Iterated Belief Change and Exogenous Actions in the Situation Calculus

Steven Shapiro [1] and **Maurice Pagnucco** [2]

Abstract. We present a novel theory of action and change capable of dealing with discord between an agent's beliefs and the results of its sensing. Previous work by Scherl and Levesque [9] and Shapiro *et al.* [10] have given accounts of iterated belief update and revision, and the ability to deal with mistaken beliefs. However, they assume that all actions, including exogenous actions beyond the agent's control, are accessible to the agent for the purposes of reasoning. Our approach abandons this idealistic stance and allows the agent to hypothesise the occurrence of exogenous actions to account for any discrepancy between belief and sensing.

1 Introduction

In this paper, we introduce an approach to reasoning about action and change in the situation calculus [6, 8] capable of dealing with iterated belief change. More significantly, our approach is capable not only of handling mistaken beliefs when current beliefs do not accord with sensory readings but also of hypothesising the occurrence of exogenous actions to account for such beliefs. The basis of this approach is built upon foundations developed by Scherl and Levesque [9] and by Shapiro *et al.* [10]. Specifically, belief change in our account is handled through the same mechanism as change to other fluents, and thus among other things, we inherit a solution to the frame problem.

As in Scherl and Levesque [9], the approach developed here is capable of dealing with both belief *update* (in the sense of Katsuno and Mendelzon [4]) and *expansion* and provides for belief *introspection*. Following Shapiro *et al.* [10], we are also able to handle belief *revision* and mistaken beliefs. A natural byproduct of these approaches is an account of iterated belief change.

One shortcoming of these approaches is that they assume all actions are accessible to the reasoner. That is, the reasoner is aware of their occurrence. This idealisation is largely unattainable in most scenarios. Certainly, a reasoner has access to the actions that it performs. On the other hand, actions performed by other agents or by the environment itself (i.e., nature) are not always accessible to the reasoner. In fact, one can argue that the only way in which such actions become accessible to a reasoner is through sensing of its environment.

The main contribution of this paper is to introduce an account of reasoning about action and change in the situation calculus in which sensing that does not accord with beliefs can be resolved in one of two ways: deciding a belief in the initial situation was in error (mistaken belief), or hypothesising the occurrence of a sequence of exogenous actions that would account for the sensing results. While the account of iterated belief change by Shapiro *et al.* can deal with mistaken beliefs it suffers from the inability to suggest the occurrence of exogenous actions to explain the discord between beliefs and observation. In fact, the only way to deal with such discord is to appeal to mistaken belief or stumble into inconsistency. As with Shapiro *et al.*, our approach assumes the integrity of the reasoner's sensors. That is, sensing is correct. Our point of departure is the ability to hypothesise the occurrence of exogenous actions to correct discrepancies between belief and sensing. We do so through a notion of minimal change applied to histories of action occurrence.

In the next section, we briefly review the situation calculus including the Scherl and Levesque [9] model of belief expansion, explaining the notions of belief revision, belief update and iterated belief change. Section 3 motivates and defines a new belief operator as a small modification to the one used by Scherl and Levesque. In Section 4, we show the operator in action on a simple example, and how an agent can change its mind repeatedly. In Section 5, we prove some properties of this operator, justifying the points made above. In the final section, we draw some conclusions and discuss future work.

2 Background

2.1 Situation Calculus

The basis of our framework for belief change is an action theory [8] based on the situation calculus [6], and extended to include a belief operator in [9]. The situation calculus is a predicate calculus language for representing dynamically changing domains. A situation represents a possible state of the domain. There is a set of initial situations corresponding to the ways the agent 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 doing 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 sequence of actions that produces a situation is called the *history* of the situation. $s \preceq s'$ ($s \prec s'$, resp.) means there is a (nonempty, resp.) path from situation s to situation s'. The initial situations are defined as those having an empty history:

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

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 $\textsc{InRoom}(s)$ to represent the fact that the agent is in the room in situation s. The effects of actions on fluents are defined using successor state axioms [8], which provide a

[1] School of Computer Science and Engineering, The University of New South Wales, NSW 2052, Australia. Email: steven@cse.unsw.edu.au

[2] ARC Centre of Excellence for Autonomous Systems, School of Computer Science and Engineering, The University of New South Wales, NSW 2052, Australia. Email: morri@cse.unsw.edu.au

succinct representation for both effect axioms and frame axioms [6]. For example, here is the successor state axiom for $\textsc{InRoom}$:[3]

$$\textsc{InRoom}(\mathit{do}(a,s)) \equiv (a = \textsc{enter} \vee (\textsc{InRoom}(s) \wedge a \neq \textsc{leave})).$$

This axiom asserts that the agent will be in the room after doing an action iff either the action is entering the room, or the agent is currently in the room and the action is not leaving the room.

Moore [7] defined a possible-worlds semantics for a logic of knowledge in the situation calculus by treating situations as possible worlds. Scherl and Levesque [9] adapted this to Reiter's action theories [8]. 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 convention in modal logic for accessibility relations.

Levesque [5] 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 value of the sensor associated with a is 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* [2]. For example, the axiom:

$$\textsc{InRoom}(s) \supset (\mathit{SF}(\textsc{senseLight}, s) \equiv \textsc{Light}(s))$$

can be used to specify that if the agent is in the room in situation s, then the $\textsc{senseLight}$ action senses whether the light is on in s. If the agent is not in the room in s, then nothing can be said about the sensor, i.e., its value is arbitrary. If a is a non-sensing (i.e., physical action) $\mathit{SF}(a, s)$ is true by definition.

Scherl and Levesque [9] gave a successor state axiom for B that states how actions, including sensing actions, affect agent beliefs.[4]

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

The situations s'' that are B-related to $\mathit{do}(a, s)$ are the ones that result from doing a in a situation s', such that the sensor associated with 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. The agent's beliefs will therefore include the property associated with a in the guarded sensed fluent axiom for a, and the physical effects of a as specified by the successor state axioms.

To describe the domain in the example to be given in Section 4, we use an action theory of the same form as the one described in [2]. That is, in addition to successor state axioms[5] and guarded sensed fluent axioms, discussed above, we have initial state axioms, which describe the initial state of the domain and the initial beliefs of the agent. These are axioms that only talk about initial situations. We also need foundational axioms, which are domain independent — including axioms about natural numbers — and unique names axioms for the primitive actions, but we omit them due to space constraints.

2.2 Belief Change

Simply put, belief change studies the manner in which an agent's epistemic (belief) state should change when the agent receives new information. In the literature,[6] there is often a clear distinction between two forms of belief change: *revision* and *update*. Both forms can be characterised by an axiomatic approach (in terms of rationality postulates) or through various constructions (e.g., epistemic entrenchment, possible worlds, etc.). The AGM theory [3] is the standard example of belief revision while the KM framework [4] is identified with belief update.

Due to lack of space we shall not present postulates or constructions. For the purposes of this paper, it is sufficient to note that belief revision is appropriate for modelling static environments about which the agent does not have full information. New information is used to fill in these gaps, but the environment itself does not change. Belief update, on the other hand, is intended for situations in which the environment itself is changing due to the performing of actions. New information results from actions and indicates a potential change in the environment. One of the major issues in this area is that of *iterated belief change* [1], i.e., modelling how the agent's beliefs change after multiple belief revisions or updates occur.

3 Definition of the Belief Operator

Scherl and Levesque [9], define a modal operator for belief in terms of an accessibility relation on situations ($B(s', s)$). In [9], the believed sentences are the ones true in all accessible situations, i.e.:

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

The problem with this framework is that the agent cannot change its mind. Once a proposition is believed, it is believed thereafter. Shapiro *et al.* [10], extended this framework to handle belief change. They added a function that describes how *plausible* the agent considers a situation to be. The beliefs of the agent were those formulae that were true in all *most plausible* accessible situations. As sensing occurs and situations are dropped from the accessibility relation, a new set of situations can become most plausible and therefore beliefs of the agent can change to contradict previous beliefs. In that framework, sensing was assumed to be accurate and there were no exogenous actions. If the agent senses the same formula more than once and gets different answers, the agent's beliefs will become inconsistent. In this paper, we relax the second constraint. We assume that sensing is accurate, but we allow multiple agents to change the world.[7] We divide the actions into *endogenous*, i.e., ones that are performed by the agent, and *exogenous*, i.e., ones that are performed by other agents. The agent is directly aware of the endogenous actions but not the exogenous actions; it can only become aware of exogenous actions indirectly by sensing their effects. The occurrence of exogenous actions does not directly affect the mental state of the agent, therefore *SF* should be identically true for these actions:

Axiom 1 $\mathit{Exo}(a) \supset \forall s.\mathit{SF}(a, s)$,

where $\mathit{Exo}(a)$ denotes that a is an exogenous action.

The ordering of situations takes into account not only the initial plausibility of a situation as before, but also the history of actions of a situation. We achieve this by relaxing the constraint in Shapiro *et al.* [10] that accessible situations have the same histories. However, we still require that histories of accessible situations have the same

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

[4] For simplicity, we leave out any concerns about when actions can be executed from this and subsequent axioms. That is, we make no mention of the *Poss* predicate [8]. We simply assume that all actions are always executable.

[5] We could use the more general *guarded successor state axioms* of [2], but regular successor state axioms suffice for the simple domain we consider.

[6] We shall restrict our attention to approaches in the AGM [3, 4] style although there are many others.

[7] However, we continue to model the mental state of only one agent, which we will continue to call *the agent*.

endogenous actions in the same order (but can have additional sequences of exogenous actions interspersed among endogenous ones).

As in Shapiro *et al.*, plausibility is assigned to situations using a function $\mathit{pl}(s)$, whose range is the natural numbers. $\mathit{pl}(s)$ indicates how plausible the agent thinks s is — the lower the value the more plausible the situation. The *pl* function only has to be specified over initial situations, using an initial state axiom. The plausibility of successor situations is inherited from their predecessors using the following successor state axiom:

Axiom 2 $\mathit{pl}(\mathit{do}(a,s)) = \mathit{pl}(s)$.

We define the *height* of a situation s to be the number of actions in the history of s. For example, $\mathit{height}(S_0) = 0$ because S_0 is an initial situation. We also define the *root* of a situation s to be the (unique) initial situation in the history of s. E.g., $\mathit{root}(\mathit{do}(\text{LEAVE}, S_0)) = S_0$. We omit formal definitions of these functions due to space constraints.

We want the agent to hypothesise the occurrence of exogenous actions only when necessary. Therefore, the agent will prefer situations that have less exogenous actions. Situation s is preferred to s' ($s \sqsubseteq s'$), if s is more plausible than s', or if s and s' are equally plausible then the one with the least exogenous actions in its history will be preferred:

$$s' \sqsubseteq s \stackrel{\text{def}}{=} \mathit{pl}(s') < \mathit{pl}(s) \vee (\mathit{pl}(s') = \mathit{pl}(s) \wedge \mathit{height}(s') \leq \mathit{height}(s))$$

We will only be comparing situations that are B-related and so have the same endogenous actions in their histories, therefore the height of the situation suffices to measure the number of exogenous actions in the history.

$\mathit{Exo}(a)$ holds if a is an exogenous action, and we assume this predicate is defined by the user. For convenience, we define $\mathit{Endo}(a) \stackrel{\text{def}}{=} \neg \mathit{Exo}(a)$.

We say that s, s' define an exogenous sequence of actions, $\mathit{ExoSeq}(s, s')$, if s precedes s' and only exogenous actions occur between s and s':

$$\mathit{ExoSeq}(s, s') \stackrel{\text{def}}{=} s \preceq s' \wedge \forall a, s_1 . s \prec \mathit{do}(a, s_1) \preceq s' \supset \mathit{Exo}(a).$$

$\mathit{LastEndo}(a, s, s')$ holds if a is the *last endogenous action* in the sequence defined by s and s':

$$\mathit{LastEndo}(a, s, s') \stackrel{\text{def}}{=} \mathit{Endo}(a) \wedge \mathit{ExoSeq}(\mathit{do}(a, s), s').$$

The successor state axiom for B is more complex than in past approaches. We want a situation s' to be accessible from s iff s' and s have the same endogenous actions in the same order in their histories, and the sensing results of the endogenous actions were the same.

Axiom 3

$$\begin{aligned} B(s', s) \equiv {} & [(\mathit{ExoSeq}(\mathit{root}(s), s) \wedge \mathit{ExoSeq}(\mathit{root}(s'), s')) \vee \\ & (\exists s_1', s_1, a . \mathit{LastEndo}(a, s_1', s') \wedge \mathit{LastEndo}(a, s_1, s) \wedge \\ & (\mathit{SF}(a, s_1') \equiv \mathit{SF}(a, s_1)) \wedge B(s_1', s_1))]. \end{aligned}$$

In other words, if there are only exogenous actions in the history of s and the same is true of s' then $B(s', s)$ holds. Otherwise, if the last endogenous action in the history of s is a (and it occurred in situation s_1), the last endogenous action in the history of s' is also a (and it occurred in situation s_1'), the sensing result of a in s_1' is the same as in s_1, and s_1' is accessible from s_1, then $B(s', s)$ holds.

Note that this successor state axiom for B differs from previous ones in that it specifies the accessible situations from every situation *including* the initial ones. Previously, the axiomatizer was allowed to specify which situations were accessible from initial situations. Here, the axiomatizer only has to specify the initial plausibilities of the situations to specify the agent's beliefs. Let Σ be Axioms 1, 2, 3, together with the foundational axioms. It can be formally shown that these axioms are consistent. We also have:

Theorem 1 *Σ entails that B is an equivalence relation.*

Note that this does not mean that we have an S5 logic, since not all the B-related situations are used to determine the beliefs of the agent, but only the *minimal* ones. As we will see later, this logic is weak S5.

We say that situation s' is *minimal* with respect to situation s if in all situations s'' accessible from s, s' is preferred to s'':

$$\mathit{Min}(s', s) \stackrel{\text{def}}{=} \forall s'' . B(s'', s) \supset s' \sqsubseteq s''.$$

We are now in a position to define the beliefs of the agent. Since ϕ will usually contain fluents, we introduce a special symbol now as a placeholder for the situation argument of these fluents, e.g., $\mathit{Bel}(\text{INROOM}(now), s)$. $\phi[s]$ 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., $\mathit{Bel}(\text{INROOM}, s)$. We say that a formula is *uniform* in s iff s is the only situation term in that formula.

We will say that the agent believes ϕ in situation s, if ϕ holds in the minimal B-related situations:

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

This definition of belief entails that exogoneous actions do not affect the beliefs of the agent:

Theorem 2 *Let ϕ be a formula uniform in s. Then,*

$$\Sigma \models \forall a, s . \mathit{Exo}(a) \supset (\mathit{Bel}(\phi, s) \equiv \mathit{Bel}(\phi, \mathit{do}(a, s))).$$

4 Example

The following example illustrates the two ways in which our approach deals with discord between belief and sensing. Both arise as a natural result of our notion of minimal change applied to action histories. The first is to suggest that a belief in the initial situation was in error. The second is to hypothesise the occurrence of a sequence of exogenous actions to account for the discrepancy between belief and perception.

The scenario we consider, depicted below in Figure 1, is one in which there is a room with a light. In the initial situation S_0, the agent is not in the room ($\neg$INROOM) and the light is off ($\neg$LIGHT). This situation is shown B-related (the dashed line) to three other situations: S_0', S_0'' and S_0''' (to keep the exposition simple, we do not show all B-related situations). The most plausible of these is S_0' (note that plausibility levels are indicated on the left axis). Minimal situations are shown by shaded circles and non-minimal ones by open circles. Actual situations are shown filled. Situation S_0' has a plausibility level of 0 (i.e., $\mathit{pl}(S_0') = 0$) while situations S_0'' and S_0''' have plausibility level 1. Therefore, the agent believes that it is not in the room ($\mathit{Bel}(\neg\text{INROOM}, S_0)$) but that the light is on ($\mathit{Bel}(\text{LIGHT}, S_0)$).

The agent performs the endogenous action ENTER; a physical action leading to a belief update. The B relation is simply projected forwarded to the successor situations. Our figure simplifies the picture by omitting situations in which exogenous actions occur in order to avoid clutter. However, one should keep in mind that such situations would be B-related to S_1. We have shown situations S_1', S_1'' and S_1''' only. Of these S_1' is minimal (preferred) and so the agent believes LIGHT and INROOM in S_1.

At this point the agent performs the endogenous sensing action SENSELIGHT; leading to belief revision. According to the sensed fluent axiom for SENSELIGHT, if the agent is in the room in a situation $S^{\#}$, then $\mathit{SF}(\text{SENSELIGHT}, S^{\#})$ holds iff the light is on. In our example, the light is off in situations S_1, S_1'' and S_1''' and on in situation S_1'. The successor state axiom for B dictates 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 our example S_1 is the actual situation. Since S_1 and S_1' disagree on the value of *SF* for SENSELIGHT, $\mathit{do}(\text{SENSELIGHT}, S_1')$ is not B-related to $S_2 = \mathit{do}(\text{SENSELIGHT}, S_1)$. S_1'' and S_1''' agree with S_1 on *SF* for SENSELIGHT and so $S_2'' = \mathit{do}(\text{SENSELIGHT}, S_1'')$ and $S_2''' = \mathit{do}(\text{SENSELIGHT}, S_1''')$ are both B-related to S_2. As a consequence, the agent believes that it is in the room and the light is off in situation S_2. Thus the agent is able to correct its mistaken belief.

Subsequently the agent performs the endogenous action LEAVE, leading to a belief update as described above. Unbeknownst to the agent, another agent, Agt_1, turns the light on by performing the exogenous SWITCH(Agt_1) action leading to situation S_4.[8] S_4 is B-related to S_4''' since they agree on the endogenous actions performed by the agent. However, S_4 is not B-related to S_4'' as the last endogenous action in the former case is LEAVE whereas in the latter it is ENTER. Note that the beliefs of the agent do not change as a result of the exogenous SWITCH(Agt_1) action, since the minimal, accessible situations are the same as before the action.

The agent now re-enters the room by performing the endogenous ENTER action leading to situation S_5 in the actual world and a belief update by projecting forward to successor situations. S_5 is B-related to S_4'' and S_5''' since they agree on all endogenous actions performed. Situation S_4'' is minimal so now the agent believes that it is in the room and that the light is still off. Performing an endogenous SENSELIGHT sensing action, the agent now realises that the light is on. The agent is now in situation S_6 which is B-related to S_6''' since they agree on the endogenous actions performed and also on the value of *SF* for SENSELIGHT. S_6 and S_5'' also agree on the endogenous actions but they disagree on the value of *SF* for SENSELIGHT and so are not B-related. Also, S_6''' is minimal because it hypothesises only one additional exogenous action; that another agent has executed the SWITCH(Agt_1) action to turn on the light. In S_6, the agent now believes that it is in the room and the light is on (and that a SWITCH(Agt_1) action occurred in the past).

5 Properties

5.1 Belief Revision

We say that in situation s formula ϕ previously held, if ϕ held before the last endogenous action:

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

We now show that belief revisions are handled correctly in our system. Suppose an endogenous sensing action A is performed in situation S, and A is a sensing action for ϕ, i.e., $\forall s'.\mathit{SF}(A, s') \equiv \phi(s')$. If the sensor indicates that ϕ holds in S, then after performing A, the agent will believe ϕ held before A was done.

Theorem 3 *Let ϕ be a formula uniform in s. Then,*

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

If the agent does not believe ϕ or $\neg\phi$ in S, then 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 the theory in [9]. However, in [9], 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 not be inconsistent in this case, in fact, the agent will never lapse into inconsistency.

Theorem 4 $\Sigma \models \forall s \neg \mathit{Bel}(\mathit{FALSE}, s)$

5.2 Belief Update

We also show that, as in [9], the agent's beliefs are updated correctly when non-sensing actions occur. Suppose A is not a sensing action. This means that $\forall s.\mathit{SF}(A, s)$ holds. In this case, the accessible situations are simply projected forward to take into account the performing of A, none are dropped. Also, because of the way we defined minimality over situations, the agent does not hypothesise any exogenous actions occurring after A. So, after performing A, the agent knows that A was the last action performed.

Theorem 5

$$\Sigma \models \forall a, s, s''.\mathit{Endo}(a) \land (\forall s' \mathit{SF}(a, s')) \land B(s'', \mathit{do}(a, s)) \land \mathit{Min}(s'', \mathit{do}(a, s)) \supset \exists s'.s'' = \mathit{do}(a, s') \land B(s', s) \land \mathit{Min}(s', s).$$

Now suppose that the agent believes ϕ in S, and that A is a non-sensing action that causes ϕ' to hold, if ϕ holds beforehand. Then after performing A in S, the agent ought to believe that ϕ' holds:

Theorem 6 *Let ϕ and ϕ' be formulae uniform in s. Then,*

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

5.3 Introspection

The agent has full introspection of its beliefs:

Theorem 7 *Let ϕ be a formula uniform in s. Then,*

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

This theorem, together with Theorem 4, shows that our logic is weak S5 (**KD45**). However, we do not have **T**, since while every situation is B-related to itself, it will not in general be minimal wrt to itself.

[8] Again, for simplicity, the SWITCH(Agt_1) action is the only one that takes an argument. We assume that the other actions can only be performed by our agent. For example, agent Agt_1 might be a "timer" that turns the light on at a particular instant.

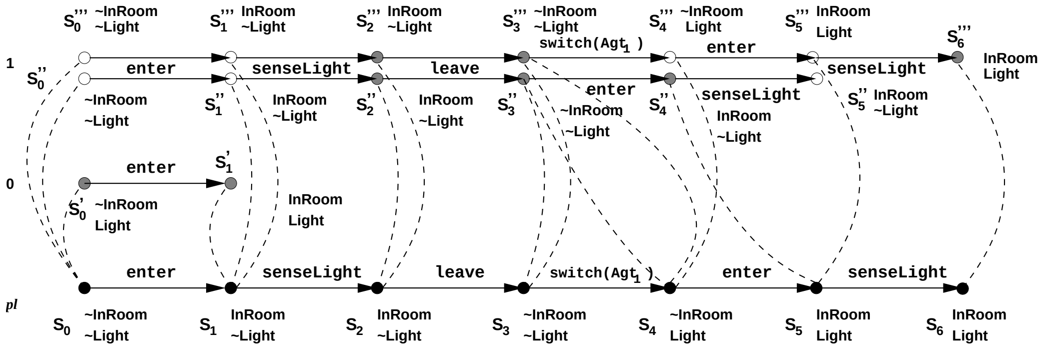

Figure 1. Illustrative example

5.4 Awareness of Mistakes

The agent also has introspection of its past beliefs. 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 $do(A, S)$, 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 be false, then there is no reason for the agent to believe that ϕ was false in the last situation. In the theorem, we rule out that case by stating in the antecedent that after doing A, the agent believes that if ϕ held before the last endogenous action then it continues to hold now.

Theorem 8 *Let ϕ be a formula uniform in s. Then,*

$$\begin{aligned}\Sigma \models \forall a, s.\mathit{Endo}(a) \wedge \mathit{Bel}(\phi, s) \wedge \mathit{Bel}(\neg\phi, \mathit{do}(a, s)) \wedge\\ \mathit{Bel}((\mathit{Previously}(\phi) \supset \phi), \mathit{do}(a, s)) \supset\\ \mathit{Bel}(\mathit{Previously}(\neg\phi \wedge \mathit{Bel}(\phi)), \mathit{do}(a, s))\end{aligned}$$

5.5 Exogenous Actions

We want the agent to refrain from hypothesising exogenous actions if it is unnecessary to do so, as is stated in the following theorem:

Theorem 9

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

This theorem says that if there is a minimal and accessible situation that agrees with s on the sensing result of (endogenous) a, then every s'' that is minimal and accessible from $do(a, s)$ is the result of doing a in a situation s' that is minimal and accessible from s. In other words, under these conditions, no new exogenous actions are hypothesised by the agent when a occurs.

5.6 Revision and Update Postulates

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

6 Conclusions and Future Work

We have presented a new theory of reasoning about action and belief change that is capable of dealing in a sensible way with the discord between the beliefs of an agent and the results of its sensing. Such discrepancies can be remedied either by suggesting that beliefs in the initial situation were in error or by hypothesising the occurrence of a sequence of exogenous actions. The resulting theory is capable of dealing with belief revision and belief update as well as iterated belief change, introspection of beliefs and awareness of mistakes.

It would be interesting to investigate other notions of minimal change. One possibility would be to take into account the likelihood of occurrence of exogenous actions given the agent's current beliefs. We would also like to develop a representation theorem that places constraints on minimal change that correspond exactly to the AGM postulates. It would also be useful to extend our framework to model beliefs of other agents in addition to their actions.

REFERENCES

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

[2] Giuseppe De Giacomo and Hector J. Levesque. Progression using regression and sensors. In *Proc. of the 16th International Joint Conference on Artificial Intelligence*, pages 160–165, 1999.

[3] Peter Gärdenfors. *Knowledge in Flux*. MIT Press, 1988.

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

[5] Hector J. Levesque. What is planning in the presence of sensing? In *Proc. of the 13th National Conf. on Art. Int.*, pages 1139–1146, 1996.

[6] John McCarthy and Patrick J. Hayes. Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer and D. Michie, editors, *Machine Intelligence 4*, 1969.

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

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

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

[10] Steven Shapiro, Maurice Pagnucco, Yves Lespérance, and Hector J. Levesque. Iterated belief change in the situation calculus. In A. G. Cohn, F. Giunchiglia, and B. Selman, editors, *Principles of Knowledge Rep. and Reasoning: Proc. of the 7th Int. Conf.*, pages 527–538, 2000.

[11] Y. Shoham. *Reasoning About Change: Time and Causation from the Standpoint of Artificial Intelligence*. MIT Press, 1988.