Proceedings of the 19th International Conference on Principles of Knowledge Representation and Reasoning
Main Track

Projection of Belief in the Presence of Nondeterministic Actions and Fallible Sensing

Jens Claßen[1], **James P. Delgrande**[2]
[1]Department of People and Technology, Roskilde University, Denmark
[2]School of Computing Science, Simon Fraser University, Burnaby, BC, Canada
classen@ruc.dk, jim@cs.sfu.ca

Abstract

We recently presented a Situation Calculus-based framework for modelling an agent who has incomplete or inaccurate knowledge about its environments, whose actions are nondeterministic, and whose sensors might give incorrect results. Generalizing earlier proposals, the approach represented the agent's epistemic state by a set of situations ranked by their respective plausibility, and which would then be updated by modifying the plausibility ranks accordingly. Here we extend our earlier work by considering *projection* in this framework, i.e. the question whether a certain (epistemic) formula will hold after a given sequence of actions. We present results on both *regression*, where the query is transformed into an equivalent one about the initial situation, as well as *progression*, where the knowledge base is updated to reflect the situation after executing the action sequence in question.

1 Introduction

In a recent paper (Claßen and Delgrande 2021), we presented a framework based on the Situation Calculus (McCarthy and Hayes 1969; Reiter 2001) for representing an agent that has incomplete or inaccurate knowledge about its environment, whose actions are non-deterministic, and whose sensors might be fallible. The approach generalizes earlier proposals (Shapiro et al. 2011; Delgrande and Levesque 2012; Delgrande and Levesque 2019) where the agent's epistemic state is presented by a set of situations ranked by their respective plausibility, and which would then be updated according to the executed action and its sensing outcome. Two notions of belief were distinguished, an extensional "bird's eye" view, and an intensional one from the agent's perspective. Here we extend this work by considering the *projection problem*, which means deciding whether a formula will hold after a given sequence of actions, which is in particular needed for the purpose of planning. We present results on both *regression*, where the query is transformed into an equivalent one about the initial situation, as well as *progression*, where the knowledge base is updated to reflect the situation after executing the action sequence in question. For brevity, we only discuss the extensional case, but it is straightforward to adapt our results to the intensional one.

As a motivating example, we adapt the extended litmus paper problem from (Hunter and Delgrande 2011). The original example is due to Moore (1985): There is a liquid in a beaker that is either an acid or a base, and the agent has a piece of litmus paper. After dipping the paper into the liquid and observing its colour, the agent will learn that the liquid is an acid, if the paper turned red, and a base, if the paper turned blue. In the extended version, the agent most plausibly *believes* that the paper is litmus, however also allows for the possibility that it is just a plain piece of white paper. After dipping it and observing that it remained white, the agent will revise its beliefs so that it now most plausibly believes that the paper is (and initially was) in fact not litmus paper.

We present preliminaries in Section 2, discuss regression in Section 3, progression in Section 4, and then conclude.

2 Preliminaries

2.1 The Situation Calculus

The Situation Calculus is a dialect of first-order logic, with some second-order features, for reasoning about action and change. There is a sort for *actions*, one for *situations*, and one for *objects* (everything else, including numbers). Situations describe possible sequences of actions, where a term $do(\alpha, \sigma)$ denotes the situation resulting from applying action α in situation σ, and S_0 is a constant representing the initial situation. We will use σ (possibly with decorations) to denote *terms* of sort situation, and s (possibly with decorations) as situation *variables*. Similarly, α and β will refer to action *terms*, whereas a is used as action variable. For convenience, we may use sequence notation for do where $do(\beta, do(\alpha, S_0))$ can be written as $do(\langle \alpha; \beta \rangle, S_0)$. Changing properties are represented by *fluents*, which are predicates that take a situation as their last argument, e.g. $Re(s)$ to express that the litmus paper is red in situation s. A formula is called *uniform in* σ if the only situation term it mentions is σ (ruling out quantification and equalities over situations).

2.2 Nondeterminism, Fallible Sensing, Belief

Here we only give a brief summary of our framework for nondeterministic actions, fallible sensing, and extensional belief, and refer to (Claßen and Delgrande 2021) for details. Nondeterminism is handled by distinguishing two types of action parameters: the ones under the control of the agent, and those under "nature's" control (we henceforth omit the scare quotes). For example, $flip(x, y)$ might denote the action of flipping a coin x, for which nature then determines outcome y (e.g., heads, tails, or some form of failure).

We use three special predicates: $Ieq(a, a')$ expresses that actions a and a' are *intensionally equivalent*, meaning they are identical except maybe for arguments set by nature. $Apl(a, p, s)$ says that action a has *argument plausibility* p in situation s, where $p = 0$ corresponds to the most plausible outcome(s), and higher values to less plausible ones. The *sensing fluent* $SF(a, s)$ holds iff action a reports sensing result "true" in situation s. Fallible sensing is again handled by a special argument set by nature, e.g. indicating that a light sensor works correctly ($sL(ok)$), or has a fault where it transiently reports the light being on ($sL(on)$) or off ($sL(off)$).

For beliefs, we allow for multiple initial situations s, denoted by $Init(s)$, in addition to S_0. We use an accessibility predicate $B(s', p, s)$ to express that in situation s, the agent considers situation s' possible with plausibility p. A formula ϕ is believed in situation s, written $Bel(\phi, s)$, iff it is true in all most plausible (MP) situations s' accessible from s:

$$MP(s', s) \doteq \exists p. B(s', p, s) \land \forall s'', p'(B(s'', p', s) \supset p \leq p')$$
$$Bel(\phi, s) \doteq \forall s'. MP(s', s) \supset \phi[s'].$$

(As convention, free variables are assumed to be universally quantified from the outside.) A successor state axiom for B then determines what is believed after executing an action a:

$$\begin{aligned} B(s', n, do(a, s)) \;\equiv\; &\exists p', s^*, n^*, a^*, p^*. B(s^*, n^*, s) \land \\ &Apl(a, p', s) \land Ieq(a, a^*) \land Apl(a^*, p^*, s^*) \land s' = do(a^*, s^*) \\ &\land (SF(a^*, s^*) \equiv SF(a, s)) \land n = n^* + p' + p^* \end{aligned} \quad (1)$$

We call formulas that don't mention SF, Apl, Ieq, or B *objective* formulas. A formula is *situation-independent* if it does not mention any fluents (including special ones such as B), but perhaps includes equalities over situation terms. Finally, we say that a (non-objective) formula is *about a ground situation term* $\sigma = do(\vec{\alpha}, S_i)$ if: (a) only σ occurs as right-hand side argument of B; (b) every other situation term in it is either a variable or a ground situation $do(\vec{\beta}, S_j)$ with $|\vec{\beta}| = |\vec{\alpha}|$; (c) if a fluent atom $F(\vec{t}, s)$ has a variable s as its situation argument, s is restricted to range only over situations such that $B(s, p, \sigma)$ for some p or $Init(s)$ holds. In particular any $Bel(\psi, \sigma)$ formula then is about σ.

2.3 Basic Action Theories

In order to describe a dynamic domain, we use action theories very similar to the ones proposed by Reiter (2001), but with additional axioms for the new special predicates:

Definition 1 (Basic Action Theory)**.** *A* basic action theory *(BAT) consists of the union of the following sets of formulas:*

- Σ_0 *is an initial theory that is given by a finite set of sentences about* S_0*, satisfying that initial situations are only* B*-related to other initial situations, and that plausibility values are a function of the pair of situations:*
$$Init(s) \land B(s', n, s) \supset Init(s')$$
$$Init(s) \land B(s', n_1, s) \land B(s', n_2, s) \supset n_1 = n_2$$
- Σ_{post} *is a finite set of successor state axioms, one for each fluent predicate* $F(\vec{x}, s)$*, of the form*
$$F(\vec{x}, do(a, s)) \;\equiv\; \gamma_F(\vec{x}, a, s),$$
where γ_F *is an objective formula uniform in* s *whose free variables are among* $\vec{x}$*,* a*, and* s*;*
- Σ_{sense} *is a singleton sentence of the form*
$$SF(a, s) \;\equiv\; \varphi_{SF}(a, s),$$
where φ_{SF} *is an objective formula uniform in* s *whose free variables are among* a *and* s*;*
- Σ_{Ieq} *is a singleton sentence of the form*
$$Ieq(a_1, a_2) \;\equiv\; \varphi_{Ieq}(a_1, a_2),$$
where $\varphi_{Ieq}(a_1, a_2)$ *is a situation-independent objective formula whose free variables are among* a_1 *and* a_2*;*
- Σ_{Apl} *is a singleton sentence of the form*
$$Apl(a, p, s) \;\equiv\; \varphi_{Apl}(a, p, s),$$
where φ_{Apl} *is an objective formula uniform in* s *whose free variables are among* a*,* p*, and* s*;*
- Σ_{una} *is a set of unique names axioms for actions;*
- Σ_{fnd} *is the set of foundational axioms that defines the set of situations to form a forest of isomorphic trees rooted at initial situations, where* $Init(s) \doteq \neg\exists a, s'. s = do(a, s')$*.*

We ignore preconditions here for brevity, but they would be straightforward to include. Compared to (Claßen and Delgrande 2021), we also make the simplifying assumption that axioms for Ieq, Apl, and SF are given as a single equivalence each, instead of multiple guarded axioms. It is conceivable, though significantly more involved, to extend our projection results to include guarded axioms along the lines of what is described in (De Giacomo and Levesque 1999).

Example 1. *We formalize the extended litmus scenario as follows. Fluents* $Ac(s)$*,* $Li(s)$*,* $Re(s)$*, and* $Bl(s)$ *denote that in situation* s*, the liquid is an acid, the paper is litmus paper, the paper is red, and the paper is blue, respectively. The action of dipping the paper in the liquid is denoted by* dip*, and the action of sensing whether the paper is white by* $sW(y)$*.*

- Σ_0*: Initially, the agent considers four situations possible: It does not know whether the liquid is acid, but believes (*$p = 0$*) that the paper is litmus, while allowing for the possibility (*$p = 1$*) that the paper is not litmus. It is certain though that the paper is white (neither red nor blue).*
$$\begin{aligned} &Init(s) \;\equiv\; s = S_0 \lor s = S_1 \lor s = S_2 \lor s = S_3, \\ &B(s, p, S_0) \;\equiv\; (s = S_1 \land p = 0) \lor (s = S_2 \land p = 0) \lor \\ &\quad (s = S_0 \land p = 1) \lor (s = S_3 \land p = 1), \\ &\neg Ac(S_0) \land \neg Li(S_0) \land \neg Re(S_0) \land \neg Bl(S_0), \\ &\quad Ac(S_1) \land \quad Li(S_1) \land \neg Re(S_1) \land \neg Bl(S_1), \\ &\neg Ac(S_2) \land \quad Li(S_2) \land \neg Re(S_2) \land \neg Bl(S_2), \\ &\quad Ac(S_3) \land \neg Li(S_3) \land \neg Re(S_3) \land \neg Bl(S_3). \end{aligned}$$
- Σ_{post}*: In this example,* $Ac(s)$ *and* $Li(s)$ *are rigid predicates, and don't change their value due to any action.* $Re(do(a, s))$ *will be true iff* a *is a* dip *action, the liquid is acid, and the paper is litmus. Similar for* Bl*.*
$$\begin{aligned} &Ac(do(a, s)) \;\equiv\; Ac(s), \quad Li(do(a, s)) \;\equiv\; Li(s), \\ &Re(do(a, s)) \;\equiv\; a = dip \land \quad Ac(s) \land Li(s) \;\lor\; Re(s), \\ &Bl(do(a, s)) \;\equiv\; a = dip \land \neg Ac(s) \land Li(s) \;\lor\; Bl(s). \end{aligned}$$

- Σ_{sense}: *When working correctly (outcome ok), sW senses whether the paper is neither red nor blue. In addition, we allow for transitory on and off failures. Thus:*

$$SF(a,s) \equiv (a = sW(ok) \wedge \neg(Re(s) \vee Bl(s))) \vee (a = sW(on))$$

- Σ_{Ieq}: *The dip action is deterministic, and so only intensionally equivalent to itself. The sensing action has a nature's argument to allow for transitory errors.*

$$Ieq(a_1,a_2) \equiv (a_1 = a_2 = dip) \vee \bigvee_{\substack{c_1,c_2 \in \\ \{ok,on,off\}}} (a_1 = sW(c_1) \wedge a_2 = sW(c_2))$$

- Σ_{Apl}: *Due to being deterministic, the dip action always has plausibility* 0*. The sensing action most plausibly returns the correct result, less plausibly has a transitory off failure, and even less plausibly a transitory on failure.*

$$\begin{aligned} &Apl(a,p,s) \equiv \\ &\quad (a = dip \wedge p = 0) \vee (a = sW(ok) \wedge p = 0) \vee \\ &\quad (a = sW(off) \wedge p = 1) \vee (a = sW(on) \wedge p = 2) \end{aligned}$$

3 Regression

The idea behind regression is to turn a query formula about a future situation into an equivalent one about the initial situation, and then test it against the initial theory of the BAT:

Definition 2 (Regression)**.** *Let* Σ *be a BAT,* $\vec{\alpha}$ *a sequence of ground action terms, and* ψ *a formula about* $do(\vec{\alpha}, S_0)$*. Then a formula* ϕ *is a* regression of ψ wrt Σ *if it is about* S_0 *and* $\Sigma \models \psi$ *iff* $\Sigma_0 \cup \Sigma_{una} \models \phi$*.*

Reiter (2001) presented a regression operator $\mathfrak{R}$ that achieves this for objective formulas by iteratively replacing subexpressions deemed equivalent by the BAT. It is straightforward to extend it to the new SF, Ieq, and Apl predicates:

- $\mathfrak{R}[F(\vec{t}, do(\alpha,\sigma))] = \mathfrak{R}[\gamma_F(\vec{t},\alpha,\sigma)]$, if F is a fluent;
- $\mathfrak{R}[P(\vec{t})] = \mathfrak{R}[\varphi_P(\vec{t})]$ for $P \in \{SF, Ieq, Apl\}$;
- $\mathfrak{R}[\phi] = \phi$ for every other atomic formula ϕ;
- $\mathfrak{R}$ distributes over $\neg, \wedge, \forall$, e.g. $\mathfrak{R}[\phi \wedge \psi] = \mathfrak{R}[\phi] \wedge \mathfrak{R}[\psi]$.

In principle, formulas involving B and Bel could be handled similarly, given that (1) is a successor state axiom for B. A technical problem is that the axiom as well as Bel formulas use quantification over situations, and due to nesting, we may end up with a formula such as $\exists s'.\, s' = do(a,s) \wedge P(s')$ that cannot immediately be regressed further, unless we rewrite it to $P(do(a,s))$. The solution below makes use of this idea, together with the fact that if situations σ and σ' are B-related, then they are the result of applying action sequences *of the same length* to B-related initial situations.

We here restrict ourselves to regressing belief atoms of the form $Bel(\phi, do(\vec{\alpha}, S_0))$, where $\vec{\alpha} = \langle \alpha_0, \ldots, \alpha_{k-1}\rangle$ is a sequence of ground actions of length k. Let $\vec{a}', \vec{a}''$ denote sequences of action variables of the same length, and $\vec{\alpha}_i$ the prefix $\langle \alpha_0, \ldots, \alpha_i\rangle$, with $\vec{\alpha}_{-1} \doteq \langle\rangle$. We add the rules

- $\mathfrak{R}[Bel(\phi, do(\vec{\alpha}, S_0))] =$
  $\mathfrak{R}[\forall \vec{a}', s'_0.\, MP(do(\vec{a}', s'_0), do(\vec{\alpha}, S_0)) \supset \phi[do(\vec{a}', s'_0)]]$;
- $\mathfrak{R}[MP(do(\vec{a}', s'_0), do(\vec{\alpha}, S_0))] =$
  $\mathfrak{R}[\exists p'.\, B(do(\vec{a}', s'_0), p', do(\vec{\alpha}, S_0)) \wedge$
  $\forall \vec{a}'', s''_0, p''\, (B(do(\vec{a}'', s''_0), p'', do(\vec{\alpha}, S_0)) \supset p' \leq p'')]$;
- $\mathfrak{R}[B(do(\vec{a}', s'_0), n, do(\vec{\alpha}, S_0))] =$
  $\mathfrak{R}[\exists n_0, \ldots, n_k.\, B(s'_0, n_0, S_0) \wedge (n = n_k) \wedge \bigwedge_{i=0}^{k-1} \Psi_i]$,

where the Ψ_i formulas in the last line are given by

$$\begin{aligned} \Psi_i \doteq\ &\exists p_i, p'_i.\, n_{i+1} = n_i + p_i + p'_i \wedge Ieq(\alpha_i, a'_i) \wedge \\ &Apl(\alpha_i, p_i, do(\vec{\alpha}_{i-1}, S_0)) \wedge Apl(a'_i, p'_i, do(\vec{a}'_{i-1}, s'_0)) \wedge \\ &(SF(a'_i, do(\vec{a}'_{i-1}, s'_0)) \equiv SF(\alpha_i, do(\vec{\alpha}_{i-1}, S_0))). \end{aligned}$$

Theorem 1. *For any belief atom* $\psi = Bel(\phi, do(\vec{\alpha}, S_0))$*,* $\mathfrak{R}[\psi]$ *is a regression of* ψ *wrt* Σ*.*

Proof. (Sketch) $\mathfrak{R}[\psi]$ clearly has the right form, and hence only depends on $\Sigma_0 \cup \Sigma_{una}$. Correctness follows from the fact that subformulas are only substituted by equivalent ones, except for the $\forall$-quantified s' and s'' in Bel and MP being replaced by $do(\vec{a}', s'_0)$ and $do(\vec{a}'', s''_0)$, respectively. The latter is shown to be "safe" by an induction over $\vec{\alpha}$. ☐

Example 2. *To decide if* $Bel(\neg Li, do(\langle dip; sW(ok)\rangle, S_0))$ *is entailed by the BAT from Ex. 1, we need to determine* $\mathfrak{R}[B(do(\langle a'_0; a'_1\rangle, s'_0), n, do(\langle dip; sW(ok)\rangle, S_0))]$*, i.e.*

$$\begin{aligned} \mathfrak{R}[\ &\exists n_0, n_1, n_2.\, B(s'_0, n_0, S_0) \wedge n = n_2 \wedge \\ &\exists p_0, p'_0.\, n_1 = n_0 + p_0 + p'_0 \wedge Ieq(dip, a'_0) \wedge \\ &\quad Apl(dip, p_0, S_0) \wedge Apl(a'_0, p'_0, s'_0) \wedge \\ &\quad (SF(a'_0, s'_0) \equiv SF(dip, S_0)) \wedge \\ &\exists p_1, p'_1.\, n_2 = n_1 + p_1 + p'_1 \wedge Ieq(sW(ok), a'_1) \wedge \\ &\quad Apl(sW(ok), p_1, do(dip, S_0)) \wedge \\ &\quad Apl(a'_1, p'_1, do(a'_0, s'_0)) \wedge \\ &\quad (SF(a'_1, do(a'_0, s'_0)) \equiv SF(sW(ok), do(dip, S_0)))\ ]. \end{aligned}$$

During further evaluation, many subexpressions can be simplified. For example, $SF(sW(ok), do(dip, S_0))$ *reduces to* TRUE*, and* $Ieq(dip, a'_0)$ *to* $(a'_0 = dip)$*. We thus obtain*

$$\begin{aligned} &\exists n_0.\, B(s'_0, n_0, S_0) \wedge a'_0 = dip \wedge \\ &\quad (a'_1 = sW(ok) \wedge n = n_0 \wedge \neg Li(s'_0) \vee \\ &\quad\ a'_1 = sW(on) \wedge n = n_0 + 2). \end{aligned}$$

$MP(\ldots)$ *now describes those* $do(\langle a'_0; a'_1\rangle, s'_0)$ *situations with minimal* n*, and* $Bel(\ldots)$ *requires* Li *to be false there. Both* $do(\langle dip; sW(ok)\rangle, S_0)$ *and* $do(\langle dip; sW(ok)\rangle, S_3)$ *satisfy this with* $n = 1$*. The agent hence believes the paper is not litmus after dipping and sensing that it is white.*

4 Progression

We follow (Vassos and Levesque 2013) and define:

Definition 3 (Progression)**.** *Let* Σ *be a BAT,* α *a ground action term, and* Σ_α *a set of sentences about* $do(\alpha, S_0)$*. Then* Σ_α *is a* progression *of* Σ_0 *wrt* α*,* Σ *iff for every sentence* ϕ *about* $do(\alpha, S_0)$*,* $\Sigma \models \phi$ *iff* $(\Sigma - \Sigma_0) \cup \Sigma_\alpha \models \phi$*.*

Lin and Reiter (1997) showed that even in the objective case, the progression of a theory may require second-order logic. However, they also presented special cases for which progression is first-order definable, one of them being the *relatively complete databases*. Here, we generalize this class to the case of extensional belief. Let us call a BAT relatively complete if Σ_0 comprises exactly the following formulas:

1. The formula
$$Init(s) \equiv s = S_0 \vee \cdots \vee s = S_k,$$
where $S_0, \ldots, S_k$ denote all and only initial situations;
2. one formula of the form
$$B(s, p, S_0) \equiv \Pi_B^{S_0}(s, p),$$
where $\Pi_B^{S_0}$ is a situation-independent objective formula whose free variables are among s, p;
3. for each fluent F and each S_i, one formula of the form
$$F(\vec{x}, S_i) \equiv \Pi_F^{S_i}(\vec{x}),$$
where $\Pi_F^{S_i}(\vec{x})$ is a situation-independent objective formula whose free variables are among $\vec{x}$.

Also assume that a ground action α has only finitely many β with $\Sigma_{una} \models Ieq(\alpha, \beta)$. Let $\mathfrak{P}$ now be an operator replacing B and F subformulas by corresponding Π expressions:

- $\mathfrak{P}[B(\sigma, t, S_0)] = \Pi_B^{S_0}(\sigma, t)$;
- $\mathfrak{P}[F(\vec{t}, \sigma)] = \bigvee_{0 \leq i \leq k} (\sigma = S_i) \wedge \Pi_F^{S_i}(\vec{t})$;
- $\mathfrak{P}[P(\vec{t})] = \mathfrak{P}[\varphi_P(\vec{t})]$ for $P \in \{SF, Ieq, Apl\}$;
- $\mathfrak{P}[\phi] = \phi$ for every other atomic formula ϕ;
- $\mathfrak{P}$ distributes over $\neg, \wedge, \forall$, e.g. $\mathfrak{P}[\phi \wedge \psi] = \mathfrak{P}[\phi] \wedge \mathfrak{P}[\psi]$.

Theorem 2. *Let Σ be a relatively complete BAT, α a ground action, and $\gamma_B(s', p, a, s)$ the right-hand side of axiom (1). We construct a set Σ_α consisting of the following formulas:*

1. The formula
$$B(s, p, do(\alpha, S_0)) \equiv \mathfrak{P}[\gamma_B(s, p, \alpha, S_0)].$$
2. For each fluent F, each S_i, and each ground action β such that $\Sigma_{una} \models Ieq(\alpha, \beta)$, the formula
$$F(\vec{x}, do(\beta, S_i)) \equiv \mathfrak{P}[\gamma_F(\vec{x}, \beta, S_i)].$$

Then Σ_α is a progression of Σ_0 through α wrt Σ.

Proof. (Sketch) Σ_α obviously has the right form. Soundness follows from the fact that we only replace equivalent subformulas, completeness from the fact that the set gives a complete description for all fluents and situations. □

Notice that Σ_α has the right form to apply progression repeatedly, if we treat the finitely many $do(\beta, S_i)$ as new initial situations (e.g. by replacing each by a new constant S_β^i).

Example 3. *Suppose we want to progress the BAT from Example 1 through $\langle dip; sW(ok)\rangle$, i.e. determine what the agent believes after dipping the paper and then performing a sense-white action. Notice that the BAT has the right form if we understand expressions like $\neg Ac(S_2)$ as shorthand for $Ac(S_2) \equiv$ FALSE. After the dip action, each initial situation is updated by the results of the dip (let $S_i' \doteq do(dip, S_i)$):*

$$\begin{aligned}
\neg Ac(S_0') \wedge \neg Li(S_0') \wedge \neg Re(S_0') \wedge \neg Bl(S_0')\\
Ac(S_1') \wedge Li(S_1') \wedge Re(S_1') \wedge \neg Bl(S_1')\\
\neg Ac(S_2') \wedge Li(S_2') \wedge \neg Re(S_2') \wedge Bl(S_2')\\
Ac(S_3') \wedge \neg Li(S_3') \wedge \neg Re(S_3') \wedge \neg Bl(S_3')
\end{aligned}$$

The situations' plausibilities remain unchanged, but for those situations where Li is true, the agent will believe the paper is red if the liquid is acid, and blue if base. Consequently the agent believes the paper is either red or blue.

$$\begin{aligned}
B(s, p, S_0') \equiv\ & (s = S_1' \wedge p = 0) \vee (s = S_2' \wedge p = 0) \vee\\
& (s = S_0' \wedge p = 1) \vee (s = S_3' \wedge p = 1)
\end{aligned}$$

After the sense-white action, the physical fluents retain their previous value, but the belief state changes as follows:

$$\begin{aligned}
B(s, p, do(\langle dip; sW(ok)\rangle, S_0)) \equiv\ &\\
& (s = do(\langle dip; sW(ok)\rangle, S_0) \wedge p = 1) \vee\\
& (s = do(\langle dip; sW(ok)\rangle, S_3) \wedge p = 1) \vee\\
& (s = do(\langle dip; sW(on)\rangle, S_1) \wedge p = 2) \vee\\
& (s = do(\langle dip; sW(on)\rangle, S_2) \wedge p = 2) \vee\\
& (s = do(\langle dip; sW(on)\rangle, S_0) \wedge p = 3) \vee\\
& (s = do(\langle dip; sW(on)\rangle, S_3) \wedge p = 3)
\end{aligned}$$

The two situations with minimum ($p = 1$) implausibility are those where sensing worked correctly. In both, the paper is not litmus, but Ac only holds in one. Other situations ($p > 1$) record less plausible circumstances where the sensing action failed. Hence the agent believes the paper is not litmus and does not know whether the liquid is acid or base.

5 Conclusion

We presented results on both regression and progression in a Situation Calculus-based framework for reasoning about the beliefs of an agent who is dealing with conflicting information, nondeterministic actions, and fallible sensing. As the two main solutions to the projection problem, they are paramount ingredients for any practical application of the formalism. A prototype implementation of our methods in SWI-Prolog with the examples from this paper is available on GitHub (Claßen 2022). For future work, we want to investigate their use as basis for an epistemic planner, possibly in combination with first-order knowledge compilation techniques (Claßen 2018). Without going into detail, we believe that our approach lends itself very well for this purpose when compared to similar work (Fang, Liu, and Wen 2015; Schwering and Lakemeyer 2015; Arenas et al. 2016). While (Shapiro et al. 2011) do not even consider projection, (Hunter and Delgrande 2011) only keep track of the most plausible set of situations, albeit in an action language (Gelfond and Lifschitz 1998) framework. In case of a contradiction, they regress to the initial situation, carry out belief revision, and progress through the action sequence again. We avoid this by keeping track of all possible outcomes, and in case of a sensing conflict, only need to look at the most plausible situations compatible with the sensing outcome.

Acknowledgements

We thank the reviewers for their helpful and pertinent comments. Financial support was gratefully received from the Natural Sciences and Engineering Research Council of Canada.

References

Arenas, M.; Baier, J. A.; Navarro, J. S.; and Sardiña, S. 2016. Incomplete causal laws in the situation calculus using free fluents. In Kambhampati, S., ed., *Proceedings of the Twenty-Fifht International Joint Conference on Artificial Intelligence (IJCAI 2016)*, 907–914. AAAI Press.

Claßen, J., and Delgrande, J. 2021. An account of intensional and extensional actions, and its application to belief, nondeterministic actions, and fallible sensors. In Bienvenue, M.; Erdem, E.; and Lakemeyer, G., eds., *Proceedings of the Eighteenth International Conference on the Principles of Knowledge Representation and Reasoning (KR 2021)*, 194–204. ijcai.org.

Claßen, J. 2018. Symbolic verification of Golog programs with first-order BDDs. In Thielscher, M.; Toni, F.; and Wolter, F., eds., *Proceedings of the Sixteenth International Conference on the Principles of Knowledge Representation and Reasoning (KR 2018)*, 524–529. AAAI Press.

Claßen, J. 2022. Belief Projection Prototype in SWI-Prolog. https://github.com/jens-classen/belief_projection.

De Giacomo, G., and Levesque, H. J. 1999. Projection using regression and sensors. In Dean, T., ed., *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI 1999)*, 160–165. Morgan Kaufmann Publishers Inc.

Delgrande, J. P., and Levesque, H. J. 2012. Belief revision with sensing and fallible actions. In Brewka, G.; Eiter, T.; and McIlraith, S. A., eds., *Proceedings of the Thirteenth International Conference on the Principles of Knowledge Representation and Reasoning (KR 2012)*. AAAI Press.

Delgrande, J. P., and Levesque, H. J. 2019. An epistemic approach to nondeterminism: Believing in the simplest course of events. *Studia Logica* 107(5):859–886.

Fang, L.; Liu, Y.; and Wen, X. 2015. On the progression of knowledge and belief for nondeterministic actions in the situation calculus. In Yang, Q., and Wooldridge, M., eds., *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015)*, 2955–2963. AAAI Press.

Gelfond, M., and Lifschitz, V. 1998. Action languages. *Electronic Transactions on AI* 3.

Hunter, A., and Delgrande, J. P. 2011. Iterated belief change due to actions and observations. *Journal of Artificial Intelligence Research* 40:269–304.

Lin, F., and Reiter, R. 1997. How to progress a database. *Artificial Intelligence* 92(1–2):131–167.

McCarthy, J., and Hayes, P. 1969. Some philosophical problems from the standpoint of artificial intelligence. In *Machine Intelligence 4*. New York: American Elsevier. 463–502.

Moore, R. C. 1985. A formal theory of knowledge and action. In Hobbs, J. R., and Moore, R. C., eds., *Formal Theories of the Commonsense World*. Norwoord, NJ, USA: Ablex Publishing. 319–358.

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

Schwering, C., and Lakemeyer, G. 2015. Projection in the epistemic situation calculus with belief conditionals. In Bonet, B., and Koenig, S., eds., *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI 2015)*, 1583–1589. AAAI Press.

Shapiro, S.; Pagnucco, M.; Lespérance, Y.; and Levesque, H. J. 2011. Iterated belief change in the situation calculus. *Artificial Intelligence* 175(1):165–192.

Vassos, S., and Levesque, H. J. 2013. How to progress a database III. *Artificial Intelligence* 195:203–221.