Belief Change with Noisy Sensing and Introspection*

Steven Shapiro
Computer Science Inst.
University of Leipzig
Leipzig 04109
Germany
shapiro@informatik.uni-leipzig.de

Abstract

We model belief change due to noisy sensing, and belief introspection in the framework of the situation calculus. We give some properties of our axiomatization and show that it does not suffer from the problems with combining sensing, introspection, and plausibility update described in Shapiro *et al.* [2000].

1 Introduction

In this paper, we generalize the framework of Shapiro *et al.* [2000], where belief change due to sensing was combined with belief introspection in the situation calculus. In that framework, sensing was assumed to be infallible and the plausibilities of alternate situations (i.e., possible worlds) were fixed in the initial state, never to be updated. Here, we relax both assumptions. That is, we model noisy sensors whose readings may stray from reality and may return different values in subsequent readings. We also allow the plausibilities of situations to change over time, bringing the framework more in line with traditional models of belief change. We give some properties of our axiomatization and show that it does not suffer from the problems with combining sensing, introspection, and plausibility update described in Shapiro *et al.* In the next section, we present the situation calculus including the representation of beliefs, and Shapiro *et al.*'s framework. In Sec. 4, we present the formal details of our axiomatization of belief change. In Sec. 5, we present some properties of our axiomatization, and in Sec. 6, we conclude and discuss future work.

2 Situation Calculus

The basis of our framework for belief change is an action theory [Reiter, 2001] based on the situation calculus [McCarthy and Hayes, 1969], and extended to include a belief operator [Scherl and Levesque, 1993]. 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. The term $do(a,s)$ 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 $Init(s) \stackrel{\text{def}}{=} \neg\exists a, s'.s = do(a, s')$.

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 $\text{InR}_1(s)$ 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 [Reiter, 2001], which provide a succinct representation for both effect axioms and frame axioms [McCarthy and Hayes, 1969]. For example, assume that there are only two rooms, R_1 and R_2, and that the action LEAVE takes the agent from the current room to the other room. Then, the successor state axiom for InR_1 is:[2]

$$\text{InR}_1(do(a,s)) \equiv \\ ((\neg\text{InR}_1(s) \wedge a = \text{LEAVE}) \vee (\text{InR}_1(s) \wedge a \neq \text{LEAVE})).$$

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

Moore [1985] defined a possible-worlds semantics for a modal logic of knowledge in the situation calculus by treating situations as possible worlds. Scherl and Levesque [1993] adapted the semantics to the action theories of Reiter [2001]. 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 [1996] introduced a predicate, $SF(a,s)$, to describe the result of performing the binary-valued sensing action a. $SF(a,s)$ holds iff the sensor associated with a returns

*We would like to thank an anonymous reviewer for helpful comments on an earlier version of this paper.

[1] Here we assume that there is a single agent, however it would not be difficult to generalize the framework to handle multiple agents.

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

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* [De Giacomo and Levesque, 1999]. For example, suppose that there are lights in R_1 and R_2 and that $\text{LIGHT}_1(s)$ ($\text{LIGHT}_2(s)$, resp.) holds if the light in R_1 (R_2, resp.) is on. Then:

$$\text{INR}_1(s) \supset (\mathit{SF}(\text{SENSELIGHT}, s) \equiv \text{LIGHT}_1(s))$$
$$\neg\text{INR}_1(s) \supset (\mathit{SF}(\text{SENSELIGHT}, s) \equiv \text{LIGHT}_2(s))$$

can be used to specify that the SENSELIGHT action senses whether the light in the room where the agent is currently located is on.

Shapiro *et al.* [2000] adapted Spohn's ordinal conditional functions [Spohn, 1988; Darwiche and Pearl, 1997] to the situation calculus by introducing plausibilities over situations using a function $\mathit{pl}(s)$ which returns a natural number representing plausibility of situation s. The lower the number, the more plausible the situation is considered. The plausibilities were fixed in the initial situation and were not allowed to change, i.e., they used this successor state axiom for *pl*: $\mathit{pl}(\mathit{do}(a, s)) = \mathit{pl}(s)$. They adopted Scherl and Levesque's [2003] successor state axiom for B:[3]

$$B(s'', \mathit{do}(a, s)) \equiv$$
$$\exists s'.B(s', s) \wedge s'' = \mathit{do}(a, s') \wedge (\mathit{SF}(a, s') \equiv \mathit{SF}(a, s)).$$

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. In other words, after doing a, the agent's beliefs will be expanded to include what the value of the sensor associated with a is in s. If a is a sensing action, the agent's beliefs will also include the property associated with a in the guarded sensed fluent axiom for a. If a is a physical action, then the agent's beliefs will also include the effects of a as specified by the successor state axioms.

Shapiro *et al.* defined the beliefs of the agent to be the formula true in the most plausible accessible situations:[4]

$$\mathit{Bel}_S(\phi, s) \stackrel{\text{def}}{=}$$
$$\forall s'[B(s', s) \wedge (\forall s''.B(s'', s) \supset \mathit{pl}(s') \le \mathit{pl}(s''))] \supset \phi[s'].$$

Shapiro *et al.* thus modelled belief change with infallible sensors. If the agent senses a property ϕ, and ϕ actually holds, then all the situations that satisfy $\neg\phi$ become inaccessible. For example, if the agent believes $\neg\phi$ and senses ϕ, then all the most plausible, accessible situations will become inaccessible. A new set of accessible situations will become most plausible, all of which satisfy ϕ, yielding belief in ϕ. However, Shapiro *et al.* did not allow for the possibility of the agent subsequently sensing $\neg\phi$.

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 [1999] consisting of: (1) successor state axioms for each fluent, and guarded sensed fluent axioms for each action, as discussed above; (2) unique names axioms for the actions, and domain-independent foundational axioms (we adopt the ones given by Levesque *et al.* [1998] which accommodate multiple initial situations, but we do not describe them further here); and (3) initial state axioms, which describe the initial state of the domain and the initial beliefs of the agent. In what follows, we will use Σ to refer to a guarded action theory of this form.

3 Belief Change

Before formally defining a belief operator in this language, we briefly review the notion of belief change. Belief change, simply put, aims to study the manner in which an agent's doxastic (belief) state should change when the agent is confronted by new information. In the literature,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 [Gärdenfors, 1988] is the prototypical example of belief revision while the KM framework [Katsuno and Mendelzon, 1991] 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.

4 Belief Change with Noisy Sensors

Shapiro *et al.* [2000] modelled belief change due to sensing but it was assumed that the sensors were always accurate. This is quite a strong assumption which we will relax here. If sensing is exact, then the sensors will never be contradicted and so belief revision is limited to revising the agents *initial* beliefs. But once an initial belief is corrected, it will never change again. In this context it seems reasonable to have a fixed plausibility relation. However, if the sensors can return different results over time, this approach will not work because after sensing two contradicting values for the same formula, the agent will have contradictory beliefs (i.e., an empty accessibility relation).

To model noisy sensing, we add another distinguished predicate $\mathit{SR}(a, s)$, which is similar to *SF* described previously. The idea is that while $\mathit{SF}(a, s)$ describes the property of the world *ideally* sensed by action a, the actual values returned by the sensor may not correspond exactly to the property described by *SF*. So, we will use $\mathit{SR}(a, s)$ to describe the

[3] 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.

[4] We use ϕ to denote a formula that may contain a distinguished situation constant, *Now*, as a placeholder for a situation, e.g., $\text{INR}_1(\mathit{Now})$. $\phi[s]$ denotes the formula that results from substituting s for *Now* in ϕ. Where the intended meaning is clear, we omit the placeholder.

value *actually* returned by the sensor associated with action a. Another way of describing *SR* is that it is the result of adding noise to the sensor described by *SF*. How to specify *SR* is still an unresolved issue. We want *SR* to be related to *SF* but perhaps only related by a stochastic relation. This problem is reserved for future work.

As with Shapiro *et al.*, we assume that the agent knows the history of actions it has taken. By that we mean the agent only considers a situation possible if it agrees with the history of actions in the actual situation.[5] We further assume that the agent has privileged access to its sensors. That is, after the agent reads its sensor, it knows the value of the sensor and it remembers the sequence of sensor readings it has made to date. That is, in addition to knowing the history of actions that have occurred, the agent knows the history of sensor readings it has taken, and it only considers possible those situations that agree with the actual situation on the history of sensor readings.

5 Axiomatization

To model plausibilities that can change, we dispense with the *pl* predicate used by Shapiro *et al.* [2000], and instead add a plausibility to the accessibility relation. So, $B(s', n, s)$ will denote that s' is considered a possible situation by the agent with plausibility n in situation s. As before, the lower the plausibility level the more plausible the agent considers the situation to be. The beliefs of the agent are determined by the situations with plausibility 0:

$$\mathit{Bel}(\phi, s) \stackrel{\text{def}}{=} \forall s'. B(s', 0, s) \supset \phi[s']$$

As previously mentioned, we have two further distinguished predicates: $\mathit{SF}(a, s)$ and $\mathit{SR}(a, s)$, both of which take an action and a situation as arguments. The former holds if the property ideally sensed by sensing action a holds in situation s, and the latter holds if the sensor associated with a actually returns the value 1 in s. We adopt the convention that if A is a non-sensing action, then $\forall s. \mathit{SF}(A, s) \wedge \mathit{SR}(A, s)$ holds.

The dynamics of the agent's beliefs are formalized by the successor state axiom for B:

Axiom 1

$$\begin{aligned} & B(s'', n'', do(a, s)) \equiv \\ & \exists s', n'. B(s', n', s) \wedge s'' = do(a, s') \wedge \\ & \qquad (\mathit{SR}(a, s') \equiv \mathit{SR}(a, s)) \wedge \mathit{Update}(n'', n', a, s', s), \end{aligned}$$

where $\mathit{Update}(n'', n', a, s', s)$ (defined below) holds if n'' is the updated plausibility level due to action a for situation s' whose plausibility with respect to s is n'.[6] In other words, s'' will be accessible from $do(a, s)$ with plausibility n'', if there exist s' and n' such that s' was accessible from s with plausibility n', s' and s agree on the value of the sensor associated with a, and n'' is the result of updating the plausibility of s' with respect to s due to a. Note that situations that disagree with s on the value of the sensor associated with a, (and those whose last action is not a) are discarded altogether from the accessibility relation. This means that the agent will never come to believe that it was mistaken about its sensor readings (or about the history of action occurrences).

We update the plausibilities as follows. We say a situation's (s) sensor reading is correct with respect to the sensor associated with a, if its *SF* and *SR* values agree, i.e., $\mathit{SR}(a, s) = \mathit{SF}(a, s)$. Those situations whose sensor readings are correct will have their plausibility levels decreased (i.e., they will become *more* plausible) and the others will have their plausibility levels increased. For concreteness, we will use Darwiche and Pearl's [1997] update function, but others are possible.

$$\mathit{Correct}(a, s') \stackrel{\text{def}}{=} \mathit{SR}(a, s') \equiv \mathit{SF}(a, s')$$

$$\begin{aligned} \mathit{Good}(s', n', a, s) \stackrel{\text{def}}{=} & B(s', n', s) \wedge (\mathit{SR}(a, s') \equiv \mathit{SR}(a, s)) \wedge \\ & \mathit{Correct}(a, s') \end{aligned}$$

$$\begin{aligned} \mathit{Min}(n, a, s) \stackrel{\text{def}}{=} & (\exists s^* \mathit{Good}(s^*, n, a, s)) \wedge \\ & \forall s', n'. \mathit{Good}(s', n', a, s) \supset n' \geq n \end{aligned}$$

$$\begin{aligned} & \mathit{Update}(n'', n', a, s', s) \stackrel{\text{def}}{=} \\ & \quad (\mathit{Correct}(a, s') \supset \exists n^*. \mathit{Min}(n^*, a, s) \wedge n'' = n' - n^*) \wedge \\ & \quad (\neg \mathit{Correct}(a, s') \supset n'' = n' + 1) \end{aligned}$$

In other words, the situations whose sensor readings are incorrect have their plausibilities increased by 1. The situations whose readings are correct are updated by subtracting the *Min* value, which is the lowest plausibility among the accessible situations that agree with the actual situation on the sensor reading and the sensor reading is correct. The result is that the agent believes that its sensor reading is correct, since this will hold in all the 0-plausibility situations.

Following Shapiro *et al.*, we want $B(s', n, s)$ to hold only if s and s' have the same histories. This means that the agent knows what actions have occurred. To enforce this, we need the following axiom which says that the situations accessible from an initial situation are also initial.

Axiom 2 $\mathit{Init}(s) \wedge B(s', n, s) \supset \mathit{Init}(s')$.

We have B as a relation so that we can exclude certain situations altogether. We can think of these situations as completely implausible. However, for the situations that are assigned some plausibility, we want their plausibility to be unique:

Axiom 3 $\mathit{Init}(s) \wedge B(s', n, s) \wedge B(s', n', s) \supset n = n'$.

To ensure that the agent has positive and negative introspection, we need to impose a constraint on the situations accessible from initial situations. As is well known (see, eg.,

[5] A treatment of exogenous actions that are hidden from the agent was given by Shapiro and Pagnucco [2004].

[6] *Update* could be a function, however we found it more convenient to formulate it as a relation.

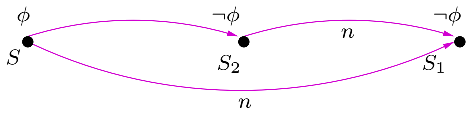

Figure 1: Introspection, exact sensing, and updating plausibilities clash

Fagin *et al.* [1995]), to get positive and negative introspection in contexts without plausibilities, it suffices for the accessibility relation to be transitive and Euclidean. Our constraint is a generalization of the combination of transitivity and Euclideanness that takes plausibilities into account. To get positive and negative introspection, we only need the accessibility relation to be transitive and Euclidean over situations with plausibility 0. However, since we are dealing with a dynamic framework, situations with higher plausibility levels could later have plausibility 0, therefore we enforce these constraints over all plausibility levels.

Axiom 4

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

In other words, for initial s, if s' is accessible from s with some plausibility, then s and s' have the same accessible situations with the same plausibilities.

Shapiro *et al.* described a conflict in their framework between preserving this constraint and updating plausibilities which is illustrated in Fig. 1. In this example, S_2 is accessible from S with some unspecified plausibility, and S_1 is accessible from both S and S_2 with plausibility n. Note that this example satisfies the constraint described in Axiom 4 (if we assume the situations are all initial). Now recall that for Shapiro *et al.* sensing was assumed to be accurate. Therefore, if the agent senses ϕ, the plausibility level of S_1 with respect to S should increase because they disagree on the value of ϕ, whereas, the plausibility level of S_1 with respect to S_2 should decrease because they agree on the value of ϕ. Therefore, the generalization of the constraint described in Axiom 4 that omits the condition that s be initial will not be satisfied after sensing ϕ. This means that the agent may loose full introspection.

The problem here is that in S, the sensor says that ϕ holds, while in S_2, the sensor says that ϕ does not hold. So, loosely speaking, the agent in S_1 is being told to revise its beliefs with ϕ by S and with $\neg\phi$ by S_2. In our framework, this problem is avoided because all the situations that disagree with S on the value of the sensor will be dropped from the accessibility relation. In effect, the beliefs of the agent in all the surviving accessible situation are revised by the same formula. In other words, to avoid this problem, we (and Shapiro *et al.*) had to model agents that have privileged access to their sensors, i.e., they *know* the results of sensing. In the next section, we give a theorem which says that we have indeed avoided this problem.

Note that this problem only arises when beliefs are changed due to sensing (and the agent is introspective). When an agent senses ϕ, it is told *whether* ϕ holds. In the traditional belief change setting, the agent is informed *that* ϕ holds. The subtle, but crucial, difference is that in the former case, the content of the belief-producing action depends on the actual situation, but not in the latter. As we stated earlier, the problem illustrated in Fig. 1 is that in S, the sensor says that ϕ holds, while in S_2, the sensor says that ϕ does not hold. If we were to model informing instead of sensing using the action INFORM(ϕ), the value of ϕ in S and S_2 would be irrelevant. In both situations, the agents beliefs would be revised with ϕ. While there have been previous approaches to belief revision with unreliable observations, e.g., [Aucher, 2005; Bacchus *et al.*, 1999; Boutilier *et al.*, 1998; Laverny and Lang, 2004] almost all of them use informing as the belief-producing action rather than sensing. Bacchus *et al.* [1999] model sensing (also in the framework of the situation calculus) as the nondeterministic choice of inform actions, one for each possible value returned by the sensor, but they do not address introspection. We think that there may be a problem modelling sensing this way, in the presence of introspection about future beliefs. If an agent believes ϕ then it should also believe that it will believe ϕ after sensing ϕ. This would not seem to hold in an approach like Bacchus *et al.*'s. Furthermore, we think it is more natural to model sensing as a primitive action rather than a nondeterministic choice of actions.

One issue that remains to be resolved is how to ensure that there is always at least one accessible situation. Since we are modelling noisy sensing, the agent's sensors could say that ϕ holds and later say that ϕ does not hold. How do we then prevent the agent's beliefs from lapsing into inconsistency? We need to ensure that regardless of the history of sensing results, for each action a, there is always an accessible situation (but not necessarily a most plausible one) that agrees with actual situation on the value returned by the sensor associated with a, and that value is correct, i.e.:

$$\forall a, s \exists s', n.B(s', n, s) \wedge (\mathit{SR}(a, s') \equiv \mathit{SR}(a, s)) \wedge \mathit{Correct}(a, s').$$

We believe we can achieve this using an axiom similar to the one given by Lakemeyer and Levesque [1998], and we will investigate this in future work.

6 Properties

In this section, we give some properties of our axiomatization of belief change and show that it does not suffer from the problem discussed by Shapiro *et al.* Let Σ be the foundational axioms together with the axioms of the previous section. First, we can show that the constraints imposed on the initial state given in Axioms 3 and 4 are preserved over all sequences of actions.

Theorem 1

$$\Sigma \models \forall n, n', s', s.B(s', n, s) \wedge B(s', n', s) \supset n = n',$$
$$\Sigma \models \forall n, n'', s, s', s''.$$
$$B(s', n, s) \supset \forall s'', n''.B(s'', n'', s') \equiv B(s'', n'', s).$$

The latter property ensures that the agent always has full introspection, and shows that we do not suffer from the problem of combining sensing, introspection, and updating plausibilities discussed in the previous section.

Corollary 2

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

Shapiro *et al.* discussed a possible solution to their problem with updating plausibilities by setting the plausibility levels from all accessible situations to be the same as they are in the actual situation. However, they showed that this solution was unsatisfactory by giving an example using this scheme that entailed a counterintuitive property, namely, that the agent believes ϕ, but thinks that after sensing ϕ, it will believe $\neg\phi$. We can show that it is not possible to construct such an example in our framework that is reasonable. In particular, we show that in any such example, the agent believes that either its sensor is incorrect or that its beliefs will be inconsistent after sensing ϕ. The second alternative is clearly not reasonable. We would not want to model an agent that believes ϕ but also believes that after sensing ϕ its beliefs will become inconsistent. The first alternative does not make sense either because we are modelling agents that revise their beliefs according to what their sensors tell them. If the agent were to believe that its sensor is not correct, then it would not make sense to revise its beliefs according to what the sensor said. So, while the agent might be aware that its sensors are not always correct, we want to avoid situations where the agent actually believes that its sensor will return the wrong value. Accordingly, the next theorem says that if the agent believes ϕ and it thinks that it will believe $\neg\phi$ after sensing ϕ, and a is a sensing action for ϕ that does not change the value of ϕ if it holds initially, then the agent believes that either its sensor is incorrect or that its beliefs will be inconsistent after sensing ϕ.

Theorem 3

$$\begin{aligned}\Sigma \models \forall a, s. & \mathit{Bel}(\phi, s) \wedge \mathit{Bel}(\mathit{Bel}(\neg\phi, \mathit{do}(a, \mathit{Now})), s) \wedge \\ & (\forall s'. \mathit{SF}(a, s') \equiv \phi(s')) \wedge \\ & (\forall s'. \phi[s'] \supset \phi[\mathit{do}(a, s')]) \supset \\ & \quad \mathit{Bel}([\neg \mathit{Correct}(a, \mathit{Now}) \vee \\ & \qquad \mathit{Bel}(\mathit{FALSE}, \mathit{do}(a, \mathit{Now}))], s)\end{aligned}$$

Next, we show that the agent will revise its beliefs appropriately. If an action a (ideally) senses a property ϕ, and the sensor indicates that ϕ holds, then after sensing, the agent will believe that ϕ held before the sensing occurred. We first define what it means for ϕ to hold in the previous situation: $\mathbf{Prev}(\phi, s) \stackrel{\text{def}}{=} \exists a, s'. s = \mathit{do}(a, s') \wedge \phi[s']$.

Theorem 4

$$\begin{aligned}\Sigma \models \forall a, s. (\forall s'. \mathit{SF}(a, s') \equiv \phi[s']) \wedge \mathit{SR}(a, s) \supset \\ \mathit{Bel}(\mathbf{Prev}(\phi), \mathit{do}(a, s)).\end{aligned}$$

If the agent also believes that a does change the value of ϕ, then the agent will believe ϕ after doing a.

Since the basis of our framework is a theory of action, belief updates are handled naturally as resulting from physical actions. We show that (as with Shapiro *et al.*) belief updates are handled appropriately. If a is a physical action (i.e., *SF* and *SR* are identically true) and situation s has at least one accessible situation with 0-plausibility, and the agent believes that a causes ϕ' to hold if ϕ holds initially, then the agent will believe that ϕ' holds after doing a in s, if it believes that ϕ holds in s.

Theorem 5

$$\begin{aligned}\Sigma \models \forall a, s. & (\exists s'. B(s', 0, s)) \wedge \\ & (\forall s'. \mathit{SF}(a, s')) \wedge (\forall s'. \mathit{SR}(a, s')) \wedge \\ & \mathit{Bel}((\phi(\mathit{Now}) \supset \phi'(\mathit{do}(a, \mathit{Now}))), s) \wedge \mathit{Bel}(\phi, s) \supset \\ & \quad \mathit{Bel}(\phi', \mathit{do}(a, s))\end{aligned}$$

Finally, we can show that Shapiro *et al.*'s framework is a special case of ours. If we assume that sensing is always accurate, and for every action a, every situation has an accessible situation that agrees with it on the value of the sensor associated with a, then Shapiro *et al.*'s axioms for B and *pl* combined and translated into our notation follow from our axioms.

Theorem 6

$$\begin{aligned}\Sigma \models & (\forall a, s. \mathit{Correct}(a, s)) \wedge \\ & [\forall a, s \exists s'. B(s', 0, s) \wedge (\mathit{SR}(a, s') \equiv \mathit{SR}(a, s))] \supset \\ & \quad \forall a, s'', n'', s. B(s'', n'', \mathit{do}(a, s) \equiv \\ & \qquad \exists s'. B(s', n'', s) \wedge s'' = \mathit{do}(a, s') \wedge \\ & \qquad (\mathit{SF}(a, s') \equiv \mathit{SF}(a, s)).\end{aligned}$$

7 Conclusions and Future Work

In this paper, we introduced a framework for modelling belief change as a result of noisy sensing in the situation calculus, where the agent has full introspection of its beliefs. Our framework allows the updating of plausibilities of situations, and we showed that we resolved the difficulty with combining all these elements discussed by Shapiro *et al.* [2000], and to achieve this, we had to endow the agent with infallible knowledge of the results of its sensing. As previously mentioned, there are some issues that are as yet unresolved. One is how to specify the *SR* predicate. Another is how to ensure that there are enough situations with the right properties to prevent the agent's beliefs from lapsing into inconsistency. We would also like to investigate the extent to which our framework satisfies the standard belief change postulates [Darwiche and Pearl, 1997; Gärdenfors, 1988; Katsuno and Mendelzon, 1991]. Lastly, we would like to extend the framework to handle unreliable physical actions, i.e., physical actions whose outcomes may be different from those expected by the agent.

References

[Aucher, 2005] Guillaume Aucher. A combined system for update logic and belief revision. In Michael Wayne Barley and Nik Kasabov, editors, *Intelligent Agents and Multi-Agent Systems: 7th Pacific Rim International Workshop* *on Multi-Agents, PRIMA 2004, Auckland, New Zealand, August 8-13, 2004, Revised Selected Papers*, volume 3371 of *LNAI*, pages 1–17. Springer-Verlag, Berlin, 2005.

[Bacchus *et al.*, 1999] 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.

[Boutilier *et al.*, 1998] Craig Boutilier, Nir Friedman, and Joseph Y. Halpern. Belief revision with unreliable observations. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI-98)*, pages 127–134, 1998.

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

[De Giacomo and Levesque, 1999] 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.

[Fagin *et al.*, 1995] Ronald Fagin, Joseph Y. Halpern, Yoram Moses, and Moshe Y. Vardi. *Reasoning about Knowledge*. MIT Press, 1995.

[Gärdenfors, 1988] Peter Gärdenfors. *Knowledge in Flux*. The MIT Press: Cambridge, MA, 1988.

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

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

[Laverny and Lang, 2004] Noël Laverny and Jérôme Lang. From knowledge-based programs to graded belief-based programs — Part I: On-line reasoning. In *Proceedings of the 16th European Conference on Artificial Intelligence (ECAI-04)*, pages 368–372, Amsterdam, 2004. IOS Press.

[Levesque *et al.*, 1998] Hector J. Levesque, Fiora Pirri, and Raymond Reiter. Foundations for a calculus of situations. *Electronic Transactions of AI (ETAI)*, 2(3–4):159–178, 1998.

[Levesque, 1996] 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, Portland, OR, August 1996.

[McCarthy and Hayes, 1969] 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.

[Moore, 1985] 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.

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

[Scherl and Levesque, 1993] 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, Washington, DC, July 1993. AAAI Press/The MIT Press.

[Scherl and Levesque, 2003] Richard B. Scherl and Hector J. Levesque. Knowledge, action, and the frame problem. *Artificial Intelligence*, 144(1–2):1–39, 2003.

[Shapiro and Pagnucco, 2004] Steven Shapiro and Maurice Pagnucco. Iterated belief change and exogenous actions in the situation calculus. In R. López de Mántaras and L. Saitta, editors, *Proceedings of the 16th European Conference on Artificial Intelligence (ECAI-04)*, pages 878–882, Amsterdam, 2004. IOS Press.

[Shapiro *et al.*, 2000] Steven Shapiro, Maurice Pagnucco, Yves Lespérance, and Hector J. Levesque. Iterated belief change in the situation calculus. In A. G. Cohn, Fausto Giunchiglia, and Bart Selman, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Seventh International Conference (KR2000)*, pages 527–538, San Francisco, CA, 2000. Morgan Kaufmann Publishers.

[Spohn, 1988] 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, 1988.