Defaults in Action: Non-monotonic Reasoning About States in Action Calculi

Hannes Strass and **Michael Thielscher**
Department of Computer Science
Dresden University of Technology
{hannes.strass, mit}@inf.tu-dresden.de

Abstract

We propose a mechanism for default reasoning in action formalisms that allows to make useful assumptions unless information to the contrary. The mechanism is shown to behave properly when actions are performed, in particular we show that it suffices to apply defaults to the initial state. This allows for very simple reasoning, since the defaults need only be applied once and monotonic entailment can thence be used to solve projection problems. We finally consider two simple, natural generalizations of the approach and show that they admit unintuitive conclusions.

Introduction

This paper is concerned with the combination of two successful approaches to the logical formalization of commonsense reasoning: logics for actions and non-monotonic logics. The present work is by no means the first to join the two; non-monotonic logics have already been used by the reasoning about actions community in the past. After (McCarthy and Hayes 1969) discovered the fundamental problem of determining the non-effects of actions, the frame problem, it was widely believed that non-monotonic reasoning were necessary to solve it. Then (Hanks and McDermott 1987) gave a (by now famous) example of how straightforward use of non-monotonic logics in reasoning about actions and change can lead to counter-intuitive results. When monotonic solutions to the frame problem were found (Reiter 1991; Thielscher 1999), non-monotonic reasoning again seemed to be obsolete.

In this paper, we argue that utilizing default logic still is of use when reasoning about actions. We will not use it to solve the frame problem, however, the solution to the frame problem we use here is monotonic and similar to the one of (Thielscher 1999), but to make useful default assumptions about states.

The approach we propose uses deterministic actions without conditional effects and a restricted form of default assumptions. The main reasoning task we are interested in is the projection problem, that is, given an initial situation and a sequence of actions, the question whether a certain condition holds in the resulting state. The approach can be used to draw intuitive conclusions that are not possible to draw in a monotonic way. As the main result of this paper, we show that default applications can be restricted to the initial state without losing any inferences, thus giving way to a simple reasoning mechanism.

In the second half of the paper, we consider two generalizations of our approach and show how they permit counterintuitive conclusions, which justifies the restrictions made earlier. The first generalization allows for more general defaults: they are still supernormal, that is, prerequisite-free and normal, but enable default conclusions to be "carried back in time." This clearly disqualifies for solving projection problems, since we would have to take an infinite number of future time points into account. The second generalization allows for more general effect axioms: they are still deterministic but permit to express conditional effects. This again causes conclusions that rely on time points that are intrinsically irrelevant for the question to be answered.

Background

This section introduces the foundations upon which our work rests. Firstly, a unifying action calculus that we will use to axiomatize action domains. Secondly, a restricted version of one of the most prominent non-monotonic logics, Raymond Reiter's Default Logic (Reiter 1980).

The Unifying Action Calculus

Recently, (Thielscher 2009) proposed a unifying action calculus (UAC) with the objective of bundling research efforts in action formalisms. It does not confine to a particular time structure and can thus be instantiated with situation-based action calculi, like the Situation Calculus (McCarthy 1963) or the Fluent Calculus (Thielscher 1999), as well as with formalisms using a linear time structure, like the Event Calculus (Kowalski and Sergot 1986).

The UAC uses only the sorts FLUENT, ACTION, and TIME along with the predicates $< : \text{TIME} \times \text{TIME}$ (denoting an ordering of time points), $Holds : \text{FLUENT} \times \text{TIME}$ (stating whether a fluent evaluates to true at a given time point), and $Poss : \text{ACTION} \times \text{TIME} \times \text{TIME}$ (indicating whether an action is applicable for particular starting and ending time points). Uniqueness-of-names is assumed for all (finitely many) functions into sorts FLUENT and ACTION.

The following definition introduces the most important types of formulas of the unifying action calculus: they allow to express properties of states and applicability conditions and effects of actions.

Definition 1. Let $\vec{s}$ be a sequence of variables of sort TIME.

- A *state formula* $\Phi[\vec{s}]$ *in* $\vec{s}$ is a first-order formula with free variables $\vec{s}$ where
  - for each occurrence of $Holds(\varphi, s)$ in $\Phi[\vec{s}]$ we have $s \in \vec{s}$ and
  - predicate $Poss$ does not occur.

Let s, t be variables of sort TIME and A be a function into sort ACTION.

- A *precondition axiom* is of the form

$$Poss(A(\vec{x}), s, t) \equiv \pi_A[s] \quad (1)$$

  where $\pi_A[s]$ is a state formula in s with free variables among $s, t, \vec{x}$.

- An *effect axiom* is of the form

$$Poss(A(\vec{x}), s, t) \supset$$
$$(\forall f)((\gamma_A^+ \vee (Holds(f, s) \wedge \neg\gamma_A^-)) \equiv Holds(f, t)) \quad (2)$$

  where

$$\gamma_A^+ = \bigvee_{0 \le i \le n_A^+} f = \varphi_i \text{ and } \gamma_A^- = \bigvee_{0 \le i \le n_A^-} f = \psi_i$$

  and the φ_i and ψ_i are terms of sort FLUENT with free variables among $\vec{x}$.

Readers may be curious as to why the predicate $Poss$ carries two time arguments instead of just one: $Poss(a, s, t)$ is to be read as "action a is possible starting at time s and ending at time t." The formulas γ_A^+ and γ_A^- enumerate the positive and negative effects of the action, respectively. This definition of effect axioms is a restricted version of the original definition of (Thielscher 2009)—it only allows for deterministic actions with unconditional effects.

A few words on notation and naming conventions: lower-case letters will denote object-level variables, we usually use f for sort FLUENT, a for sort ACTION, and s, t for sort TIME. Capital letters and words will denote object level functions of all sorts. Lower-case Greek letters will serve as meta-level variables for fluent and action terms. Capital Greek letters denote formulas or sets of formulas. As usual, $s \le t$ abbreviates $s < t \vee s = t$. Formulas with occurrences of free variables are assumed universally prenex-quantified.

Next, we formalize the concept of an (action) domain axiomatization with its notion of time and action laws.

Definition 2. A *(UAC) domain axiomatization* consists of a finite set of foundational axioms Ω (that define the underlying time structure), a set Π of precondition axioms (1), and a set Υ of effect axioms (2); the latter two for all functions into sort ACTION.

A domain axiomatization is *progressing*, if

- $\Omega \models (\exists s : \text{TIME})(\forall t : \text{TIME}) s \le t$ and
- $\Omega \cup \Pi \models Poss(a, s, t) \supset s < t$.

A domain axiomatization is *sequential*, if it is progressing and

$$\Omega \cup \Pi \models Poss(a, s, t) \wedge Poss(a', s', t') \supset$$
$$(t < t' \supset t \le s') \wedge (t = t' \supset (a = a' \wedge s = s'))$$

That is, a domain axiomatization is progressing if there exists a least time point and time always increases when applying an action. A sequential domain axiomatization furthermore requires that no two actions overlap.

Lastly, we formalize the intuition of a time point that is reachable via a finite sequence of actions.

Definition 3. Let Σ be a domain axiomatization. A time point τ is *finitely reachable in* Σ iff $\Sigma \models Reach(\tau)$, where the predicate $Reach : \text{TIME}$ is macro-defined by

$$Reach(r) \stackrel{\text{def}}{=} (\forall R)((\forall s)(Init(s) \supset R(s))$$
$$\wedge (\forall a, s, t)(R(s) \wedge Poss(a, s, t) \supset R(t)) \supset R(r))$$
$$Init(t) \stackrel{\text{def}}{=} \neg(\exists s) s < t$$

Note that these macros allow us to perform induction on reachable time points as follows: to show that a certain property $\Psi[s]$ holds for all reachable time points, we show that all minimal time points satisfy the property and that it is preserved by action application to reachable time points.

The examples of this paper will employ situations as their underlying time structure, so we briefly recall the corresponding foundational axioms from (Pirri and Reiter 1999):

$$\neg(s < S_0) \quad (3)$$
$$s < Do(a, s') \equiv s \le s' \quad (4)$$
$$Do(a, s) = Do(a', s') \equiv (a = a' \wedge s = s') \quad (5)$$
$$(\forall P)((P(S_0) \wedge (P(s) \supset P(Do(a, s)))) \supset P(s')) \quad (6)$$

The above axioms shall henceforth be referred to as Ω_{sit}. Whenever we use them as underlying time structure, we stipulate that for each action function A with right hand side $\pi_A[s]$ of precondition axiom (1), we have $\pi_A[s] \equiv \pi'_A[s] \wedge t = Do(A(\vec{x}), s)$ for some π'_A.

Since we are mainly interested in the projection problem, our domain axiomatizations will usually include a set Σ_0 of state formulas in S_0 that characterize the initial situation.

To illustrate the intended usage of the introduced notions, we make use of a variant of a well-known example already mentioned earlier: the Yale Shooting scenario (Hanks and McDermott 1987).

Example 1. Consider the domain axiomatization $\Sigma = \Omega_{sit} \cup \Pi \cup \Upsilon \cup \Sigma_0$. The precondition axioms say that the action Shoot is possible if the gun is loaded and the actions Load and Wait are always possible.

$$\Pi = \{Poss(\mathsf{Shoot}, s, t) \equiv$$
$$(Holds(\mathsf{Loaded}, s) \wedge t = Do(\mathsf{Shoot}, s)),$$
$$Poss(\mathsf{Load}, s, t) \equiv t = Do(\mathsf{Load}, s),$$
$$Poss(\mathsf{Wait}, s, t) \equiv t = Do(\mathsf{Wait}, s)\}$$

With these preconditions and foundational axioms (3)–(6), the domain axiomatization is sequential. The effect of shooting is that the turkey ceases to be alive, loading the gun causes it to be loaded, and waiting does not have any effect. All effect axioms in Υ are of the form (2), we state only the $\gamma^\pm$ different from the empty disjunction.

$$\gamma^-_{\mathsf{Shoot}} = (f = \mathsf{Alive})$$
$$\gamma^+_{\mathsf{Load}} = (f = \mathsf{Loaded})$$

Finally, we state that the turkey is alive in the initial situation S_0.

$$\Sigma_0 = \{Holds(\mathsf{Alive}, S_0)\}$$

We can now employ logical entailment to answer the question whether the turkey is still alive after applying the actions Load, Wait, and Shoot, respectively. With the notation $Do([a_1, \ldots, a_n], s)$ as abbreviation for $Do(a_n, Do(\ldots, Do(a_1, s)\ldots))$, it is easy to see that

$$\Sigma \models \neg Holds(\mathsf{Alive}, Do([\mathsf{Load}, \mathsf{Wait}, \mathsf{Shoot}], S_0).$$

Default Logic

Introduced in the seminal paper (Reiter 1980), Default Logic has become one of the most important formalisms for non-monotonic reasoning. The semantics for supernormal defaults used here is taken from (Brewka and Eiter 1999), which is itself an enhancement of a notion developed in (Brewka 1989)[1]. Here, a default rule always comes without a prerequisite, and justification and consequence always coincide. A default rule can thus also be seen as a hypothesis that we are willing to assume, but prepared to give up in case of contradiction. A default theory then adds a set of formulas, the *indefeasible knowledge*, that we are not willing to give up for any reason.

Definition 4. A *supernormal default rule*, or, for short, *default*, is a closed first-order formula. Any formulas with occurrences of free variables are taken as representatives of their ground instances.

For a set of closed formulas S, we say the default δ is *active in* S if both $\delta \notin S$ and $\neg\delta \notin S$.

A *(supernormal) default theory* is a pair $(W, \mathcal{D})$, where W is a set of sentences and $\mathcal{D}$ a set of default rules.

An extension for a default theory can be seen as a way of assuming as many defaults as possible without creating inconsistencies. It should be noted that, although the definition differs, our extensions are extensions in Reiter's (1980) sense.

Definition 5. Let $(W, \mathcal{D})$ be a default theory where all default rules are supernormal and $\prec\!\!\!\prec$ be a total order on $\mathcal{D}$. Define $E_0 := Th(W)$ and for all $i > 0$,

$$E_{i+1} = \begin{cases} E_i & \text{if no default is active in } E_i \\ Th(E_i \cup \{\delta\}) & \text{otherwise, where } \delta \text{ is the } \prec\!\!\!\prec\text{-} \\ & \text{minimal default active in } E_i. \end{cases}$$

Then the set $E := \bigcup_{i>0} E_i$ is called the *extension generated by* $\prec\!\!\!\prec$. A set of formulas E is a *preferred extension* for $(W, \mathcal{D})$ if there exists a total order $\prec\!\!\!\prec$ that generates E. The set of all preferred extensions for a default theory $(W, \mathcal{D})$ is denoted by $Ex(W, \mathcal{D})$.

[1]Readers familiar with these works will note that they are concerned with prioritized default logics while we do not use priorities at all. We however use the more general definition because we intend to incorporate prioritized defaults into our framework later on.

Extensions need not be unique: if there are two contradicting defaults δ and $\neg\delta$, either both or none of them are active in $Th(W)$. Applying one of them makes the other inactive, thus they give rise to two different extensions.

Based on extensions, one can define skeptical and credulous conclusions for default theories: skeptical conclusions are formulas that are contained in every extension, credulous conclusions are those that are contained in at least one extension.

Definition 6. Let $(W, \mathcal{D})$ be a supernormal default theory and Ψ be a first-order formula.

$$W \approx_{\mathcal{D}}^{skept} \Psi \stackrel{\text{def}}{\equiv} \Psi \in \bigcap_{E \in Ex(W,\mathcal{D})} E$$

$$W \approx_{\mathcal{D}}^{cred} \Psi \stackrel{\text{def}}{\equiv} \Psi \in \bigcup_{E \in Ex(W,\mathcal{D})} E$$

In the present work, we will primarily be concerned with skeptical reasoning.

Action Domains with Static Defaults

We now combine the hitherto introduced concepts into the notion of a domain axiomatization with defaults. It is essentially a default theory where the set containing the indefeasible knowledge is a domain axiomatization. The defaults are of a restricted form since we allow only static defaults about states.

Definition 7. A *domain axiomatization with defaults* is a pair $(\Sigma, \mathcal{D}[s])$, where Σ is a UAC domain axiomatization and $\mathcal{D}[s]$ is a set of supernormal defaults of the form $Holds(\varphi, s)$ or $\neg Holds(\varphi, s)$ for a fluent φ.

By $\mathcal{D}[\sigma]$ we denote the set of defaults in $\mathcal{D}[s]$ where s has been instantiated by the term σ.

Example 1 (continued)**.** We add a fluent Broken that indicates if the gun does not function properly. Shooting is now only possible if the gun is loaded *and* not broken:

$$\begin{aligned} & Poss(\mathsf{Shoot}, s, t) \equiv \\ & \quad (Holds(\mathsf{Loaded}, s) \land \neg Holds(\mathsf{Broken}, s) \\ & \quad \land t = Do(\mathsf{Shoot}, s)) \end{aligned}$$

Unless there is information to the contrary, it should be assumed that the gun has no defects. This is expressed by the following default rule:

$$\mathcal{D}[s] = \{\neg Holds(\mathsf{Broken}, s)\}$$

Without the default assumption, it cannot be concluded that the action Shoot is possible after performing Load and Wait since it cannot be inferred that the gun is not broken. Using the abbreviations $S_1 = Do(\mathsf{Load}, S_0)$, $S_2 = Do(\mathsf{Wait}, S_1)$, and $S_3 = Do(\mathsf{Shoot}, S_2)$, we illustrate how the non-monotonic entailment relation defined earlier enables us to use the default rule to draw the desired conclusion:

$$\Sigma \approx_{\mathcal{D}[S_0]}^{skept} \neg Holds(\mathsf{Broken}, S_2),$$

$$\Sigma \approx_{\mathcal{D}[S_0]}^{skept} Poss(\mathsf{Shoot}, S_2, S_3), \text{ and}$$

$$\Sigma \approx_{\mathcal{D}[S_0]}^{skept} \neg Holds(\mathsf{Alive}, S_3).$$

The default conclusion that the gun works correctly, drawn in S_0, carries over to S_2 and allows to conclude applicability of Shoot in S_2 and its effects on S_3.

In the example just seen, default reasoning could be restricted to the initial situation. As it turns out, this is sufficient for the type of action domain considered here: effect axiom (2) never "removes" information about fluents and thus never makes more defaults active after executing an action. This observation is formalized by the following lemma. It essentially says that to reason about a time point in which an action ends, it makes no difference whether we apply the defaults to the resulting time point or to the time point when the action starts. This holds of course only due to the restricted nature of effect axiom (2).

Lemma 1. *Let $(\Sigma, \mathcal{D}[s])$ be a domain axiomatization with defaults, α be a ground action such that $\Sigma \models Poss(\alpha, \sigma, \tau)$ for some σ, τ : TIME, and let $\Psi[\tau]$ be a state formula in τ. Then*

$$\Sigma \approx^{skept}_{\mathcal{D}[\sigma]} \Psi[\tau] \text{ iff } \Sigma \approx^{skept}_{\mathcal{D}[\tau]} \Psi[\tau]$$

Proof. (Sketch.) The proof uses structural induction on $\Psi[\tau]$ with $\Psi[\tau] = Holds(\varphi, \tau)$ being the only interesting case. The result is immediate if Σ is inconsistent, so for the following assume that Σ is consistent. If φ is amongst the positive effects of α, then $\Sigma \models Holds(\varphi, \tau)$ and we are done. If φ is no positive effect of α, the conclusion $Holds(\varphi, \tau)$ relies on a default $Holds(\varphi, s) \in \mathcal{D}[s]$ and φ cannot be a negative effect of α (since the conclusion would be impossible otherwise). Since φ is not changed by α, we have that $Holds(\varphi, \sigma) \in E$ iff $Holds(\varphi, \tau) \in E$ for any extension E for $(\Sigma, \mathcal{D}[\sigma])$ or $(\Sigma, \mathcal{D}[\tau])$. □

We next introduce a helpful regression operator which is inspired by the one from (Reiter 1991). It uses the structure of the effect axioms to reduce reasoning about a time point that is the result of applying an action to reasoning about the time point in which the action started.

Definition 8. The operator $\mathcal{R}_\alpha$ maps, for a given action α, a state formula in τ into a state formula in σ as follows.

$$\mathcal{R}_\alpha(Holds(\varphi, \tau)) \stackrel{\text{def}}{=} (\gamma^+_\alpha\{f \mapsto \varphi\} \vee (Holds(\varphi, \sigma) \wedge \neg\gamma^-_\alpha\{f \mapsto \varphi\}))$$

The operator does not change atomic formulas other than $Holds$ statements, and distributes over the first order connectives in the obvious way.

Now whenever an action α is possible and its effect axiom is available, a state formula in the resulting time point and its regression are indeed equivalent.

Proposition 2. *Let α be a ground term of sort* ACTION *and S be a consistent set of closed formulas that contains an effect axiom (2) for action α and where $S \models Poss(\alpha, \sigma, \tau)$ for some σ, τ :* TIME *and let $\Psi[s]$ be a state formula. Then*

$$S \models \Psi[\tau] \equiv \mathcal{R}_\alpha(\Psi)[\sigma]$$

Proof. By structural induction on Ψ. The only interesting case is $\Psi = Holds(\varphi, \tau)$ for some fluent φ. Let I be a model for S.

$I \models Holds(\varphi, \tau)$
iff $I \models (\gamma^+_\alpha\{f \mapsto \varphi\} \vee (Holds(\varphi, \sigma) \wedge \neg\gamma^-_\alpha\{f \mapsto \varphi\}))$
(since $I \models Poss(\alpha, \sigma, \tau)$ and I is a model for α's effect axiom)
iff $I \models \mathcal{R}_\alpha(Holds(\varphi, \tau))$ (by definition) □

The next theorem says that all *local* conclusions about a finitely reachable time point σ (that is, all conclusions about σ using defaults from $\mathcal{D}[\sigma]$) are exactly the conclusions about σ that we can draw by instantiating the defaults only with the least time point.

Theorem 3. *Let $(\Sigma, \mathcal{D}[s])$ be a progressing domain axiomatization with defaults, λ its least time point, σ :* TIME *be finitely reachable, and $\Psi[\sigma]$ be a state formula. Then*

$$\Sigma \approx^{skept}_{\mathcal{D}[\sigma]} \Psi[\sigma] \text{ iff } \Sigma \approx^{skept}_{\mathcal{D}[\lambda]} \Psi[\sigma]$$

Proof. By induction on σ. The base case is trivial. For the induction step, assume that $\Sigma \models Poss(\alpha, \sigma, \tau)$.

$\Sigma \approx^{skept}_{\mathcal{D}[\tau]} \Psi[\tau]$
iff $\Sigma \approx^{skept}_{\mathcal{D}[\sigma]} \Psi[\tau]$ (Lemma 1)
iff $\Sigma \approx^{skept}_{\mathcal{D}[\sigma]} \mathcal{R}_\alpha(\Psi)[\sigma]$ (Proposition 2)
iff $\Sigma \approx^{skept}_{\mathcal{D}[\lambda]} \mathcal{R}_\alpha(\Psi)[\sigma]$ (induction hypothesis)
iff $\Sigma \approx^{skept}_{\mathcal{D}[\lambda]} \Psi[\tau]$ (Proposition 2) □

It thus remains to show that local defaults are indeed exhaustive with respect to local conclusions. The next lemma takes a step into this direction: it states that action application does not increase default knowledge about past time points.

Lemma 4. *Let $(\Sigma, \mathcal{D}[s])$ be a domain axiomatization with defaults, α be a ground action such that $\Sigma \models Poss(\alpha, \sigma, \tau)$ for some σ, τ :* TIME*, and let $\Psi[\rho]$ be a state formula in ρ :* TIME *where $\rho \leq \sigma$. Then*

$$\Sigma \approx^{skept}_{\mathcal{D}[\tau]} \Psi[\rho] \text{ implies } \Sigma \approx^{skept}_{\mathcal{D}[\sigma]} \Psi[\rho]$$

Proof. (Sketch.) We prove the contrapositive. Let $\Sigma \not\approx^{skept}_{\mathcal{D}[\sigma]} \Psi[\rho]$. Then there is an extension E for $(\Sigma, \mathcal{D}[\sigma])$ where $\Psi[\rho] \notin E$. We generate an extension F for $(\Sigma, \mathcal{D}[\tau])$ as follows. Set the ordering $\prec\!\!\prec$ on $\mathcal{D}[\tau]$such that defaults from $\mathcal{D}[\tau] \cap E$ get higher priority than the ones from $\mathcal{D}[\tau] \setminus E$. None of the latter gets applied during generation of F: roughly, if $\delta[\tau] \notin E$ although there is a default $\delta[s] \in \mathcal{D}[s]$, then $\neg\delta[\tau] \in E$. This can be due to either (1) a contradicting action effect or (2) a contradicting default $\neg\delta[s] \in \mathcal{D}[s]$. In case (1), $\neg\delta[\tau] \in Th(\Sigma)$ and $\delta[\tau]$ is inapplicable. For (2), α does not affect $\neg\delta[\sigma]$, thus $\neg\delta[\tau]$ is applicable in $Th(\Sigma)$ and by construction applied in F, which makes $\delta[\tau]$ inapplicable. Now there exists an $E' \subseteq \mathcal{D}[\tau] \cap E$ such that $F = Th(\Sigma \cup E')$, thus any model for E is a model for F. Hence, $\Psi[\rho] \notin F$ and $\Sigma \not\approx^{skept}_{\mathcal{D}[\tau]} \Psi[\rho]$. □

The converse of the lemma does not hold, since an action effect might preclude a default conclusion about the past. The following theorem now says that no sequence of future actions whatsoever can have an impact on conclusions about the present.

Theorem 5. *Let* $(\Sigma, \mathcal{D}[s])$ *be a progressing domain axiomatization with defaults, let* $\Psi[s]$ *be a state formula,* $\sigma \leq \tau$ *be time points, and* σ *be finitely reachable. Then*

$$\Sigma \mathrel{|\!\!\approx}^{skept}_{\mathcal{D}[\tau]} \Psi[\sigma] \text{ implies } \Sigma \mathrel{|\!\!\approx}^{skept}_{\mathcal{D}[\sigma]} \Psi[\sigma]$$

Proof. If τ is not finitely reachable, we have $\Sigma \models \Psi[\sigma]$ and the claim is immediate, so let τ be finitely reachable. We use induction on τ. The base case, $\tau = \sigma$, is obvious. For the induction step, $\Sigma \models Poss(\alpha, \tau, \tau')$ and $\Sigma \mathrel{|\!\!\approx}^{skept}_{\mathcal{D}[\tau']} \Psi[\sigma]$ imply $\Sigma \mathrel{|\!\!\approx}^{skept}_{\mathcal{D}[\tau]} \Psi[\sigma]$ by Lemma 4. The induction hypothesis then yields $\Sigma \mathrel{|\!\!\approx}^{skept}_{\mathcal{D}[\sigma]} \Psi[\sigma]$. □

The final theorem, our main result, now combines Theorems 3 and 5.

Theorem 6. *Let* $(\Sigma, \mathcal{D}[s])$ *be a progressing domain axiomatization with defaults,* λ *be its least time point,* $\Psi[s]$ *be a state formula, and* $\sigma \leq \tau$ *be terms of sort* TIME *where* σ *is finitely reachable. Then*

$$\Sigma \mathrel{|\!\!\approx}^{skept}_{\mathcal{D}[\tau]} \Psi[\sigma] \text{ implies } \Sigma \mathrel{|\!\!\approx}^{skept}_{\mathcal{D}[\lambda]} \Psi[\sigma]$$

Proof. $\Sigma \mathrel{|\!\!\approx}^{skept}_{\mathcal{D}[\tau]} \Psi[\sigma]$ implies $\Sigma \mathrel{|\!\!\approx}^{skept}_{\mathcal{D}[\sigma]} \Psi[\sigma]$ by Theorem 5. By Theorem 3, this is the case iff $\Sigma \mathrel{|\!\!\approx}^{skept}_{\mathcal{D}[\lambda]} \Psi[\sigma]$. □

Generalizations with Undesired Side Effects

In this section, we show some generalizations of the thus far introduced notion of a domain axiomatization with defaults and show how these generalizations clash with our intuitive notion of relevance. The first subsection generalizes the default hypotheses used, and the second subsection generalizes the effect axioms.

Unrestricted Supernormal Defaults

Concluding atomic propositions about the world is not always enough. Sometimes we wish to express defaults of the form "in general, x are y", for example, "in general, paper airplanes fly[2]." Surely, we could instantiate a default $Holds(\mathsf{Flies}(x), s)$ by all objects x which are known to be paper airplanes. But this is by no means elaboration tolerant (McCarthy 1998) and furthermore does not account for previously unknown paper airplanes. We would much rather have a default rule

$$Holds(\mathsf{PaperAirplane}(x), s) \supset Holds(\mathsf{Flies}(x), s) \quad (7)$$

which is still supernormal and will let us draw the desired conclusion whenever there is no contradicting information. But, unfortunately, allowing disjunctive defaults has unintuitive side effects:

Example 2. Imagine an action $\mathsf{Fold}(x)$ that transforms a sheet of paper x into a paper airplane:

$$Poss(\mathsf{Fold}(x), s, t) \equiv Holds(\mathsf{SheetOfPaper}(x), s) \wedge t = Do(\mathsf{Fold}(x), s)$$

$$\gamma^+_{\mathsf{Fold}} = (f = \mathsf{PaperAirplane}(x))$$

$$\gamma^-_{\mathsf{Fold}} = (f = \mathsf{SheetOfPaper}(x))$$

[2] Yes, paper airplanes. Birds are not the only objects that should fly by default.

Let the domain axiomatization be $\Sigma = \Omega_{sit} \cup \Pi \cup \Upsilon \cup \Sigma_0$ where Π contains the precondition axiom above, Υ contains effect axiom (2) with γ^+_{Fold} and γ^-_{Fold} stated above, and the initial situation is characterized by $\Sigma_0 = \{Holds(\mathsf{SheetOfPaper}(\mathsf{T}), S_0)\}$. The set of defaults $\mathcal{D}[s]$ contains the single default rule (7). Now after folding T into a paper airplane (using the abbreviation $S_1 = Do(\mathsf{Fold}(\mathsf{T}), S_0)$), we can indeed make the desired conclusion that it flies:

$$\Sigma \mathrel{|\!\!\approx}^{skept}_{\mathcal{D}[S_1]} Holds(\mathsf{Flies}(\mathsf{T}), S_1)$$

So far, so good. But there is another conclusion that we can draw in S_1 and that refers to the past:

$$\Sigma \mathrel{|\!\!\approx}^{skept}_{\mathcal{D}[S_1]} Holds(\mathsf{Flies}(\mathsf{T}), S_0)$$

Spelled out, the sheet of paper *already flew before it was folded!* Moreover, this conclusion about the initial situation could not be drawn in the initial situation itself without utilizing a future situation:

$$\Sigma \not\mathrel{|\!\!\approx}^{skept}_{\mathcal{D}[S_0]} Holds(\mathsf{Flies}(\mathsf{T}), S_0)$$

This line of argument could be read as: "If I folded the sheet of paper into a paper airplane, it would fly. Therefore, it flies." This is counterfactual reasoning gone awry. So what happened?

The problem stems from effect axiom (2) and its incorporated solution to the frame problem: since $\mathsf{Flies}(\mathsf{T})$ holds after $\mathsf{Fold}(\mathsf{T})$ but was not a positive effect of the action, according to the effect axiom it must have held beforehand. This example shows that disjunctive defaults can have unintended effects in the presence of actions: they are, locally instantiated, not exhaustive with respect to local conclusions. A proposition similar to Theorem 5 can thus not be made when using default rules with disjunctions.

Conditional Effects

Let us get back to defaults that are $Holds$ statements or negations thereof, but instead increase the expressiveness of the action domain by allowing *conditional effects* (also called *alternative results* (Sandewall 1994)). They are modelled as a case distinction on the right hand side of the effect axiom. For each case, the actual formula expressing the effects is identical to (2).

Definition 9. An *effect axiom with conditional effects* is of the form

$$Poss(A(\vec{x}), s, t) \supset \bigvee_{1 \leq i \leq k} (\Phi_i[s] \wedge \Upsilon_i[s, t]) \quad (8)$$

where $k \geq 1$, and for each $1 \leq i \leq k$,

$$\Upsilon_i[s, t] = (\forall f)(Holds(f, t) \equiv (\gamma_i^+ \vee (Holds(f, s) \wedge \neg\gamma_i^-))) \quad (9)$$

$$\gamma^+ = \bigvee_{0 \leq j \leq n_i^+} f = \varphi_{ij} \text{ and } \gamma^- = \bigvee_{0 \leq j \leq n_i^-} f = \psi_{ij}$$

and the φ_{ij} and ψ_{ij} are terms of sort FLUENT with free variables among $\vec{x}$. The $\Phi_i[s]$ are state formulas in s that define the conditions for case i to apply. They are mutually exclusive and the disjunction of them is a tautology—the actions are thus still deterministic.

Conditional effects allow us to further "inspect" a state and base effects upon state properties. This was not possible with effect axiom (2) where all effects were unconditional and the only possibility to inspect the starting state of an action was by precondition axioms.

Example 3. We slightly modify Example 1: the action Shoot is now always possible but breaks an unloaded gun (that works as expected if loaded and not broken).

$$\begin{array}{l} Poss(\mathsf{Shoot}, s, t) \equiv t = Do(\mathsf{Shoot}, s) \\ Poss(\mathsf{Shoot}, s, t) \supset \\ \quad (\neg Holds(\mathsf{Broken}, s) \wedge Holds(\mathsf{Loaded}, s)) \wedge \\ \qquad (\forall f)((Holds(f, s) \wedge f \neq \mathsf{Alive}) \equiv Holds(f, t)) \\ \quad \vee \\ \quad (Holds(\mathsf{Broken}, s) \vee \neg Holds(\mathsf{Loaded}, s)) \wedge \\ \qquad (\forall f)(Holds(f, s) \vee f = \mathsf{Broken} \equiv Holds(f, t)) \end{array}$$

With the gun still being not broken by default and $S_1 = Do(\mathsf{Shoot}, S_0)$, we get the following conclusions: by default, the gun is not broken, even after shooting:

$$\Sigma \mathrel{|\!\!\approx}^{skept}_{\mathcal{D}[S_1]} \neg Holds(\mathsf{Broken}, S_1)$$

But then, it must have been loaded in the initial situation (otherwise it would be broken, which it is not):

$$\Sigma \mathrel{|\!\!\approx}^{skept}_{\mathcal{D}[S_1]} Holds(\mathsf{Loaded}, S_0),$$

although this was not known without utilizing a default about a situation in the future:

$$\Sigma \mathrel{|\!\!\not\approx}^{skept}_{\mathcal{D}[S_0]} Holds(\mathsf{Loaded}, S_0).$$

It might appear rather contrived to conclude the value of a fluent after applying an action that possibly affects it, but the point of the example should become clear: it is a counterexample for a "conditional effects" version of Theorem 5.

Conclusions and Future Work

The paper investigated the combination of two successful approaches to the logical formalization of commonsense reasoning, logics for actions and non-monotonic logics, and introduced a framework for default reasoning in action formalisms. Due to the restricted nature of the employed effect axioms and defaults, the proposed mechanism behaves in an intuitive way. It is even enough to apply default assumptions only to a single time point, namely the initial situation, without losing any of the conclusions. The restrictions made in the definitions were not arbitrary—loosening them results in counter-intuitive inferences, which has been shown via illustrative examples.

In the future, we aim at integrating the results into the concept of *Agent Logic Programs* (Drescher, Schiffel, and Thielscher 2009). Agent Logic Programs are definite logic programs with two special predicates that are evaluated with respect to an underlying domain axiomatization. We intend to augment ALPs by a negation-as-failure operator and combine the answer set semantics for general logic programs (Gelfond and Lifschitz 1991) with a background theory of action to provide a semantics for the augmented language.

References

Brewka, G., and Eiter, T. 1999. Prioritizing Default Logic: Abridged Report. In *Festschrift on the occasion of Prof. Dr. W. Bibel's 60th birthday*. Kluwer.

Brewka, G. 1989. Preferred Subtheories: An Extended Logical Framework for Default Reasoning. In *Proceedings of the Eleventh International Conference on Artificial Intelligence*, 1043–1048.

Drescher, C.; Schiffel, S.; and Thielscher, M. 2009. Agent Logic Programs. In *Proceedings of the Twenty-first International Joint Conference on Artificial Intelligence*. (Submitted).

Gelfond, M., and Lifschitz, V. 1991. Classical Negation in Logic Programs and Disjunctive Databases. *New Generation Computing* 9:365–385.

Hanks, S., and McDermott, D. 1987. Nonmonotonic Logic and Temporal Projection. *Artificial Intelligence* 33(3):379–412.

Kowalski, R. A., and Sergot, M. J. 1986. A Logic-based Calculus of Events. *New Generation Computing* 4(1):67–95.

McCarthy, J., and Hayes, P. J. 1969. Some Philosophical Problems from the Standpoint of Artificial Intelligence. In *Machine Intelligence*, 463–502. Edinburgh University Press.

McCarthy, J. 1963. Situations and Actions and Causal Laws. Stanford Artificial Intelligence Project: Memo 2.

McCarthy, J. 1998. Elaboration Tolerance. In progress.

Pirri, F., and Reiter, R. 1999. Some Contributions to the Metatheory of the Situation Calculus. *Journal of the ACM* 46(3):325–361.

Reiter, R. 1980. A Logic for Default Reasoning. *Artificial Intelligence* 13:81–132.

Reiter, R. 1991. The Frame Problem in the Situation Calculus: A Simple Solution (Sometimes) and a Completeness Result for Goal Regression. In *Artificial Intelligence and Mathematical Theory of Computation – Papers in Honor of John McCarthy*, 359–380. Academic Press.

Sandewall, E. 1994. The Range of Applicability for some Nonmonotonic Logics for Strict Inertia. *Journal of Logic and Computation* 4:581–616.

Thielscher, M. 1999. From Situation Calculus to Fluent Calculus: State Update Axioms as a Solution to the Inferential Frame Problem. *Artificial Intelligence* 111(1–2):277–299.

Thielscher, M. 2009. A Unifying Action Calculus. Artificial Intelligence. To appear.