR-2000)*, pages 527–538. Morgan Kaufmann Publishing, 2000.

[Thi98] Michael Thielscher. Introduction to the fluent calculus. *Linköping Electronic Articles in Computer and Information Science*, 3, 1998.

[Thi99] Michael Thielscher. From Situation Calculus to Fluent Calculus: State update axioms as a solution to the inferential frame problem. *Artificial Intelligence*, 111(1–2):277–299, 1999.

[Thi00] Michael Thielscher. Representing the knowledge of a robot. In A. Cohn, F. Giunchiglia, and B. Selman, editors, *Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning (KR)*, pages 109–120. Morgan Kaufmann, 2000.

[Win88] M. Winslett. Reasoning about action using a possible models approach. In *Proc. of 7th national conference on AI*, pages 89–93, 1988.