Iterated Belief Change in the Situation Calculus

Steven Shapiro
Dept. of Computer Science
University of Toronto
Toronto, ON
M5S 3G4 Canada
steven@ai.toronto.edu

Maurice Pagnucco
Dept. of Computing
Macquarie University
NSW 2109
Australia
morri@ics.mq.edu.au

Yves Lespérance
Dept. of Computer Science
York University
Toronto, ON
M3J 1P3 Canada
lesperan@cs.yorku.ca

Hector J. Levesque
Dept. of Computer Science
University of Toronto
Toronto, ON
M5S 3G4 Canada
hector@ai.toronto.edu

Abstract

The ability to reason about action and change has long been considered a necessary component for any intelligent system. Many proposals have been offered in the past to deal with this problem. In this paper, we offer a new approach to belief change associated with performing actions that addresses some of the shortcomings of these approaches. In particular, our approach is based on a well-developed theory of action in the situation calculus extended to deal with belief. Moreover, our account handles nested belief, belief introspection, mistaken belief, and handles belief revision and belief update together with iterated belief change.

1 Introduction

An agent acting in its environment must be capable of reasoning about the state of its environment and keeping track of any changes to the environment due to the performing of actions. Various theories have been developed to give an account of how this can be achieved. Foremost among these are theories of belief change and theories for reasoning about action. While originating from different initial motivations, the two are united in their aim to have agents maintain a model of the environment that matches the actual environment as closely as possible given the available information. An important consideration is the ability to deal with more than one change; known as the problem of *iterated belief change*.

In this paper, we consider a new approach for modeling iterated belief change using the language of the *situation calculus* [15]. While our approach is limited in its applicability, we feel that it is conceptually very simple and offers a number of useful features not found in other approaches:

- It is completely integrated with a well-developed theory of action in the situation calculus [18] and its extension to handle knowledge expansion [19]. Specifically, how beliefs change in our account is simply a special case of how other fluents change as the result of actions, and thus among other things, we inherit a solution to the frame problem.
- Like Scherl and Levesque [19], our theory accommodates both belief *update* and belief *expansion*. The former concerns beliefs that change as the result of the realization that the world has changed; the latter concerns beliefs that change as the result of new information acquired.
- Unlike Scherl and Levesque, however, our theory is not limited to belief expansion; rather it deals with the more general case of belief *revision*. It will be possible in our model for an agent to believe some formula ϕ, acquire information that causes it to change its mind and believe $\neg\phi$ (without believing the world has changed), and later go back to believing ϕ again. In Scherl and Levesque and in other approaches based on this work such as [12, 13], new information that contradicts previous beliefs cannot be consistently accommodated.
- Because belief change in our model is always the result of action, our account naturally supports *iterated* belief change. This is simply the result of a sequence of actions. Moreover, each individual action can potentially cause both an update (by changing the world) and a revision (by providing sensing information) in a seamless way.
- Like Scherl and Levesque and unlike many previous approaches to belief change, e.g., [9, 11], our approach supports belief *introspection*: an agent will know what it believes and does not believe. Furthermore, it has information about the past, and so will also know what it used to believe and not believe. Finally, an agent will be able to predict what it will believe after it acquires information through sensing.

- Unlike Scherl and Levesque, our agents will be able to introspectively tell the difference between an update and a revision as it moves from believing ϕ to believing $\neg\phi$. In the former case, the agent will believe that it believed ϕ in the past, and that it was correct to do so; in the latter case, it will believe that it believed ϕ in the past but that it was *mistaken*.

The rest of the paper is organized as follows: in the next section, we briefly review the situation calculus including the Scherl and Levesque [19] model of belief expansion, and we review the most popular accounts of belief revision, belief update and iterated belief change; in Section 3, we motivate and define a new belief operator as a modification to the one used by Scherl and Levesque; in Section 4, we prove some properties of this operator, justifying the points made above; in Section 5, we show the operator in action on a simple example, and how an agent can change its mind repeatedly; in Section 6, we consider the importance of our work and compare it to some of the existing approaches to belief change; in the final section, we draw some conclusions and discuss future work.

2 Background

The basis of our framework for belief change is an action theory [18] based on the situation calculus [15], and extended to include a belief operator [19]. In this section, we begin with a brief overview of the situation calculus and follow it with a short review of belief change.

2.1 Situation Calculus

The situation calculus is a predicate calculus language for representing dynamically changing domains. A situation represents a snapshot of the domain. There is a set of initial situations corresponding to the ways the agent[1] believes the domain might be initially. The actual initial state of the domain is represented by the distinguished initial situation constant, S_0, which may or may not be among the set of initial situations believed possible by the agent. The term $\mathit{do}(a, s)$ denotes the unique situation that results from the agent performing action a in situation s. Thus, the situations can be structured into a set of trees, where the root of each tree is an initial situation and the arcs are actions. The initial situations are defined as those situations that do not have a predecessor:

Definition 1

$$\mathit{Init}(s) \overset{\text{def}}{=} \neg\exists a, s'.s = \mathit{do}(a, s').$$

Predicates and functions whose value may change from situation to situation (and whose last argument is a situation) are called *fluents*. For instance, we use the fluent $\textsc{InR}_1(s)$ to represent that the agent is in room R_1 in situation s. The effects of actions on fluents are defined using successor state axioms [18], which provide a succinct representation for both effect axioms and frame axioms [15]. For example, assume that there are only two rooms, R_1 and R_2, and that the action $\textsc{leave}$ takes the agent from the current room to the other room. Then, the successor state axiom for $\textsc{InR}_1$ is:[2]

$$\begin{aligned}&\textsc{InR}_1(\mathit{do}(a, s)) \equiv \\ &\quad ((\neg\textsc{InR}_1(s) \land a = \textsc{leave}) \lor (\textsc{InR}_1(s) \land a \neq \textsc{leave})).\end{aligned}$$

This axiom asserts that the agent will be in R_1 after doing some action iff either the agent is in R_2 ($\neg\textsc{InR}_1(s)$) and leaves it or the agent is currently in R_1 and the action is anything other than leaving it.

Moore [16] defined a possible-worlds semantics for a modal logic of knowledge in the situation calculus by treating situations as possible worlds. Scherl and Levesque [19] adapted the semantics to the action theories of Reiter [18]. The idea is to have an accessibility relation on situations, $B(s', s)$, which holds if in situation s, the situation s' is considered possible by the agent. Note, the order of the arguments is reversed from the usual convention in modal logic.

Levesque [13] introduced a predicate, $\mathit{SF}(a, s)$, to describe the result of performing the binary-valued sensing action a. $\mathit{SF}(a, s)$ holds iff the sensor associated with a returns the sensing value 1 in situation s. Each sensing action senses some property of the domain. The property sensed by an action is associated with the action using a *guarded sensed fluent axiom* [10]. For example, suppose that there are lights in R_1 and R_2 and that $\textsc{Light}_1(s)$ ($\textsc{Light}_2(s)$, resp.) holds if the light in R_1 (R_2, resp.) is on. Then:

$$\begin{aligned}&\textsc{InR}_1(s) \supset (\mathit{SF}(\textsc{senseLight}, s) \equiv \textsc{Light}_1(s)) \\ &\neg\textsc{InR}_1(s) \supset (\mathit{SF}(\textsc{senseLight}, s) \equiv \textsc{Light}_2(s))\end{aligned}$$

can be used to specify that the $\textsc{senseLight}$ action senses whether the light in the room where the agent is currently located is on.

Scherl and Levesque [19] defined a successor state axiom for B that shows how actions, including sensing actions, affect the beliefs of the agent. We use the same axiom (with some notational variation) here:

Axiom 1 *(Successor State Axiom for B)*

$$\begin{aligned}&B(s'', \mathit{do}(a, s)) \equiv \\ &\exists s'[B(s', s) \land s'' = \mathit{do}(a, s') \land (\mathit{SF}(a, s') \equiv \mathit{SF}(a, s))].\end{aligned}$$

The situations s'' that are B-related to $\mathit{do}(a, s)$ are the ones that result from doing action a in a situation s', such that the sensor associated with action a has the same value in s' as it does in s. We will see in Section 3 how a modal belief operator can be defined in terms of this fluent.

[1] The situation calculus can accommodate multiple agents, but for the purposes of this paper we assume that there is a single agent, and all actions are performed by that agent.

[2] We adopt the convention that unbound variables are universally quantified in the widest scope.

There are various ways of axiomatizing dynamic applications in the situation calculus. Here we adopt a simple form of the guarded action theories described by De Giacomo and Levesque [10] consisting of: (1) successor state axioms[3] for each fluent (including B and *pl* introduced below), and guarded sensed fluent axioms for each action, as discussed above; (2) unique names axioms for the actions, and domain-independent foundational axioms (similar to the ones given by Lakemeyer and Levesque [12]), which we do not describe further here; and (3) initial state axioms, which describe the initial state of the domain and the initial beliefs of the agent.[4] For simplicity, we assume here that all actions are always executable and omit the action precondition axioms and references to a *Poss* predicate that are normally included in situation calculus action theories.

In what follows, we will use Σ to refer to a guarded action theory of this form. By a *domain-dependent fluent*, we mean a fluent other than B or *pl*, and a *domain-dependent formula* is one that only mentions domain-dependent fluents. Finally, we say that a domain-dependent formula is *uniform* in s iff s is the only situation term in that formula.

2.2 Belief Change

Before formally defining a belief operator in this language, we briefly review the notion of belief change as it exists in the literature. Belief change, simply put, aims to study the manner in which an agent's epistemic (belief) state should change when the agent is confronted by new information. In the literature,[5] there is often a clear distinction between two forms of belief change: *revision* and *update*. Both forms can be characterized by an axiomatic approach (in terms of rationality postulates) or through various constructions (e.g., epistemic entrenchment, possible worlds, etc.). The AGM theory [9] is the prototypical example of belief revision while the KM framework [11] is often identified with belief update.

Intuitively speaking, belief revision is appropriate for modeling static environments about which the agent has only partial and possibly incorrect information. New information is used to fill in gaps and correct errors, but the environment itself does not undergo change. Belief update, on the other hand, is intended for situations in which the environment itself is changing due to the performing of actions.

For completeness and later comparison, we list here the AGM postulates [1, 9] for belief revision. By $K*\phi$ we mean the revision of belief state K by new information ϕ.[6]

(K*1) $K * \phi$ is deductively closed
(K*2) $\phi \in K * \phi$
(K*3) $K * \phi \subseteq K + \phi$
(K*4) If $\neg\phi \notin K$, then $K + \phi \subseteq K * \phi$
(K*5) $K * \phi = \mathcal{L}$ iff $\models \neg\phi$
(K*6) If $\models \phi \equiv \psi$, then $K * \phi = K * \psi$
(K*7) $K * (\phi \wedge \psi) \subseteq (K * \phi) + \psi$
(K*8) If $\neg\psi \notin K * \phi$, then $(K * \phi) + \psi \subseteq K * (\phi \wedge \psi)$

Katsuno and Mendelzon provide the following postulates for belief update, where $K \diamond \phi$ denotes the update of K by formula ϕ.[7]

(K$\diamond$1) $K \diamond \phi$ is deductively closed
(K$\diamond$2) $\phi \in K \diamond \phi$
(K$\diamond$3) If $\phi \in K$, then $K \diamond \phi = K$
(K$\diamond$4) $K \diamond \phi = \mathcal{L}$ iff $K \models \perp$ or $\phi \models \perp$
(K$\diamond$5) If $\models \phi \equiv \psi$, then $K \diamond \phi = K \diamond \psi$
(K$\diamond$6) $K \diamond (\phi \wedge \psi) \subseteq (K \diamond \phi) + \psi$
(K$\diamond$7) If K is complete and $\neg\psi \notin K \diamond \phi$,
then $(K \diamond \phi) + \psi \subseteq K \diamond (\phi \wedge \psi)$
(K$\diamond$8) If $[K] \neq \emptyset$, then $K \diamond \phi = \bigcap_{w \in [K]} w \diamond \phi$

One of the major issues in this area is that of *iterated belief change*, i.e., modeling how the agent's beliefs change after multiple belief revisions or updates occur. Two of the main developments in this area are the work of Darwiche and Pearl [6] and Boutilier [4]. Darwiche & Pearl put forward the following postulates as a way of extending the AGM revision postulates to handle *iterated revision*.[8]

(DP1) If $\psi \models \phi$, then $(K * \phi) * \psi = K * \psi$
(DP2) If $\psi \models \neg\phi$, then $(K * \phi) * \psi = K * \psi$
(DP3) If $\phi \in K * \psi$, then $\phi \in (K * \phi) * \psi$
(DP4) If $\neg\phi \notin K * \psi$, then $\neg\phi \notin (K * \phi) * \psi$

In Section 6.2, we return to consider the extent to which our framework satisfies these postulates.

[3] We could use the more general *guarded successor state axioms* of De Giacomo and Levesque [10], but regular successor state axioms suffice for the simple domain we consider here.

[4] These are axioms that only describe initial situations. Reiter [18] has S_0 as the only initial situation, but to formalize belief, we need additional ones.

[5] We shall restrict our attention to approaches in the AGM vein [1, 9, 11] although there are many others.

[6] In the AGM theory, K is a set of formulae and ϕ is a formula taken from an object language $\mathcal{L}$ containing the standard boolean connectives and the logical constant $\perp$ (falsum). Furthermore, K is a set of formulae (from $\mathcal{L}$) closed under the deductive consequence operator Cn associated with the underlying logic. The operation $K + \phi$ denotes the belief expansion of K by ϕ and is defined as $K + \phi = Cn(K \cup \{\phi\})$. $[K]$ denotes the set of all consistent complete theories of $\mathcal{L}$ containing K.

[7] To facilitate comparison with the AGM postulates, we have reformulated the original postulates of Katsuno and Mendelzon into an equivalent set using AGM-style terminology [17]. For renderings of these postulates and the AGM postulates above in the KM-style, refer to Katsuno & Mendelzon [11].

[8] Again, we have translated the Darwiche and Pearl postulates into AGM-style terminology rather than KM-style terminology used in the original paper.

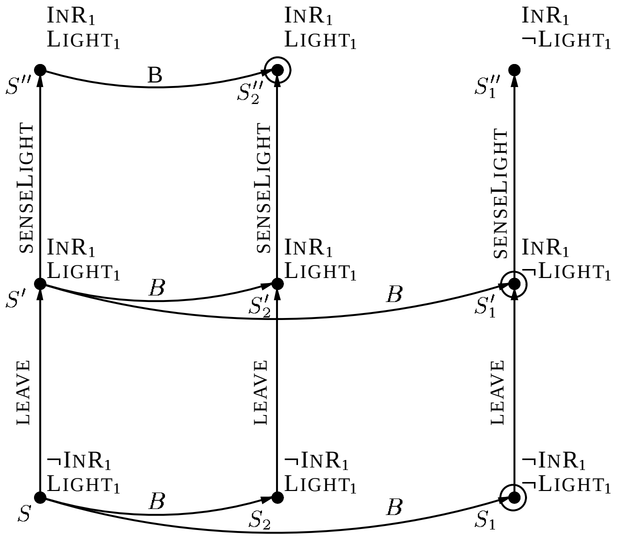

Figure 1: An example of belief update and revision.

3 Definition of the Belief Operator

In this section, we define what it means for an agent to believe a formula ϕ in a situation s, i.e., $\mathit{Bel}(\phi, s)$. Since ϕ will usually contain fluents, we introduce a special symbol *now* as a placeholder for the situation argument of these fluents, e.g., $\mathit{Bel}(\mathrm{INR}_1(\mathit{now}), s)$. $\phi[s]$ denotes the formula that results from substituting s for *now* in ϕ. To make the formulae easier to read, we will often suppress the situation argument of fluents in the scope of a belief operator, e.g., $\mathit{Bel}(\mathrm{INR}_1, s)$.

Scherl and Levesque [19], define a modal operator for belief in terms of the accessibility relation on situations, $B(s', s)$. For Scherl and Levesque, the believed formulae are the ones true in all accessible situations:

Definition 2

$$\mathit{Bel}_{\mathrm{SL}}(\phi, s) \overset{\text{def}}{=} \forall s'(B(s', s) \supset \phi[s']).$$

To understand how belief change works, both in Scherl and Levesque and here, consider the example illustrated in Figure 1. In this example, we have three initial situations S, S_1, and S_2. S_1 and S_2 are B-related to S (i.e., $B(S_1, S)$ and $B(S_2, S)$), as indicated by the arrows labeled B. (Ignore the circles around certain situations for now.) In all three situations, the agent is not in the room R_1. In S and S_2 the light in R_1 is on, and in S_1 the light is off. So at S, the agent believes it is not in R_1 (i.e., that it is in R_2), but it has no beliefs about the status of the light in R_1. We first consider the action of leaving R_2, which will lead to a belief update. By the successor state axiom for B, both $\mathit{do}(\mathrm{LEAVE}, S_1)$ and $\mathit{do}(\mathrm{LEAVE}, S_2)$ are B-related to $\mathit{do}(\mathrm{LEAVE}, S)$. In the figure, these three situations are called S'_1, S'_2 and S', respectively. The successor state axiom for INR_1 causes INR_1 to hold in these situations. Therefore, the agent believes INR_1 in S'. By the successor state axiom for LIGHT_1, which we state below, the truth value of LIGHT_1 would not change as the result of LEAVE.

Now the agent performs the sensing action SENSELIGHT. According to the sensed fluent axioms for SENSELIGHT, $\mathit{SF}(\mathrm{SENSELIGHT}, S^*)$ holds for situation S^* iff the light is on in the room in which the agent is located in S^*. In the figure, the light in R_1 is on in S' and S'_2, but not in S'_1. So, *SF* holds for SENSELIGHT in the former two situations but not in the latter. The successor state axiom for B ensures that after doing a sensing action A, any situation that disagrees with the actual situation on the value of *SF* for A is dropped from the B relation in the successor state. In the figure, S' is the actual situation. Since S'_1 disagrees with S' on the value of *SF* for SENSELIGHT, $\mathit{do}(\mathrm{SENSELIGHT}, S'_1)$ (labeled S''_1 in the figure) *is not* B-related to $\mathit{do}(\mathrm{SENSELIGHT}, S')$ (labeled S''). On the other hand, S'_2 and S' agree on the value of *SF* for SENSELIGHT, so $\mathit{do}(\mathrm{SENSELIGHT}, S'_2)$ (labeled S''_2 in the figure) *is* B-related to S''. The result is that the agent believes the light is on in S''. This is an example of belief expansion because the belief that the light is on was simply added to the belief state of the agent.

Our definition of *Bel* is similar to the one in Scherl and Levesque, but we are going to generalize their account in order to be able to talk about how *plausible* the agent considers a situation to be. Plausibility is assigned to situations using a function $\mathit{pl}(s)$, whose range is the natural numbers, where lower values indicate higher plausibility. The *pl* function only has to be specified over initial situations, using an initial state axiom. The plausibility of successor situations is left unchanged using the following successor state axiom:

Axiom 2 *(Successor State Axiom for pl)*

$$\mathit{pl}(\mathit{do}(a, s)) = \mathit{pl}(s).$$

Unlike Scherl and Levesque, we will say that the agent believes a proposition ϕ in situation s, if ϕ holds in the *most plausible* B-related situations. Here is our definition of the belief operator:

Definition 3

$$\begin{aligned}&\mathit{Bel}(\phi, s) \overset{\text{def}}{=}\\ &\quad \forall s'[B(s', s) \wedge (\forall s''.B(s'', s) \supset \mathit{pl}(s') \leq \mathit{pl}(s''))] \supset\\ &\qquad \phi[s'].\end{aligned}$$

That is, ϕ is believed at s precisely when it holds at all the most plausible situations B-related to s. Note that the actual numbers assigned to the situations are not relevant. All that is important is the ordering of the situations by plausibility. We could have used any total pre-order on situations for this purpose, but using $\leq$ on natural numbers simplifies the presentation of our framework.

We now return to the initial situations in Figure 1, and add a plausibility structure to the belief state of the agent by supposing that S_1 is more plausible than S_2 (indicated by the circle surrounding S_1). For example, suppose that $\mathit{pl}(S_1) = 0$ and $\mathit{pl}(S_2) = 1$. Now, the beliefs of the agent are determined only by S_1. Therefore, the agent now has a belief about the light in R_1 in S, namely that the light is off. After leaving R_2 and entering R_1, the agent continues to believe that the light is off. After doing $\textsc{senseLight}$, S_1'' is dropped from B as before, so now S_2'' is the most plausible accessible situation, which means that it determines the beliefs of the agent. Since the light is on in S_2'', the agent believes it is on in S''. Since the agent goes from believing the light is off to believing it is on, this is a case of belief revision.

Both accounts of belief handle *belief introspection* of current and past beliefs. In order to obtain positive and negative introspection of beliefs, we require B to be initially transitive and euclidean. For notational simplicity, we combine the two constraints into a single constraint, which says that any situation that is B-related to an initial situation s is B-related to the same situations as s. We assert this constraint as an initial state axiom:

Axiom 3

$$\mathit{Init}(s) \wedge B(s', s) \supset (\forall s''.B(s'', s') \equiv B(s'', s)).$$

As in Scherl and Levesque, the successor state axiom for B ensures that this constraint is preserved over all situations:

Theorem 1

$$B(s', s) \supset (\forall s''.B(s'', s') \equiv B(s'', s)).$$

In order to clarify how this constraint ensures that introspection is handled properly, we will show that in the example illustrated in Figure 1, the agent positively introspects its past beliefs. First, we need some notation that allows us to talk about the past. We use $\mathit{Previously}(\phi, s)$ to denote that ϕ held in the situation immediately before s:

Definition 4

$$\mathit{Previously}(\phi, s) \overset{\text{def}}{=} \exists a, s'.s = do(a, s') \wedge \phi[s'].$$

We want to show that $\mathit{Bel}(\mathit{Previously}(\mathit{Bel}(\neg\textsc{Light}_1)), S'')$[9] holds, i.e., in S'', the agent believes that in the previous situation it believed that the light in R_1 was off. Consider a situation S^* that is among the most plausible B-related situations to S''. In this example, there is only one such situation, namely, S_2''. We need to show that $\mathit{Previously}(\mathit{Bel}(\neg\textsc{Light}_1), S_2'')$ holds, i.e., that $\mathit{Bel}(\neg\textsc{Light}_1, S_2')$ holds. By Theorem 1, S_2' is B-related to the same situations as S', i.e., S_1' and S_2'. Since S_1' is more plausible than S_2', we only require that $\neg\textsc{Light}_1(S_1')$ holds. Since this is true, we see that $\mathit{Bel}(\mathit{Previously}(\mathit{Bel}(\neg\textsc{Light}_1)), S'')$ is also true.

The specification of *pl* and B over the initial situations is the responsibility of the axiomatizer of the domain in question. This specification need not be complete. Of course, a more complete specification will yield more interesting properties about the agent's current and future belief states.

We have another constraint on the specification of B over the initial situations: the situations B-related to an initial situation are themselves initial, i.e., the agent believes that initially nothing has happened. We assert this constraint as an initial state axiom:

Axiom 4

$$\mathit{Init}(s) \wedge B(s', s) \supset \mathit{Init}(s').$$

4 Properties

In this section, we highlight some of the more interesting properties of our framework. In order to clarify our explanations and facilitate a comparison with previous approaches to belief change, it will be important for us to attach a specific meaning to the use of the terms *revision* and *update*, which we shall do here.

4.1 Belief Revision

Recall (Section 2.2) that belief revision is suited to the acquisition of information about static environments for which the agent may have mistaken or partial information. In our framework, this can only be achieved through the use of sensing actions. We suppose that to revise by a formula ϕ, there is a corresponding sensing action capable of determining the truth value of ϕ. Moreover, we assume that this sensing action has no effect on the environment; the only fluent it changes is B.[10]

More formally, we define a revision action as follows:

Definition 5 *(Revision Action for ϕ)*
A revision action A for a formula ϕ (uniform in now) wrt action theory Σ is a sensing action satisfying $\Sigma \models [\forall s.SF(A, s) \Leftrightarrow \phi[s]] \wedge [\forall s \forall \vec{x}.F(\vec{x}, s) \Leftrightarrow F(\vec{x}, do(A, s))]$ (for every domain-dependent fluent F).

We now show that belief revisions are handled appropriately in our system in the sense that if the sensor indicates that ϕ holds, then the agent will indeed believe ϕ after performing A. Similarly, if the sensor indicates that ϕ is false, then the agent will believe $\neg\phi$ after doing A.

Theorem 2
Let A be a revision action for formula ϕ (uniform in now).

[9] Recall that we omit the situation argument of fluents in the scope of a *Bel* operator whenever possible.

[10] This is not an overly strict imposition for we can capture sensing actions that modify the domain by "decomposing" the action into a sequence of non-sensing actions and sensing actions.

It follows that:

$$\Sigma \models [\forall s.\phi[s] \supset \mathit{Bel}(\phi, \mathit{do}(A, s))] \wedge$$
$$[\forall s.\neg\phi[s] \supset \mathit{Bel}(\neg\phi, \mathit{do}(A, s))]$$

If the agent is indifferent towards ϕ before doing the action, i.e., does not believe ϕ or $\neg\phi$, this is a case of belief expansion. If, before sensing, the agent believes the opposite of what the sensor indicates, then we have belief revision.

Note that this theorem also follows from Scherl and Levesque's theory. However, for Scherl and Levesque, if the agent believes ϕ in S and the sensor indicates that ϕ is false, then in $\mathit{do}(A, S)$, the agent's belief state will be inconsistent. The agent will then believe all propositions, including $\neg\phi$. In our theory, the agent's belief state will be consistent in this case, as long as there is some situation S', accessible from S that agrees with S on the value of the sensor associated with A (here, A can be any action, not just a revision action):

Theorem 3

$$\Sigma \models \forall a, s\{[\exists s'.B(s', s) \wedge (\mathit{SF}(a, s') \equiv \mathit{SF}(a, s))] \supset$$
$$\neg\mathit{Bel}(\mathit{FALSE}, \mathit{do}(a, s))\}$$

Since S' is not necessarily among the *most plausible* accessible situations, the agent can consistently believe ϕ in S and $\neg\phi$ in $\mathit{do}(A, S)$. As a direct corollary to this result, if we restrict our attention to revision actions for ϕ where the agent considers ϕ is possible, it will not hold inconsistent beliefs after performing A.

Corollary 4
Let A be a revision action for a formula ϕ (uniform in now). It follows that:

$$\Sigma \models (\exists s'.B(s', s) \wedge (\phi[s'] \equiv \phi[s]) \supset$$
$$\neg\mathit{Bel}(\mathit{FALSE}, \mathit{do}(A, s)).$$

4.2 Belief Update

Belief update refers to the belief change that takes place due to a change in the environment. In analogy to revision, we introduce the notion of an update action. (Recall that we assume that actions are always possible.)

Definition 6 *(Update Action for ϕ)*
An update action A for a formula ϕ (uniform in now) wrt action theory Σ is a non-sensing action that always makes ϕ true in the environment. That is, $\Sigma \models \forall s.\phi[do(A, s)] \wedge \forall s.\mathit{SF}(A, s)$.

As with Scherl and Levesque's theory, the agent's beliefs are updated appropriately when an update action A for ϕ occurs, in the sense that the agent will believe ϕ after A is performed.

Theorem 5
Let A be an update action for a formula ϕ (uniform in now). It follows that:

$$\Sigma \models \forall s.\ \mathit{Bel}(\phi,\ \mathit{do}(A,\ s))$$

In our framework, we can represent actions that do not fall under the category of update actions. Of particular interest are ones whose effects depend on what is true in the current situation. We can prove an analogous theorem for such actions. Let A be a non-sensing action, i.e., $\forall s.\mathit{SF}(A, s)$. Further suppose that A is an action that causes ϕ' to hold, if ϕ holds beforehand, and that the agent believes ϕ in S. Then after performing A in S, the agent ought to believe that ϕ' holds:

Theorem 6

$$\Sigma \models \mathit{Bel}(\phi, s) \wedge (\forall s'.\mathit{SF}(A, s')) \wedge$$
$$(\forall s'.\phi[s'] \supset \phi'[\mathit{do}(A, s')]) \supset$$
$$\mathit{Bel}(\phi', \mathit{do}(A, s)).$$

4.3 Introspection

In Section 3, we claimed that the agent can introspect its beliefs. We do indeed have this as a theorem.

Theorem 7

$$\Sigma \models [\mathit{Bel}(\phi, s) \supset \mathit{Bel}(\mathit{Bel}(\phi), s)] \wedge$$
$$[\neg\mathit{Bel}(\phi, s) \supset \mathit{Bel}(\neg\mathit{Bel}(\phi), s)].$$

This is a straightforward consequence of Theorem 1.

4.4 Awareness of Mistakes

In Section 3, we also claimed that the agent can introspect its past beliefs. Now suppose that the agent believes ϕ in S, and after performing a sensing action A in S, the agent discovers that ϕ is false. In $\mathit{do}(A, S)$, the agent should believe that in the previous situation ϕ was false, but it believed ϕ was true. In other words, the agent should believe that it was mistaken about ϕ. We now state a theorem that says that the agent will indeed believe that it was mistaken about ϕ. First note that this only holds if A does not affect ϕ. If A causes ϕ to become false, then there is no reason for the agent to believe that ϕ was false in the previous situation. In the theorem, we rule out this case by stating in the antecedent that for any situation S', ϕ holds in S' iff ϕ holds in $\mathit{do}(A, S')$.

Theorem 8

$$\Sigma \models \mathit{Bel}(\phi, s) \wedge \mathit{Bel}(\neg\phi, \mathit{do}(a, s)) \wedge$$
$$(\forall s'.\phi[s'] \equiv \phi[\mathit{do}(a, s')]) \supset$$
$$\mathit{Bel}(\mathit{Previously}(\neg\phi \wedge \mathit{Bel}(\phi)), \mathit{do}(a, s)).$$

In Section 6.2, we will discuss to what extent standard AGM revision and KM update postulates are satisfied in our framework.

5 Example

We now present an example to illustrate how this theory of belief change can be applied. We model a world in which there are two rooms, R_1 and R_2. The agent can move between the rooms. Each room contains a light that can be on

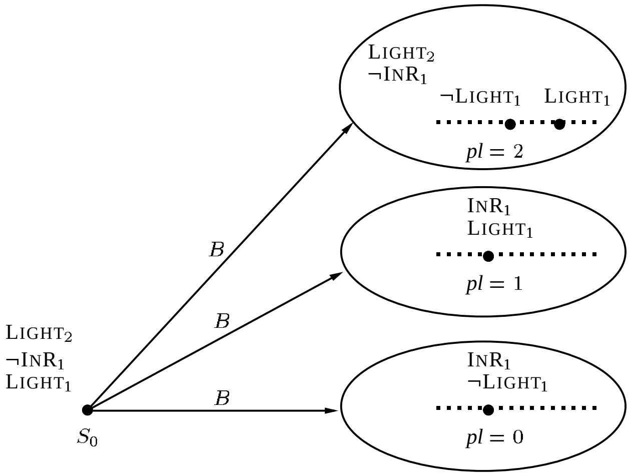

Figure 2: The initial state of the example domain.

or off. The agent has two binary sensors. One sensor detects whether or not the light is on in the room in which the agent is currently located. The other sensor detects whether or not the agent is in R_1.

We have three fluents: $\textsc{Light}_1(s)$ ($\textsc{Light}_2(s)$, resp.), which holds iff there is light in R_1 (R_2, resp.) in situation s, and $\textsc{InR}_1(s)$, which holds if the agent is in R_1 in s. If the agent is not in R_1, then it is assumed to be in R_2. There are three actions: the agent leaves the room it is in and enters the other room ($\textsc{leave}$), the agent senses whether it is in R_1 ($\textsc{senseInR}_1$), and the agent senses whether the light is on in the room in which it is currently located ($\textsc{senseLight}$).

The successor state axioms and guarded sensed fluent axioms for our example, which we will call E, are as follows:

$$\textsc{Light}_1(do(a,s)) \equiv \textsc{Light}_1(s)$$
$$\textsc{Light}_2(do(a,s)) \equiv \textsc{Light}_2(s)$$
$$\textsc{InR}_1(do(a,s)) \equiv$$
$$\quad((\neg\textsc{InR}_1(s) \wedge a = \textsc{leave}) \vee (\textsc{InR}_1(s) \wedge a \neq \textsc{leave}))$$
$$\text{TRUE} \supset (SF(\textsc{leave},s) \equiv \text{TRUE})$$
$$\textsc{InR}_1(s) \supset (SF(\textsc{senseLight},s) \equiv \textsc{Light}_1(s))$$
$$\neg\textsc{InR}_1(s) \supset (SF(\textsc{senseLight},s) \equiv \textsc{Light}_2(s))$$
$$\text{TRUE} \supset (SF(\textsc{senseInR}_1,s) \equiv \textsc{InR}_1(s))$$

Next we must specify the initial state. This includes both the physical state of the domain and the belief state of the agent. First we describe the initial physical state of the domain, by saying which domain-dependent fluents hold in the actual initial situation, S_0. Initially, the lights in both rooms are on and the agent is in R_2 (this is illustrated on the left side of Figure 2):

$$\textsc{Light}_1(S_0) \wedge \neg\textsc{InR}_1(S_0) \wedge \textsc{Light}_2(S_0).$$

The initial belief state of the agent is illustrated in Figure 2. It shows that in the most plausible (the ones with plausibility 0 in the figure) B-related situations to S_0, $\neg\textsc{Light}_1$ and $\textsc{InR}_1$ hold. In the next most plausible (the ones with plausibility 1) B-related situations to S_0, $\textsc{Light}_1$ and $\textsc{InR}_1$ hold. In the third most plausible (the ones with plausibility 2) B-related situations to S_0, $\textsc{Light}_2$ and $\neg\textsc{InR}_1$ hold. There is also at least one situation in the latter group in which $\textsc{Light}_1$ holds and one in which $\neg\textsc{Light}_1$ holds. Specifying this belief state directly can be cumbersome. For example, the axiom for the situations with plausibility 1 is:

$$(\exists s.Init(s) \wedge B(s,S_0) \wedge pl(s) = 1) \wedge$$
$$(\forall s.Init(s) \wedge pl(s) = 1 \supset \textsc{Light}_1(s) \wedge \textsc{InR}_1(s)).$$

For now, we will not enumerate the set of axioms that specify the belief state shown in Figure 2. But we assume that we have such a set which, together with the axioms for the initial physical state, we refer to as I. After we have discussed the example, we will show that there is a more elegant way to specify the initial belief state of the agent. So for this example, Σ consists of the foundational axioms, unique names axioms, Axioms 1–4, E, and I, for which we get the following:

Theorem 9 *The following formulae are entailed by* Σ*:*

i. $Bel(\neg\textsc{Light}_1 \wedge \textsc{InR}_1, S_0)$

ii. $Bel(\textsc{Light}_1 \wedge \textsc{InR}_1, do(\textsc{senseLight}, S_0))$

iii. $Bel(\neg\textsc{InR}_1, do(\textsc{senseInR}_1, do(\textsc{senseLight}, S_0)))$

iv. $Bel(Previously(\neg\textsc{InR}_1 \wedge Bel(\textsc{InR}_1)),$
$\quad do(\textsc{senseInR}_1, do(\textsc{senseLight}, S_0)))$

v. $\neg Bel(\textsc{Light}_1, do(\textsc{senseInR}_1, do(\textsc{senseLight}, S_0))) \wedge$
$\neg Bel(\neg\textsc{Light}_1, do(\textsc{senseInR}_1, do(\textsc{senseLight}, S_0)))$

vi. $Bel(\textsc{InR}_1,$
$\quad do(\textsc{leave}, do(\textsc{senseInR}_1, do(\textsc{senseLight}, S_0))))$

vii. $Bel(\textsc{Light}_1,$
$\quad do(\textsc{senseLight},$
$\qquad do(\textsc{leave}, do(\textsc{senseInR}_1,$
$\qquad\qquad do(\textsc{senseLight}, S_0))))).$

We shall now give a short, informal explanation of why each part of the previous theorem holds.

i. In the most plausible situations B-related to S_0, $\neg\textsc{Light}_1 \wedge \textsc{InR}_1$ holds.

ii. Even though the agent believes that it is in R_1 initially, it is actually in R_2. Therefore, its light sensor is measuring whether there is light in R_2, even though the agent thinks that it is measuring whether there is light in R_1. It turns out that there is light in R_2 in S_0, so the sensor returns 1. Since the agent believes that the light sensor is measuring whether there is light in R_1 and in all the situations with plausibility 0, there is no light in R_1, those situations are dropped from the B relation. In the situations with plausibility 1, the light is on in R_1, so those situations are retained. In those situations $\textsc{Light}_1 \wedge \textsc{InR}_1$ holds and those fluents are not affected by the $\textsc{senseLight}$ action, so the agent believes $\textsc{Light}_1 \wedge \textsc{InR}_1$ after doing $\textsc{senseLight}$.

iii. Now the agent senses whether it is in R_1. Again the agent's most plausible situations conflict with what is actually the case, so they are dropped from the B relation. The situations with plausibility 2 become the most plausible situations, so the agent believes it is not in R_1.

iv. By Theorem 8, the agent realizes that it was mistaken about being in R_1.

v. Among the situations with plausibility 2, there is one in which the light is on in R_1 and one in which it is not on. Therefore, the agent is unsure as to whether the light is on.

vi. Now the agent leaves R_2 and enters R_1. This happens in all the B-related situations as well. Therefore, the agent believes that it is in R_1. This is an example of an update.

vii. The light in R_1 was on initially, and since no action was performed that changed the state of the light, the light remains on. After checking its light sensor, the agent believes that the light is on in R_1.

This example shows that the agent's beliefs change appropriately after both revision actions and update actions. The example also demonstrates that our formalism can accommodate iterated belief change. The agent goes from believing that the light is not on, to believing that it is on, to not believing one way or the other, and then back to believing that it is on.

To facilitate the specification of the initial belief state of the agent, we find it convenient to define another belief operator $\Rightarrow$, in the spirit of the conditional logic connective [14]:

Definition 7

$$\begin{aligned} &\phi \Rightarrow_s \psi \overset{\text{def}}{=} \\ &\quad \forall s'[B(s',s) \wedge \phi[s'] \wedge \\ &\qquad \forall s''(B(s'',s) \wedge \phi[s''] \supset pl(s') \leq pl(s'')) \supset \\ &\qquad\quad \psi[s']]. \end{aligned}$$

$\phi \Rightarrow_s \psi$ holds if in the most plausible situations B-related to s where ϕ holds, ψ also holds. Note that for any situation S, $Bel(\phi, S)$ is equivalent to $(\text{TRUE} \Rightarrow_S \phi)$.

We can use this operator to specify the initial belief state of the agent without having to explicitly mention the plausibility of situations. To obtain the results of Theorem 9, it suffices to let I be the following set of axioms:

$$\begin{aligned} &\textsc{Light}_1(S_0) \wedge \neg\textsc{InR}_1(S_0) \wedge \textsc{Light}_2(S_0) \\ &\text{TRUE} \Rightarrow_{S_0} \neg\textsc{Light}_1 \wedge \textsc{InR}_1 \\ &\textsc{Light}_1 \Rightarrow_{S_0} \textsc{InR}_1 \\ &\neg(\textsc{Light}_2 \wedge \neg\textsc{InR}_1 \Rightarrow_{S_0} \textsc{Light}_1) \\ &\neg(\textsc{Light}_2 \wedge \neg\textsc{InR}_1 \Rightarrow_{S_0} \neg\textsc{Light}_1) \end{aligned}$$

It is easy to see that the belief state depicted in Figure 2 satisfies these axioms. In the most plausible worlds, $(\neg\textsc{Light}_1 \wedge \textsc{InR}_1)$ holds. In the most plausible worlds where the light in R_1 is on, the agent is in R_1. Finally, the last two axioms state that among the most plausible worlds where the light is on in R_2 and the agent is in R_2, there is one where the is light is off in R_1 and one in which the light is on (resp.).

6 Discussion

There are various aspects of our framework that deserve further consideration. We address what we consider to be some of the more important issues here.

6.1 Plausibility Ordering

Our plausibility function is based on ordinal conditional (κ) functions [6, 21]. However, our assignment of plausibilities to situations is fixed, whereas the plausibility assigned to a world using a κ-function can change when revisions occur. The dynamics of belief in our framework derives from the dynamics of the B relation, rather than that of the plausibility assignment.

In Darwiche and Pearl's framework [6], the κ-ranking of a world that does not satisfy the formula in a revision increases by 1. However, if the world satisfies the revision formula in future revisions, the world's κ-ranking decreases, and if it decreases to 0, the world will help determine the beliefs of the agent. In our framework, when a sensing action occurs, any situation S' that disagrees with the actual value of the sensor is *removed* from the B relation (actually, its successor is removed). The successors of S' will never be readmitted to B, so they will never help determine the beliefs of the agent.

One may think that having a fixed plausibility assignment limits the applicability of our approach. Consider an example[11] where, most plausibly, a cat is asleep at home, but where after phoning home, most plausibly, the cat is awake. (Nothing is certain in either case.) This might seem to require adjustment of the plausibility assignment.

To handle this example, we need first to observe that in the action theory we are using, actions are taken to be *deterministic*, with effects described by successor state axioms, quite apart from properties of belief and plausibility. If in some situations a phone action wakes the cat, and in others not, then there has to be some property M such that we can write a successor state axiom of the following form:

$$\begin{aligned} &\textsc{Awake}(do(a,s)) \equiv (a = \textsc{phone} \wedge M(s)) \\ &\quad \vee [\ldots \text{other actions that can wake cats} \ldots] \\ &\quad \vee (\textsc{Awake}(s) \wedge [a \text{ is not some put-to-sleep action}]). \end{aligned}$$

For example, M could represent that "the phone's ringer is loud enough to wake the cat". With this model, we can then arrange the B relation in the initial situation so that there are 4 groups of situations s' B-related to S_0 where

[11] We are indebted to Jim Delgrande for this example.

the following hold (in order of decreasing plausibility): $M(s') \wedge \neg\textsc{Awake}(s')$, $M(s') \wedge \textsc{Awake}(s')$, $\neg M(s') \wedge \neg\textsc{Awake}(s')$, and $\neg M(s') \wedge \textsc{Awake}(s')$. Then we obtain:

$$\mathit{Bel}(\neg\textsc{Awake}, S_0)$$

but

$$\mathit{Bel}(\textsc{Awake}, \mathit{do}(\textsc{phone}, S_0))$$

as desired.[12] Of course, we also get that

$$\mathit{Bel}(M, S_0)$$

but this is to be expected: why would we think it most likely that the cat would be awake after the phone rings if we didn't also think it most likely that the ringer was loud enough to waken it? Thus, changing our minds about the plausibility of the cat being awake does not require us to change the plausibility ordering over situations.

We can also handle a belief-revision variant of this example where we change our mind about whether phoning home wakes the cat. For example, imagine a sensing action $\textsc{examineRinger}$ that informs us that M is false initially (*e.g.*, the ringer on the phone is set to low). Then, we get

$$\mathit{Bel}(\neg\textsc{Awake}, \mathit{do}(\textsc{phone}, \mathit{do}(\textsc{examineRinger}, S_0))).$$

In fact, in the process of developing the approach described in this paper, we experimented with various schemes where the plausibility assigned to situations could be updated. But we found that this led to problems for introspection. Consider a scheme where we combine the plausibility assignment with the belief accessibility relation by adding an extra argument to the B relation, i.e., where $B(s', n, s)$ means that in situation s the agent thinks s' is plausible to degree n. In order to ensure that beliefs are properly introspected, the relation would have to satisfy a constraint similar to the one given in Theorem 1, but taking plausibilities into account. That is to say, all the B-related situations to a situation s must have the same belief structure as s, i.e., they should be B-related to the same situations with the same plausibilities as s. Unfortunately, this conflicts with some of our intuitions about how to change plausibilities to accommodate new information.

Consider an example where we have two situations S_0 and S_1, and where initially the agent considers situation S_1 more plausible than S_0, i.e., $B(S_1, 0, S_0)$, $B(S_0, 1, S_0)$, $B(S_1, 0, S_1)$, $B(S_0, 1, S_1)$. Notice that S_0 and S_1 have the same belief structure. Suppose that $\textsc{Light}_1(S_0) \wedge \mathit{SF}(\textsc{senseLight}, S_0)$ and $\neg\textsc{Light}_1(S_1) \wedge \neg\mathit{SF}(\textsc{senseLight}, S_1)$ hold. The natural way to update the plausibilities after sensing would be to make the most plausible situations from a situation $\mathit{do}(\textsc{senseLight}, s)$ be the ones that agree with s on the value of $\mathit{SF}(\textsc{senseLight})$. So, if we let S_0' denote $\mathit{do}(\textsc{senseLight}, S_0)$ and S_1' denote $\mathit{do}(\textsc{senseLight}, S_1)$, then in S_0', S_0' should be more plausible than S_1' and in S_1', S_1' should be more plausible than S_0'. But this would violate the constraint that B-related situations have the same belief structure, and cause introspection to fail.

One way to avoid this problem would be to update the plausibilities of all situations based on what holds in the 'actual' situations, i.e., S_0 and its successors (this focuses attention on beliefs that hold in actual situations, which is what we normally do anyway). Friedman and Halpern [8] essentially use this approach. For the example above, we would look at how the plausibilities should change in S_0' and adjust the plausibilities in the situations B-related to S_0' (in this case just S_1') in the same way. We would then have that S_0' is more plausible than S_1' in both S_0' and S_1', i.e., $B(S_0', 0, S_0')$, $B(S_1', 1, S_0')$, $B(S_0', 0, S_1')$, $B(S_1', 1, S_1')$. Notice that S_0' and S_1' have the same belief structure, so the constraint violation mentioned above is resolved.

Unfortunately, under this new scheme we have a problem with beliefs about future beliefs. If we were to redefine *Bel* in the obvious way to accommodate the extra argument in B, our example would entail the very counterintuitive $\mathit{Bel}(\neg\textsc{Light}_1 \wedge \mathit{Bel}(\textsc{Light}_1, \mathit{do}(\textsc{senseLight}, \mathit{now})), S_0)$, i.e., in S_0, the agent believes that the light is not on but thinks that after sensing he will believe that it is on. Our approach—which uses a fixed plausibility ordering on situations and simply drops situations that conflict with sensing results from the B relation—avoids both of these problems.

Another interesting difference between our approach and many of the proposals for modifying the plausibility ordering [4, 6, 21, 22] is that they adopt orderings over possible worlds which do not contain a history of the actions that have taken place in the world. Our approach, on the other hand, is based on situations, which do have such histories. While Friedman and Halpern [8] do not adopt situations, their possible worlds (runs) do include a history.

6.2 Comparison with AGM and KM

In order to effect a comparison with established belief change frameworks—in particular, the AGM and KM frameworks—we need to first establish a common footing. The first notion to establish is what is meant by the epistemic (or belief) state of the agent. We define a belief state (relative to a given situation) to consist of those formulae believed true at a particular situation. We limit our attention to formulae uniform in a situation since the AGM and KM are state-based methods, and so there is no need to consider beliefs regarding more than one situation, i.e., situations other than the one currently under consideration.

[12] We can also handle a variant where nothing is believed about the cat sleeping initially by making the first two groups the most plausible.

Definition 8 *(K_t)*
Let t be a ground situation term. We denote a belief state at t by K_t and define it as follows:

$$K_t = \{\phi : \Sigma \models Bel(\phi,\ t) \text{ and } \phi \text{ is uniform in now}\}$$

It is easily verified that K_t is closed under deduction.

We first define a belief expansion operator ($K_t + \phi$), which returns the set of (uniform) formulae that the agent believes are implied by ϕ at t.

Definition 9 *($K_t + \phi$)*
Let t be a ground situation term and ϕ be a formula uniform in now. We denote the expansion of K_t with ϕ by $K_t + \phi$ and define it as follows:

$$K_t{+}\phi = \{\psi : \Sigma \models Bel(\phi \supset \psi,\ t) \text{ and } \psi \text{ uniform in now}\}$$

Next, we define the revision of K_t by A for ϕ ($K_t *_A \phi$) as the belief set held by the agent in the situation that results from performing revision action A for ϕ. In the AGM setting, a revision $K * \phi$ is interpreted as the revision of beliefs K *after learning* ϕ. In our case, we do not know whether ϕ will be true until after performing A. Accordingly, we define a revision of K_t by A for ϕ only in the case that ϕ happens to be true in situation t (i.e., $\phi[t]$ holds).

Definition 10 *($K_t *_A \phi$)*
Let t be a ground situation term, ϕ be a formula uniform in now, and A be a revision action for ϕ. We define the revision of K_t by A for ϕ to be

$$K_t *_A \phi = K_{do(A,\ t)}$$

whenever $\phi[t]$. If $\neg\phi[t]$, then $K_t *_A \phi$ is undefined.

We now state the relationship with the AGM theory.

Theorem 10 *Let t be a ground situation term, ϕ be a formula uniform in now, and A be a revision action for ϕ. If $K_t *_A \phi$ is defined, then it satisfies AGM postulates (K* 1)—(K* 4) and (K* 6).*

Notice that postulate (K* 5) is not satisfied because the agent will end up in inconsistency, if in t there are no B-related situations where ϕ holds. In our framework, the agent is also not capable of recovering from inconsistency. Once everything is believed possible at a situation (i.e., it has no B-related situations), there is no action that can be performed to remedy this. Also note that it does not make sense in our framework to sense (or, for that matter, try to bring about) a formula ϕ known to be necessarily false (i.e., $\models \neg\phi$).

Now, we take the update of a belief state K_t by update action A for formula ϕ to be the set of beliefs held by the agent in the situation that results from performing A. Recall that A causes ϕ to hold.

Definition 11 *Let t be a ground situation term, ϕ be a formula uniform in now, and A be an update action for ϕ. We define the update of K_t by A for ϕ to be:*

$$K_t \diamond_A \phi = K_{do(A,\ t)}$$

If no action makes ϕ true, then $K_t \diamond_A \phi$ is undefined.

The essential difference between the definitions of revision and update is that the former is effected by sensing (revision) actions while the latter by non-sensing (update) actions and the two are dealt with quite differently in our framework. We now compare with the KM theory.

Theorem 11 *Let t be a ground situation term, ϕ be a formula uniform in now, and A be an update action for ϕ. If $K_t \diamond_A \phi$ is defined, then it satisfies KM postulates (K$\diamond$1)—(K$\diamond$2) and (K$\diamond$4)—(K$\diamond$5).*

Notice that postulate (K$\diamond$3) is not satisfied because an update action for ϕ may have other effects, so despite the fact that the agent believes ϕ beforehand, we cannot guarantee that nothing will change. Boutilier [5] has a problem with this postulate ((U2) in the KM rendering) for similar reasons. In his framework, (update) actions have plausibilities, and the most plausible action explaining the new information is assumed to have taken place. It could be that this action has other effects. To satisfy this postulate, he introduces a *null event* and considers a model in which this is the most plausible event at any world.

In our framework, iterated revision corresponds to the performing of at least two consecutive revision actions. We now show that there is some correspondence with the Darwiche and Pearl account of iterated belief revision.

Theorem 12 *Let t be a ground situation term, ϕ and ψ be formulae uniform in now, A be a revision action for ϕ, and B be a revision action for ψ. Then if $*_A$ and $*_B$ are defined, they satisfy postulates (DP1), (DP3) and (DP4).*[13]

Interestingly, changes of the type described by (DP2) are not defined according to our view of belief revision. In the case where sensing ψ allows us to conclude $\neg\phi$, it is not defined to first sense for ψ and subsequently to sense for ϕ.

6.3 Previous Work

Belief change in the situation calculus has already been dealt with by Scherl and Levesque [19]. However, as noted previously, while they can handle belief update, they are limited to belief expansion. del Val and Shoham [7] also address the issue of belief change in the situation calculus, and their theory deals with both revision and update. However, they cannot represent nested belief and consequently cannot deal with the issues of belief introspection and mistaken belief.

[13] Applying Definition 10, we have that $(K *_A) *_B \psi = K_{do(B, do(A,t))}$ and $K *_B \psi = K_{do(B,t)}$.

There are a variety of frameworks that accommodate both belief revision and belief update. As noted, this is one strength of the proposal by del Val and Shoham [7]. In a more traditional belief change setting, Boutilier [3] also provides a general framework that allows for both these forms of change. However, this framework cannot deal with introspection in the object language. One approach that supports both belief revision and update and also handles introspection is Friedman and Halpern [8]. Their approach to revision and update is fairly standard, but set within a very general modal logic framework that combines operators for knowledge, belief (interpreted a using plausibility ordering), and time. But they do not discuss interactions between revision and update and introspection. We also think that it may suffer from some of the problems mentioned in Section 6.1 that prompted us to abandon approaches based on updating plausibilities.

7 Conclusions and Future Work

We have proposed an account of iterated belief change that integrates into a well-developed theory of action in the situation calculus [18]. This has some advantages, in that previous work on the underlying theory can be exploited for dealing with issues such as solving the frame problem, performing automated reasoning about the effects of actions, specifying and reasoning about complex actions, etc. Our framework supports the introspection of beliefs and ensures that the agent is aware of when it was mistaken about its beliefs. Our account of iterated belief change differs from previous accounts in that, for us, the plausibility assignment to situations remains fixed over time. The dynamics of belief derives from the dynamics of the B modality and of the domain-dependent fluents. We showed that our theory satisfies the majority of the AGM, KM, and DP postulates.

Our approach does have some limitations. In this paper, we have only looked at cases of belief change where the sensors are accurate, so that the agent only revises its beliefs by sentences that are actually true. It is the case that our successor state axiom for B ensures that the agent believes the output of its sensor after sensing. Also, our guarded sensed fluent axioms allow only hard (but context-dependent) constraints to be specified between the output of the sensor and the associated fluent; one cannot state that the sensor is only correct with a certain probability. However, we can also use beliefs to correlate sensor values to the associated fluents instead of guarded sensed fluent axioms. Thus, we could specify that the agent prefers histories where the sensors agree with the associated fluents more often to histories where they agree less often. We will explore this approach in future work. Note that Bacchus et al. [2] have a probabilistic account of noisy sensors in the situation calculus.

In Theorem 10, we saw that our framework captures some, but not all, of the AGM revision postulates. In particular, the agent may end up believing everything after a revision by a consistent formula ϕ, if none of its B-alternatives satisfies ϕ, violating (K*5). This, together with the fact that we never update the plausibility assignment, may suggest that our account has limited expressiveness. But we maintain that this is not the case. The example of Section 6.1 shows that we can handle some cases where a plausibility assignment update seems to be required. As well, we can construct theories where the (K*5) postulate does hold. This is done by ensuring that the B relation contains enough situations initially.[14] The need to ensure that enough epistemic alternatives are initially present if one wants to avoid inconsistency is not specific to our approach. In most frameworks, a similar issue arises with respect to revision by conjunctive observations. In future work, we will investigate the expressive limits of our framework.

We could also extend the framework by having multiple agents that act independently and impart information to each other. Instead of beliefs changing only through sensing, they would also change as a result of *inform* actions. Shapiro et al. [20] provide a framework for belief expansion resulting from the occurrence of inform actions in the situation calculus, which we would like to generalize to handle belief revision.

Lakemeyer and Levesque [12] incorporate the logic of *only knowing* into the Scherl and Levesque framework of belief update and expansion. The traditional belief (and knowledge) operator specifies formulae that are believed (or known) by the agent, but there could be others. The 'only knows' operator is used to describe *all* that the agent knows, i.e., a formula that corresponds exactly to the knowledge state of the agent. In future work, we would like to define an analogous 'only believes' operator that could be used to describe exactly what the agent believes in a framework that supports belief revision as well as belief expansion.

Acknowledgments

The authors would like to gratefully acknowledge Craig Boutilier and Jim Delgrande for valuable comments on an earlier version of this paper. Financial support for the Canadian authors was gratefully received from the Natural Sciences and Engineering Research Council of Canada, and the Institute for Robotics and Intelligent Systems of Canada. The Australian author is partially supported by a Macquarie University MUNS grant.

[14] Note also that when no guarded sense fluent axiom is applicable, the value of an *SF* fluent is unconstrained and can vary freely. There can be B-alternatives that differ only in the value of the *SF* fluent after some sequence of sensing actions has been performed. By modifying the definition of $*_A$, and ensuring that there are enough B alternatives, we *can* model iterated revisions involving contradictory information where the agent's beliefs remain consistent, i.e., where $K * \phi * \neg\phi \not\models \bot$.

References

[1] Carlos E. Alchourrón, Peter Gärdenfors and David Makinson. On the the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50: 510–533, 1985.

[2] Fahiem Bacchus, Joseph Y. Halpern, and Hector J. Levesque. Reasoning about noisy sensors and effectors in the situation calculus. *Artificial Intelligence*, 111(1–2):171–208, 1999.

[3] Craig Boutilier. Generalized update: Belief change in dynamic settings. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI-95)*, 1995.

[4] Craig Boutilier. Iterated revision and minimal change of conditional beliefs. *Journal of Philosophical Logic*, 25(3):262–305, 1996.

[5] Craig Boutilier. Abduction to plausible causes: An event-based model of belief update. *Artificial Intelligence*, 83(1):143–166, 1996.

[6] Adnan Darwiche and Judea Pearl. On the logic of iterated belief revision. *Artificial Intelligence*, 89(1–2):1–29, 1997.

[7] Alvaro del Val and Yoav Shoham. A unified view of belief revision and update. *Journal of Logic and Computation*, 4:797–810, 1994.

[8] Nir Friedman and Joseph Y. Halpern. A knowledge-based framework for belief change, part II: Revision and update. In *Proceedings of the Fourth International Conference on Principles of Knowledge Representation and Reasoning (KR-94)*, pages 190–201, 1994.

[9] P. Gärdenfors. *Knowledge in Flux*. The MIT Press, Cambridge, MA, 1988.

[10] Giuseppe De Giacomo and Hector J. Levesque. Progression using regression and sensors. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI-99)*, pages 160–165, 1999.

[11] Hirofumi Katsuno and Alberto O. Mendelzon. On the difference between updating a knowledge base and revising it. In Peter Gärdenfors, editor, *Belief Revision*, pages 183–203, Cambridge University Press, 1992.

[12] Gerhard Lakemeyer and Hector J. Levesque. $\mathcal{AOL}$: a logic of acting, sensing, knowing, and only knowing. In *Proceedings of the Sixth International Conference on Principles of Knowledge Representation and Reasoning (KR-98)*, pages 316–327, 1998.

[13] Hector J. Levesque. What is planning in the presence of sensing? In *Proceedings of the Thirteenth National Conference on Artificial Intelligence (AAAI-96)*, pages 1139–1146, August 1996.

[14] David Lewis. *Counterfactuals*. Harvard University Press, Cambridge, Massachusetts, 1973.

[15] John McCarthy and Patrick J. Hayes. Some philosophical problems from the standpoint of Artificial Intelligence. In Bernard Meltzer and Donald Michie, editors, *Machine Intelligence 4*. Edinburgh University Press, 1969.

[16] Robert C. Moore. A formal theory of knowledge and action. In Jerry R. Hobbs and Robert C. Moore, editors, *Formal Theories of the Common Sense World*, pages 319–358. Ablex Publishing, Norwood, NJ, 1985.

[17] Pavlos Peppas, Abhaya C. Nayak, Maurice Pagnucco, Norman Y. Foo, Rex Kwok, and Mikhail Prokopenko. Revision vs. update: Taking a closer look. In W. Wahlster, editor, *Proceedings of the Twelfth European Conference on Artificial Intelligence (ECAI-96)*, pages 95–99, August 1996.

[18] 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.

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

[20] Steven Shapiro, Yves Lespérance, and Hector J. Levesque. Specifying communicative multi-agent systems. In Wayne Wobcke, Maurice Pagnucco, and Chengqi Zhang, editors, *Agents and Multi-Agent Systems — Formalisms, Methodologies, and Applications*, volume 1441 of *LNAI*, pages 1–14. Springer-Verlag, Berlin, 1998.

[21] Wolfgang Spohn. Ordinal conditional functions: A dynamic theory of epistemic states. In W. Harper and B. Skyrms, editors, *Causation in Decision, Belief Change, and Statistics*, pages 105–134. Kluwer Academic Publishers, Dordrecht, 1988.

[22] Mary-Anne Williams. Transmutations of knowledge systems. In *Proceedings of the Fourth International Conference on Principles of Knowledge Representation and Reasoning (KR-94)*, pages 619-629, 1994.