# Knowledge Equivalence in Combined Action Theories

**Citation for published version:**
Petrick, R & Levesque, HJ 2002, Knowledge Equivalence in Combined Action Theories. in Proceedings of the Eights International Conference on Principles and Knowledge Representation and Reasoning (KR-02). pp. 303-314.

**Link:**
Link to publication record in Edinburgh Research Explorer

**Document Version:**
Peer reviewed version

**Published In:**
Proceedings of the Eights International Conference on Principles and Knowledge Representation and Reasoning (KR-02)

# Knowledge Equivalence in Combined Action Theories

**Ronald P. A. Petrick**
Department of Computer Science
University of Toronto
Toronto, Ontario, Canada M5S 3G4
rpetrick@cs.utoronto.ca

**Hector J. Levesque**
Department of Computer Science
University of Toronto
Toronto, Ontario, Canada M5S 3G4
hector@cs.utoronto.ca

## Abstract

We investigate the relationship between two accounts of knowledge and action in the situation calculus: the Scherl and Levesque (SL) approach that models knowledge with possible worlds, and the Demolombe and Pozos Parra (DP) approach that models knowledge by a set of "knowledge fluents." We construct *combined action theories*: basic action theories that encode a correspondence between an SL and a DP theory. We prove, subject to certain restrictions, that knowledge of fluent literals are provably the same after a sequence of actions. Moreover, this knowledge equivalence extends to a rich class of formulae. These results allow us to translate certain SL theories into equivalent DP theories that avoid the computational drawbacks of possible world reasoning. They also enable us to prove the correctness of the DP treatment of knowledge and action in terms of a possible world specification.

## 1 INTRODUCTION

Reasoning about *sensing as a form of action*, for the purpose of planning or high-level agent control, requires the ability to reason effectively about knowledge. Conceptually, reasoning about knowledge and action has been extensively studied and is relatively well understood. For example, Moore (1985) shows how the situation calculus can be adapted to knowledge using the accessibility relation over possible worlds (Hintikka, 1962). Scherl and Levesque (1993) extend Reiter's theory of action (Reiter, 2001a) to handle knowledge, thus providing a solution to the frame problem (McCarthy and Hayes, 1969) for knowledge change. A similar approach is explored by Thielscher (2000), where the fluent calculus is extended to include knowledge update axioms that model an agent's changing knowledge. In (Shapiro et al., 2000), sensing actions are generalized to manage possibly inaccurate beliefs in the situation calculus. Reiter (2001b) considers knowledge-based GOLOG programs with sensing actions. In (Baral and Son, 1997) a high-level action description language is presented that models sensing actions and a distinction between the state of the world and the knowledge of the world.

The approaches mentioned above all share a common treatment of knowledge: reasoning about knowledge is understood as reasoning about the accessibility relation over possible worlds, treated as a fluent that changes due to action. Computationally, this approach is not so promising. The difficulty is that determining if a formula is known then means determining if it is true in all of the currently accessible possible worlds. With $n$ atomic formulae, this means that there are potentially $2^n$ distinguishable worlds to check truth in. In other words, model checking of formulae about knowledge looks as bad as theorem-proving of formulae without knowledge, and theorem-proving of formulae with knowledge looks even worse. Therefore, even if we were to accept that a planner for ordinary actions based on a formalism like the situation calculus could be made practical, the addition of knowledge and sensing, modelled on possible worlds, raises new concerns.

Consequently, it is not too surprising that many of the attempts to construct planners to effectively manage sensing actions that we are aware of (e.g., (Bacchus and Petrick, 1998; Peot and Smith, 1992; Etzioni et al., 1997; Weld et al., 1998; Pryor and Collins, 1996)) have relied either on variants of STRIPS (Fikes and Nilsson, 1971) or special-purpose algorithmic treatments of knowledge. The trouble with these approaches, however, is in separating any formal semantics from the implementation details of the algorithms that the systems are built on. As a result, it is often quite hard to see how the work relates to a logical specification in a more general theory of knowledge and action.

It is possible, however, to formalize a limited concept of knowledge and sensing in a logical language of action like the situation calculus without using possible worlds. For example, Funge (1998) restricts knowledge to be about the values of real-valued functional fluents (e.g., distance, temperature, height). What is known is characterized not by an accessibility relation defining possible worlds, but rather by a set of upper and lower bounds that define intervals of possible values for these fluents. More qualitatively, Demolombe and Pozos Parra (2000) characterize knowledge of relational formulae by a set of fluents known true or known false. Instead of formalizing how the set of accessible possible worlds changes as the result of action, they propose to formalize how these "knowledge fluents" change individually. Both of these approaches are very attractive for two important reasons: first, the effect of sensing actions on knowledge is now very similar in form to the effect of ordinary actions on other fluents; second, reasoning about this type of knowledge change is now computationally no worse than reasoning about ordinary fluent change.

But what exactly do we give up in these accounts? What exactly is their relationship to the standard possible world one? In this paper, we propose a partial answer to these questions. We consider *combined action theories*, basic action theories that include axioms from both the Scherl and Levesque (henceforth SL) and the Demolombe and Pozos Parra (henceforth DP) theories of knowledge and action. Our combined action theories will encode a correspondence between an SL theory (using possible worlds) and a DP one (using knowledge fluents). We prove, subject to certain restrictions, that this correspondence maintains the property that fluent literals known are provably the same after a sequence of actions. Moreover, we show that this knowledge equivalence extends to a rich class of formulae. These results are important as they allow us to translate certain SL theories into equivalent DP theories that avoid the computational drawbacks of possible world reasoning. Unfortunately, differences in the expressive nature of the two approaches mean that this correspondence is not one to one. Our results do, however, enable us to prove the correctness of the DP treatment of knowledge and action in terms of the standard possible world specification.

The rest of the paper is organized as follows. In Section 2 we review the situation calculus and the SL and DP theories of knowledge and action. In Section 3 we introduce the notion of a combined action theory. In Section 4, we establish knowledge equivalence of fluent literals for certain classes of combined action theories and extend these results to more general first-order formulae. In Section 5 we give a comprehensive example illustrating our approach. Finally, in Section 6 we discuss some of the issues and possible extensions related to our work.

# 2 BACKGROUND

## 2.1 SITUATION CALCULUS

The situation calculus (as presented in (Reiter, 2001a)) is a first-order, many-sorted language (with some second-order features), specifically designed for modelling dynamically changing worlds. All changes to the world are the result of named *actions*. A first-order term called a *situation* is used to represent a possible world history (a sequence of actions). A special constant called $S_0$ indicates the *initial situation*, that is, the situation in which no actions have yet been performed. There is also a distinguished binary function symbol *do* such that $do(a, s)$ denotes the successor situation resulting from performing action $a$ in situation $s$. Actions are denoted by function symbols and may be parameterized, while situations are first-order terms. Relations (predicates) with the property that their truth values can change from situation to situation are referred to as (relational) *fluents*.[1] A fluent is denoted by including a situation argument as its last argument, indicating the value of the fluent at that situation.

Domain theories are specified by defining the following axioms:

- For each action $A$, an *action precondition axiom* of the form
$$Poss(A(\vec{x}), s) \equiv \Pi_A(\vec{x}, s).^2$$

- For each fluent $F$, a *successor state axiom* of the form
$$F(\vec{x}, do(a, s)) \equiv \gamma_F^+(\vec{x}, a, s) \vee F(\vec{x}, s) \wedge \neg\gamma_F^-(\vec{x}, a, s),$$
characterizing the conditions under which fluent $F$ is true at situation $do(a, s)$ as a function of situation $s$. $\gamma_F^+$ (similarly $\gamma_F^-$) describes all the ways of making $F$ true (false) in the situation $do(a, s)$ by executing $a$ in situation $s$.

- A set of first-order sentences describing the initial situation that syntactically only mention the situation term $S_0$.

Together with a set of unique-names axioms for primitive actions and a set of domain-independent foundational axioms (formally defining legal situations), this collection of axioms forms a *basic action theory*.

[1] *Functional fluents* are also permitted but we will restrict our attention to relational fluents only.

[2] Axioms that contain "free" variables can be thought of as being universally quantified from outside the axiom. Also, for simplicity we will assume that $Poss(A(\vec{x}), s) \equiv$ **true** for each action $A$. We will omit any discussion of the *Poss* predicate and assume that actions are always executable.

### 2.2 A $K$ FLUENT IN THE SITUATION CALCULUS

The situation calculus formalism in (Reiter, 2001a) does not distinguish between what is true in a situation and what is known in a situation. Scherl and Levesque (1993) formalize knowledge in the situation calculus by adapting a standard possible worlds model of knowledge as was done by Moore (1985). A binary relation $K(s', s)$ is introduced, read informally as "$s'$ is accessible from $s$," and treated like any other fluent (the last argument being the "official" situation argument).

Informally, $K(s', s)$ holds when as far as an agent in situation $s$ knows, it could be in situation $s'$. The expression[3] $\mathbf{Knows}_{SL}(\phi, s)$ is used to state that $\phi$ is known in situation $s$, where $\phi$ is a situation calculus formula with a special situation term "*now*." The notation $\phi[s]$ is used to indicate the formula that results from replacing *now* in $\phi$ by $s$. The expression $\mathbf{Knows}_{SL}(\phi, s)$ is then an abbreviation defined by

$$\mathbf{Knows}_{SL}(\phi, s) \stackrel{\text{def}}{=} (\forall s').K(s', s) \supset \phi[s'].$$

As with other relational fluents, the $K$ fluent possibly changes truth values due to action. The effects that actions have on $K$ are encoded by defining a successor state axiom of the form

$$\begin{aligned} K(s'', do(a, s)) \equiv\ & (\exists s').s'' = do(a, s') \wedge K(s', s) \wedge \\ & \forall((a = \alpha_1) \supset (\phi_1[s] \equiv \phi_1[s'])) \wedge \\ & \ldots \\ & \wedge \forall((a = \alpha_n) \supset (\phi_n[s] \equiv \phi_n[s'])). \end{aligned}$$

Here the $\alpha_i$ are *knowledge-producing* or sensing actions that inform the agent whether or not $\phi_i$ holds. $K$ is updated to reflect the situations now considered possible, depending on the type of action (knowledge-producing or ordinary).

A particular modal logic is modelled by including axioms that place restrictions on the $K$ accessibility relation. For instance, the S4 modal logic is modelled by including reflexivity and transitivity axioms. Scherl and Levesque also show that provided these properties hold of the $K$ relation in initial situations, then the $K$ relation in every situation resulting from an executable sequence of actions will also satisfy the same set of properties.

Finally, a basic action theory must include axioms that define the possible world alternatives for $K$ in the initial situation $S_0$. These specifications are necessary to define what is known and what is not known initially. To refer to these initial alternative situations, we include the expression *Init*$(s)$, to indicate "$s$ is an initial situation." Formally, we define *Init*$(s)$ as the abbreviation

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

### 2.3 KNOWLEDGE FLUENTS IN THE SITUATION CALCULUS

Demolombe and Pozos Parra (2000) present an alternate approach to modelling knowledge in the situation calculus.[4] A modal operator $K$ is introduced and "combined" syntactically with a non-equality fluent literal $P$ to form a *knowledge fluent* $KP$.[5] Informally, $KP(s)$ is a fluent meaning "$P$ is known to be true in situation $s$." These modal fluents are used to explicitly model knowledge without manipulating possible worlds, but restrict the expressive power of the representation to knowledge of literals.

For each ordinary fluent $F$, a pair of modal fluents, $KF$ and $K\neg F$, are defined. In addition to specifying a standard successor state axiom for $F$, *successor knowledge state axioms* must be given for both $KF$ and $K\neg F$. These axioms have the same form as regular successor state axioms,

$$\begin{aligned} & KF(\vec{x}, do(a, s)) \equiv \\ & \quad \gamma^+_{KF}(\vec{x}, a, s) \vee KF(\vec{x}, s) \wedge \neg\gamma^-_{KF}(\vec{x}, a, s), \\ & K\neg F(\vec{x}, do(a, s)) \equiv \\ & \quad \gamma^+_{K\neg F}(\vec{x}, a, s) \vee K\neg F(\vec{x}, s) \wedge \neg\gamma^-_{K\neg F}(\vec{x}, a, s), \end{aligned}$$

but must ensure that knowledge remains consistent. That is, both $KF(\vec{x}, s)$ and $K\neg F(\vec{x}, s)$ cannot hold in the same situation $s$.

Since knowledge fluents are ordinary situation calculus fluents, a basic action theory must include axioms defining $KF$ and $K\neg F$ at $S_0$. These axioms formally define what is initially known (or not known) about an ordinary fluent $F$.

## 3 PROPERTIES OF COMBINED ACTION THEORIES

As a first step towards relating the two accounts of knowledge, we begin by defining a *combined action theory*, a basic action theory that includes axioms for the $K$ fluent, successor state axioms for ordinary fluents, a set of successor knowledge state axioms for knowledge fluents, and restrictions on the set of initial situations *Init*$(s)$. A combined action theory will be used to encode a *translation* between

[3] We are freely changing the notation used by Scherl and Levesque.

[4] In (Demolombe and Pozos Parra, 2000), *belief* is modelled in a KD axiom system. We are instead modelling knowledge and have made notational changes to reflect this difference.

[5] We will use the term *fluent literal* to refer to a fluent $F(\vec{x}, s)$ or its negation $\neg F(\vec{x}, s)$, indicating that either form may be used. Similarly, for $KP(\vec{x}, s)$, where $P$ is a fluent literal (of $F$), we mean the corresponding knowledge fluent $KF(\vec{x}, s)$ or $K\neg F(\vec{x}, s)$.

the SL and DP axioms by specifying the form of the axioms we consider and the relationship between the SL and DP axioms. In this section we concentrate on the translation of knowledge-producing actions and initial situation axioms, but describe in general how the effects of ordinary physical actions are encoded. We will define 5 properties that any combined action theory must satisfy. In Section 4 we will consider *classes* of combined action theories based on certain restrictions to the successor state axioms. These restrictions will allow us to establish an equivalence between the SL and DP forms of knowledge.

We will assume that we have a finite number of knowledge-producing actions, $\alpha_1, \alpha_2, \ldots, \alpha_m$, and a finite number of physical actions, $\beta_1, \beta_2, \ldots, \beta_n$. We will treat each action as being distinct, and the physical actions as being distinct from the knowledge-producing actions.

### 3.1 REPRESENTATION OF SENSING ACTIONS

A combined action theory will contain a successor state axiom for $K$ that has the standard SL form. The first property we consider imposes additional restrictions on the form of the sensory effects that can be modelled. These constraints will allow us to translate the effects described in $K$ into appropriate successor knowledge state axioms for which we can establish a knowledge equivalence between the SL and DP forms of representation. Even with these restrictions, we will still be able to model a number of interesting sensory effects.

For instance, consider the axiom for $K$ defined by

$$K(s'', do(a, s)) \equiv (\exists s').s'' = do(a, s') \land K(s', s) \land \\ \varphi_1(a, s, s') \land \varphi_2(a, s, s') \land \varphi_3(a, s, s').$$

Say $\varphi_1$ defines a knowledge-producing action *sense*$_1$ as

$$\varphi_1(a, s, s') \stackrel{\text{def}}{=} a = sense_1(x) \supset (F(x, s) \equiv F(x, s')).$$

Here *sense*$_1$ is a simple action that unconditionally senses the truth value of a fluent $F$ for the specified $x$. A more complex action is given by:

$$\varphi_2(a, s, s') \stackrel{\text{def}}{=} a = sense_2 \supset (\forall x)(F(x, s) \equiv F(x, s')).$$

In this case the action *sense*$_2$ has a *universal* sensory effect. The universal quantification of $x$ results in the unconditional sensing of $F$ for each possible value of $x$. One additional type of sensing action is represented by:

$$\varphi_3(a, s, s') \stackrel{\text{def}}{=} a = sense_3(x) \supset ((G(x, s) \equiv G(x, s')) \land \\ G(x, s) \supset (F(x, s) \equiv F(x, s'))).$$

The action *sense*$_3$ has a compound effect: it unconditionally senses the truth value of $G$ (for the specified $x$) and also *conditionally* senses $F$, provided $G$ is true. In general this type of sensing allows additional properties about some set of objects to be sensed, contingent on the truth of some initial property. Our representation allows finite "chains" of this type of sensing to be modelled, and also allows situation independent formulae to be specified as conditions.

Formally, we have the following definition of the $K$ axiom.

**Property 1** Let $\alpha_1, \alpha_2, \ldots, \alpha_m$ be distinct knowledge-producing action terms. The successor state axiom for the $K$ fluent has the form

$$K(s'', do(a, s)) \equiv (\exists s').s'' = do(a, s') \land K(s', s) \land \\ \varphi_1(a, s, s') \land \varphi_2(a, s, s') \land \ldots \land \varphi_m(a, s, s').$$

For each $\varphi_i$ let $F_1, F_2, \ldots, F_l$ be distinct fluents so that

$$\varphi_i \stackrel{\text{def}}{=} \forall (a = \alpha_i(\vec{y}) \supset \psi_1 \land \psi_2 \land \ldots \land \psi_l),$$
$$\psi_j \stackrel{\text{def}}{=} (\forall \vec{z}).C_j(\vec{y}, \vec{z}, s) \supset \\ F_j(\vec{y}, \vec{z}, s) \equiv F_j(\vec{y}, \vec{z}, s').$$

$\psi_j$ describes an effect with *condition* $C_j$ and fluent $F_j$. $C_1$ is a situation independent formula. For $j > 1$, $C_j$ is either a situation independent formula or a conjunction of the form:

$$C_{j-1}(\vec{y}, \vec{z}, s) \land [\neg]F_{j-1}(\vec{y}, \vec{z}, s),$$

where $C_{j-1}$ is the condition associated with $\psi_{j-1}$.

### 3.2 SUCCESSOR KNOWLEDGE STATE AXIOMS

For every ordinary fluent $F$ in our SL theory, our combined action theory will include a pair of DP successor knowledge state axioms for $KF, K\neg F$. The second property we consider concerns the form of these axioms which encode all the effects of actions on the agent's knowledge of the fluent $F$. In other words, this encoding specifies the translation of the SL successor state axioms for $K$ and $F$ into DP axioms. Since we require a translation that preserves an equivalence with the effects described by the SL theory, we must consider two different types of effects: the effects of physical actions and the effects of knowledge-producing actions.

For physical actions, the equivalence is achieved by converting ordinary successor state actions into "knowledge fluent versions," through syntactic changes to fluent literals. All references to $P$ in $\gamma_F^{\pm}$ are changed to $KP$ without changing the underlying structure of $\gamma_F^{\pm}$ (i.e., the logical connectives). In the case of a situation independent formula, the conversion leaves the formula unchanged. (In Section 4 we will apply this syntactic conversion to restricted successor state axioms.)

For knowledge-producing actions, the equivalence depends on extracting the separate effects of all knowledge-producing actions on a particular fluent (defined in $K$) and packaging them together into the pair of corresponding successor knowledge state axioms. The appropriate components of the $K$ axiom (i.e., the specific effects that sense the fluent $F$) are incorporated into the successor knowledge state axioms, maintaining the same structure of the action terms and conditions on conditional effects that are defined for $K$. Any explicit quantification becomes implicitly quantified in the successor knowledge state axiom.

Consider the actions *sense*$_1$, *sense*$_2$, and *sense*$_3$, defined in Section 3.1. Assuming no other actions sense $F$ we can generate the corresponding pair of successor knowledge state axioms for $KF, K\neg F$:

$$\begin{aligned}
&KF(x, do(a, s)) \equiv \\
&\quad (\gamma_F^+)^K \vee ((a = sense_1(x) \vee a = sense_2 \vee \\
&\quad (a = sense_3(x) \wedge G(x, s))) \wedge F(x, s)) \vee \\
&\quad KF(x, s) \wedge \neg(\gamma_F^-)^K, \\
&K\neg F(x, do(a, s)) \equiv \\
&\quad (\gamma_F^-)^K \vee ((a = sense_1(x) \vee a = sense_2 \vee \\
&\quad (a = sense_3(x) \wedge G(x, s))) \wedge \neg F(x, s)) \vee \\
&\quad K\neg F(x, s) \wedge \neg(\gamma_F^+)^K.
\end{aligned}$$

($\gamma_F^\pm$ is defined in the successor state axiom for $F$.) Note that the explicit universal quantification in $\varphi_2$ is now expressed implicitly in the successor knowledge state axioms.

We formally define the translation of successor state axioms to successor knowledge state axioms as follows.

**Property 2** For each ordinary fluent $F$, the successor knowledge state axioms for knowledge fluents $KF, K\neg F$ are of the form

$$\begin{aligned}
&KF(\vec{x}, do(a, s)) \equiv \\
&\quad \gamma_{KF}^+(\vec{x}, a, s) \vee KF(\vec{x}, s) \wedge \neg\gamma_{KF}^-(\vec{x}, a, s), \\
&K\neg F(\vec{x}, do(a, s)) \equiv \\
&\quad \gamma_{KF}^-(\vec{x}, a, s) \vee K\neg F(\vec{x}, s) \wedge \neg\gamma_{KF}^+(\vec{x}, a, s),
\end{aligned}$$

and $\gamma_{KF}^\pm(\vec{x}, a, s)$ has the form $(\gamma_F^\pm)^K \vee \xi_F^\pm$.

$(\gamma_F^\pm)^K$ is structurally identical to $\gamma_F^\pm$ with the exception that every fluent literal $P$ is syntactically replaced by $KP$. $\xi_F^\pm$ has the form

$$\bigvee_{\langle i,j\rangle \in S_F} (a = \alpha_i(\vec{y}) \wedge C_j(\vec{y}, \vec{z}, s) \wedge \pm F(\vec{y}, \vec{z}, s)),$$

where $S_F$ is defined for each fluent $F$ as

$$\begin{aligned}
S_F \stackrel{\text{def}}{=} \{\langle i, j\rangle \mid \varphi_i &\text{ has an effect } \psi_j \text{ with condition} \\
C_j &\text{ and fluent } F (\text{defined in the } K \text{axiom})\}.
\end{aligned}$$

$S_F$ indicates the components of the $K$ axiom that sense $F$. $\alpha(\vec{y})$ and $C_j$ are structurally identical to those defined in Property 1. If no knowledge-producing action has an effect on a fluent $F$, then $\gamma_{KF}^\pm$ reduces to $(\gamma_F^\pm)^K$.

### 3.3 CONSTRAINTS ON INITIAL SITUATIONS

We now consider the three final properties required of a combined action theory, dealing with initial situations. First, since we are modelling knowledge we require that a reflexivity restriction hold of the $K$ fluent. As shown in (Scherl and Levesque, 1993) we only require that this property hold of initial situations for it to hold for all situations. A consequence of reflexivity, however, is that our initial knowledge must correspond correctly to the initial values of ordinary fluents (i.e., the way the real world is initially configured). Formally, we require the following conditions hold:

**Property 3** *(reflexivity of $K$)*

$$\begin{aligned}
&\Sigma \models (\forall s).Init(s) \supset K(s, s), \\
&\Sigma \models (\forall s)(\forall \vec{x}).Init(s) \supset \\
&\qquad \mathbf{Knows}_{SL}(P(\vec{x}, now), s) \supset P(\vec{x}, s),
\end{aligned}$$

for every fluent literal $P$.

Second, we require a knowledge equivalence for initial situations to ensure that we begin with literal-based knowledge that is identical in terms of both the SL (using $K$ and possible worlds) and DP (using knowledge fluents) forms of representation. Our goal in Section 4 will be to show that this equivalence is preserved through action, subject to certain restrictions that we place on the form of the combined action theory. Formally, we require the following property:

**Property 4** *(initial knowledge equivalence)* For every fluent literal $P$,

$$\begin{aligned}
&\Sigma \models (\forall s)(\forall \vec{x}).Init(s) \supset \\
&\qquad \mathbf{Knows}_{SL}(P(\vec{x}, now), s) \equiv KP(\vec{x}, s).
\end{aligned}$$

Finally, we require a strong restriction on our initial knowledge to ensure that we can "break apart" any knowledge of disjunctions to reason about the knowledge of the individual disjuncts. In general, SL theories can model knowledge of disjunctions without requiring knowledge of individual disjuncts.

**Example 1** Consider the following axioms:

$$\begin{aligned}
&F(do(a, s)) \equiv F(s), \\
&G(do(a, s)) \equiv G(s), \\
&K(s'', s) \equiv (\exists s').s'' = do(a, s') \wedge K(s', s) \wedge \\
&\quad (a = sense \supset ((F(s) \vee G(s)) \equiv (F(s') \vee G(s')))), \\
&(\exists s_1, s_2, s_3, s_4).K(s_1, S_0) \wedge K(s_2, S_0) \wedge K(s_3, S_0) \wedge \\
&\quad K(s_4, S_0) \wedge F(s_1) \wedge G(s_1) \wedge F(s_2) \wedge \neg G(s_2) \wedge \\
&\quad \neg F(s_3) \wedge G(s_3) \wedge \neg F(s_4) \wedge \neg G(s_4).
\end{aligned}$$

Initially, nothing is known about the fluents $F$ and $G$, however, in the situation $S = do(sense, S_0)$ we have that $\mathbf{Knows}_{SL}((F(now) \vee G(now)), S)$ holds, but neither $\mathbf{Knows}_{SL}(F(now), S)$ nor $\mathbf{Knows}_{SL}(G(now), S)$ hold.

In DP theories, however, the representation is restricted to knowledge of fluent literals. Thus, we require a disjunctive knowledge restriction to ensure that we can establish an equivalence of literal-based knowledge that can be preserved after a sequence of actions.

**Property 5** *(initial disjunctive knowledge)* For all fluent literals $P_1, P_2, \ldots, P_k$ that are not complementary,[6] and any ground sequence of actions $\vec{A}_1, \vec{A}_2, \ldots, \vec{A}_k$,

$$\Sigma \models (\forall s)(\forall \vec{x}).\mathit{Init}(s) \supset \mathbf{Knows}_{SL}(\bigvee_{i=1}^{k} P_i(\vec{x}, \mathit{do}(\vec{A}_i, \mathit{now})), s) \equiv \bigvee_{i=1}^{k} \mathbf{Knows}_{SL}(P_i(\vec{x}, \mathit{do}(\vec{A}_i, \mathit{now})), s).$$

This property not only specifies that we can break apart "immediate" disjunctive knowledge (e.g., formulae such as $\mathbf{Knows}_{SL}(P(\mathit{now}), S_0)$ that include *now* but no other action terms) but that we can also do the same for knowledge of "future" disjunctions (e.g., formulae such as $\mathbf{Knows}_{SL}(P(\mathit{do}(\vec{A}, \mathit{now})), S_0)$ that include an action sequence $\vec{A}$). It is this second condition that is important for ensuring a literal-based knowledge equivalence can be maintained through action. It also means, however, that we impose strong restrictions on the structure of our initial situations. This issue will be discussed further in Section 4.4.

The strength of Property 5 allows us to extend it to hold for all situations, not just initial situations, given a successor state axiom for $K$ of the form in Property 1.

**Theorem 1** *Let $\Sigma$ be a basic action theory that satisfies Properties 1 and 5. Then Property 5 holds for all situations, not just $\mathit{Init}(s)$.*

**Proof** (By induction over situations) The base case follows directly from Property 5. In the induction step we consider the two types of actions. For physical actions, using the definitions of $\mathbf{Knows}_{SL}$, the $K$ axiom from Property 1, and the fact that if $\mathbf{Knows}_{SL}(P(\vec{c}, \mathit{do}(A, \mathit{now})), s)$ holds then $\mathbf{Knows}_{SL}(P(\vec{c}, \mathit{now}), \mathit{do}(A, s))$ holds for all $s$ and $A$, the result quickly follows. For sensing actions, the $K$ axiom specifies that all knowledge-producing effects reduce to sensing the truth of individual fluent literals, thus preserving the required property. ■

This property will also be required to extend our equivalence results to more general formulae (see Section 4.3).

### 3.4 COMBINED ACTION THEORIES

We are now able to give a formal definition of a combined action theory, based on the properties described in Sections 3.1–3.3.

[6] That is, we cannot include both $P_i$ and $\neg P_i$.

**Definition 1** A *combined action theory* $\Sigma$ is a basic action theory that satisfies Properties 1–5.

Note that our definition does not specifically define the form of the ordinary successor state axioms (with the exception of $K$). It does, however, specify how such axioms will be converted to successor knowledge state axioms. In the next section we focus on the restrictions we require of successor state axioms.

## 4 KNOWLEDGE EQUIVALENCE IN COMBINED ACTION THEORIES

In general, our combined action theory alone is not enough to establish a knowledge equivalence between SL and DP theories, even with the strong restrictions placed on the initial situations.

**Example 2** Consider the following axioms:

$$F(\mathit{do}(a, s)) \equiv (a = A \wedge \neg F(s)) \vee F(s),$$
$$\neg \mathbf{Knows}_{SL}(F(\mathit{now}), S_0),$$
$$\neg \mathbf{Knows}_{SL}(\neg F(\mathit{now}), S_0).$$

In terms of literal-based knowledge, nothing is known about $F$ at $S_0$. However, in the situation $S = \mathit{do}(A, S_0)$, $\mathbf{Knows}_{SL}(F(\mathit{now}), S)$ holds. In this case, knowledge of $F$ at $S$ does not depend on knowing individual literals (i.e., knowing $F$ holds at all possible worlds). Rather, it involves a property that is true of each possible world, in this case a "hidden" tautology (i.e., $F \vee \neg F$ holds at all possible worlds).[7] It is this general representation of knowledge, allowable in SL theories, that poses a problem for DP theories since DP theories are restricted to knowledge of literals and unable to encode such knowledge.

Thus, to ensure a translation between the SL and DP forms of knowledge can be achieved, we are faced with the task of either first removing the hidden logical constraints (such as tautologies) from a theory and constructing a new, logically equivalent theory, or restricting the form of the theories we consider to avoid such issues entirely. We adopt the latter approach and consider restrictions to the form of the successor state axioms that allow us to define *classes* of combined action theories.

### 4.1 CONTEXT FREE THEORIES

The first class of combined action theories we investigate is formed by restricting our successor state axioms to be *context free*:

[7] Note that our disjunctive knowledge restriction does not forbid this.

**Definition 2** (following (Lin and Reiter, 1997)) A successor state axiom for a fluent $F$ is *context free* iff it has the form

$$F(\vec{x}, do(a, s)) \equiv \gamma_F^+(\vec{x}, a) \vee F(\vec{x}, s) \wedge \neg\gamma_F^-(\vec{x}, a),$$

where $\gamma_F^+(\vec{x}, a)$ and $\gamma_F^-(\vec{x}, a)$ are situation independent formulae whose free variables are among those in $\vec{x}$, a.

**Definition 3** A *context free combined action theory* $\Sigma$ is a combined action theory with the property that successor state axioms for ordinary fluents $F$ are context free (i.e., $\gamma_F^+$ and $\gamma_F^-$ are situation independent).

A context free successor state axiom for a fluent $F$ prohibits any references to fluents in $\gamma_F^\pm$. Even with these restrictions, axioms of this form are common. Quantification is still permitted, provided the scope of the quantifiers only range over the situation independent formulae. By requiring successor state axioms be context free, however, we are also placing restrictions on the form of the successor knowledge state axioms (at least the part determined by the effects of physical actions on a fluent). In this case, the physical effect portion of $\gamma_{KF}^\pm$ (i.e., $(\gamma_F^\pm)^K$) will also be context free and, in fact, identical to $\gamma_F^\pm$.

For instance, consider the following context free successor state axiom for an ordinary fluent *broken*:

$$\begin{aligned} &broken(x, do(a, s)) \equiv (a = drop(x) \wedge fragile(x)) \vee \\ &\qquad broken(x, s) \wedge \neg(a = repair(x)). \end{aligned}$$

Assuming that no actions also sense *broken*, Definition 3 allows us to generate a corresponding pair of successor knowledge state axioms based solely on the definition of $\gamma_{broken}^\pm$ given above. The resulting axioms for $Kbroken, K\neg broken$ (following the form specified in Definition 1) are also context free:

$$\begin{aligned} &Kbroken(x, do(a, s)) \equiv (a = drop(x) \wedge fragile(x)) \vee \\ &\qquad Kbroken(x, s) \wedge \neg(a = repair(x)), \\ &K\neg broken(x, do(a, s)) \equiv a = repair(x) \vee \\ &\qquad K\neg broken(x) \wedge \neg(a = drop(x) \wedge fragile(x)). \end{aligned}$$

These restrictions enable us to establish our first equivalence result between the SL and DP definitions of knowledge, not just for initial situations, but for every situation:

**Theorem 2** *Let $\Sigma$ be a context free combined action theory. Then for any fluent literal $P$,*

$$\Sigma \models (\forall s)(\forall \vec{x}).\mathbf{Knows}_{SL}(P(\vec{x}, now), s) \equiv KP(\vec{x}, s).$$

**Proof** (By induction over situations) The base case follows directly from Definition 1. In the induction step we consider sensing and physical actions separately. For sensing actions, the basic argument follows from the form of the successor state and successor knowledge state axioms: successor state axioms leave the truth of all ordinary fluents unchanged and the translation in Definition 1 ensures that the corresponding components of the successor state and successor knowledge state axiom will necessarily hold in both axioms if they hold in one axiom. Reflexivity and the definition of $\mathbf{Knows}_{SL}$ ensures we can establish the truth of fluents in the "real" situation. For physical actions, we use the property that $\Sigma \models (\forall s).\mathbf{Knows}_{SL}(\psi, s) \equiv \psi$, when $\psi$ is situation independent. Since $(\gamma_F^\pm)^K$ and $\gamma_F^\pm$ are identical situation independent formulae, and the induction assumption lets us convert between SL and DP knowledge for fluent literals, the result quickly follows from the correspondence between the successor state and successor knowledge state axioms. ■

This result means that as far as knowledge of fluent literals is concerned, the SL and DP accounts are identical and will remain identical after any executable sequence of actions. In practical terms this means that we can exchange an SL theory based on possible worlds for a corresponding DP theory based on knowledge fluents (e.g., the DP axioms for $Kbroken, K\neg broken$ replace the SL axioms for *broken* and $K$), provided we can accept the limitation of literal-based knowledge.

### 4.2 LITERAL-BASED THEORIES

We now consider a much more expressive class of combined action theories, formed by extending our successor state axioms to include fluent literals.

Our definition of a literal-based combined action theory forces $\gamma_F^\pm$ to be described in a disjunctive normal form, subject to certain restrictions. As with the context free case, quantifiers are allowed, provided their scope only ranges over the situation independent formulae. (An exception is made for variables $\vec{y}$ that appear as parameters to the action.) Additional restrictions ensure that no problematic logical constraints (such as tautologies) arise.

**Definition 4** A *literal-based combined action theory* $\Sigma$ is a combined action theory where the successor state axioms for ordinary fluents $F$ have the property that

$$\gamma_F^\pm(\vec{x}, a, s) \stackrel{\text{def}}{=} \textstyle\bigvee_{i=1}^k \pi_i(\vec{x}, a, s).$$

For each $\pi_i$, let $P_1, P_2, \ldots, P_l$ be fluent literals, $\psi_i$ a situation independent formula, $\beta_i$ a physical action term, and $\vec{y}_i$ a vector of variables (possibly empty) so that

$$\begin{aligned} &\pi_i(\vec{x}, a, s) \stackrel{\text{def}}{=} (\exists \vec{y}_i).a = \beta_i(\vec{x}, \vec{y}_i) \wedge \psi_i(\vec{x}, \vec{y}_i, a) \wedge \\ &\qquad P_1(\vec{x}, \vec{y}_i, s) \wedge P_2(\vec{x}, \vec{y}_i, s) \wedge \ldots \wedge P_l(\vec{x}, \vec{y}_i, s), \end{aligned}$$

where $\vec{y}_i$ must be a parameter of $\beta_i$, and $\neg F$ (similarly $F$) can't be mentioned in $\gamma_F^+$ (similarly $\gamma_F^-$). We also require the following property hold of every $\pi_i, \pi_j, i \neq j$: for every substitution $\sigma$ of $\vec{x}, a, \vec{y}_i, \vec{y}_j$ so that

$$\Sigma \models (\beta_i(\vec{x}, \vec{y}_i) = \beta_j(\vec{x}, \vec{y}_j) \wedge \psi_i(\vec{x}, \vec{y}_i) \wedge \psi_j(\vec{x}, \vec{y}_j))\sigma,$$

then for all fluent literals $P$ in $\pi_i$ and $R$ in $\pi_j$, (i) if $\pi_i$ is in $\gamma_F^+$ and $\pi_j$ is in $\gamma_F^-$: $P\sigma$ cannot unify with $R\sigma$, otherwise (ii) $P\sigma$ cannot unify with $\neg R\sigma$.

In practice, these constraints are conservative, yet many of the successor state axioms that occur in the literature can be converted to this form. For instance, consider the following successor state axiom for a fluent *holding*:

$$\begin{aligned} &\mathit{holding}(x, \mathit{do}(a, s)) \equiv (a = \mathit{pickup}(x) \vee \\ &\quad (\exists y).a = \mathit{pickup}(y) \wedge \mathit{in}(x, y, s)) \vee \\ &\quad \mathit{holding}(x, s) \wedge \neg(a = \mathit{dropall}). \end{aligned}$$

Following our definition, successor knowledge state axioms must encode knowledge fluent versions of the ordinary successor state axioms. References to fluent literals $P$ in $\gamma_F^{\pm}$ are syntactically replaced by references to $KP$ in $(\gamma_F^{\pm})^K$. For the *holding* example, assuming no actions also sense *holding*, we have the following successor knowledge state axioms:

$$\begin{aligned} &\mathit{Kholding}(x, \mathit{do}(a, s)) \equiv (a = \mathit{pickup}(x) \vee \\ &\quad (\exists y).a = \mathit{pickup}(y) \wedge \mathit{Kin}(x, y, s)) \vee \\ &\quad \mathit{Kholding}(x, s) \wedge \neg(a = \mathit{dropall}), \\ &K\neg \mathit{holding}(x, \mathit{do}(a, s)) \equiv a = \mathit{dropall} \vee \\ &\quad K\neg \mathit{holding}(x, s) \wedge \neg(a = \mathit{pickup}(x) \vee \\ &\quad (\exists y).a = \mathit{pickup}(y) \wedge \mathit{Kin}(x, y, s)). \end{aligned}$$

This translation allows our equivalence result for context free theories to be extended to literal-based theories as well.

**Theorem 3** *Let $\Sigma$ be a literal-based combined action theory. Then for any fluent literal $P$,*

$$\Sigma \models (\forall s)(\forall \vec{x}).\mathbf{Knows}_{SL}(P(\vec{x}, \mathit{now}), s) \equiv KP(\vec{x}, s).$$

**Proof** (By induction over situations) The base case follows directly from Definition 1. In the induction step, consider two types of actions. For sensing actions, the proof is the same as in Theorem 2. For physical actions, in the if direction we repeatedly choose disjunctions of fluent literals from the successor state axiom that must be known. Our restrictions in Definition 4 allow us to apply Theorem 1 to break apart this knowledge into component parts. This process terminates with the fluents in some component of the axiom being known individually. Using the form of the corresponding successor knowledge state axioms and the induction assumption we are able to establish the result. In the only-if direction, the induction assumption applied to the appropriate components of the successor knowledge state axioms relates the knowledge fluents to SL knowledge. The result then follows by considering the form of the successor state axiom and its corresponding translation described in Definition 1. ■

### 4.3 EXTENDING KNOWLEDGE EQUIVALENCE TO FIRST-ORDER FORMULAE

Up to this point we have only established a knowledge equivalence between SL and DP theories for fluent literals. We now seek to extend that equivalence to account for more general first-order formulae. We begin by defining the expression $\mathbf{Knows}_{DP}(\phi, s)$, to indicate that $\phi$ is known (in the DP sense) in situation $s$:

**Definition 5** Let $F$ be a fluent and let $\phi$ and $\psi$ be first-order formulae that don't mention $K$ or any knowledge fluents $KF, K\neg F$. Then

1. $\mathbf{Knows}_{DP}(\phi, s) \stackrel{\text{def}}{=} \phi,$ if $\phi$ is situation independent,
2. $\mathbf{Knows}_{DP}(F(\vec{x}), s) \stackrel{\text{def}}{=} KF(\vec{x}, s),$
3. $\mathbf{Knows}_{DP}(\neg F(\vec{x}), s) \stackrel{\text{def}}{=} K\neg F(\vec{x}, s),$
4. $\mathbf{Knows}_{DP}(\neg\neg\phi, s) \stackrel{\text{def}}{=} \mathbf{Knows}_{DP}(\phi, s),$
5. $\mathbf{Knows}_{DP}(\phi \wedge \psi, s) \stackrel{\text{def}}{=}$ $\mathbf{Knows}_{DP}(\phi, s) \wedge \mathbf{Knows}_{DP}(\psi, s),$
6. $\mathbf{Knows}_{DP}(\neg(\phi \wedge \psi), s) \stackrel{\text{def}}{=}$ $\mathbf{Knows}_{DP}(\neg\phi, s) \vee \mathbf{Knows}_{DP}(\neg\psi, s),$
7. $\mathbf{Knows}_{DP}((\forall \vec{x}).\phi, s) \stackrel{\text{def}}{=} (\forall \vec{x}).\mathbf{Knows}_{DP}(\phi, s),$
8. $\mathbf{Knows}_{DP}(\neg(\forall \vec{x}).\phi, s) \stackrel{\text{def}}{=} (\exists \vec{x}).\mathbf{Knows}_{DP}(\neg\phi, s).$

Using Definition 5 we can now refer to DP knowledge beyond that of simple knowledge fluents. Since our definition of $\mathbf{Knows}_{SL}$ can already be applied to such general formulae, a reasonable question to ask is whether our equivalence results can also be extended to a more general class of formulae. We offer a partial answer to this question. First, we extend our results to disjunctive formulae:

**Lemma 1** *Let $\Sigma$ be a context free or literal-based combined action theory. Let $\phi$ be a disjunction of non-complementary ground fluent literals. Then*

$$\Sigma \models (\forall s).\mathbf{Knows}_{SL}(\phi, s) \equiv \mathbf{Knows}_{DP}(\phi, s).$$

**Proof** Let $\phi$ be a disjunction of the ground fluent literals $P_1, P_2, \ldots, P_k$ and let $s$ be any situation. By Theorem 1:
$\Sigma \models \mathbf{Knows}_{SL}(\phi, s) \equiv \bigvee_{i=1}^k \mathbf{Knows}_{SL}(P_i(\vec{c}_i, \mathit{now}), s).$
Since $\Sigma$ is a context free (similarly, literal-based) combined action theory, by Theorem 2 (similarly, Theorem 3):
$\Sigma \models \bigvee_{i=1}^k \mathbf{Knows}_{SL}(P_i(\vec{c}_i, \mathit{now}), s) \equiv \bigvee_{i=1}^k \mathbf{Knows}_{DP}(P_i(\vec{c}_i, \mathit{now}), s).$
Now, by applying Definition 5 we obtain the desired result:
$\Sigma \models \bigvee_{i=1}^k \mathbf{Knows}_{DP}(P_i(\vec{c}_i, \mathit{now}), s) \equiv \mathbf{Knows}_{DP}(\phi, s).$
■

Although Lemma 1 requires that a disjunctive formula be free of tautologies (in order to make use of our disjunctive knowledge restriction), we can use this lemma to establish the following general equivalence:

**Theorem 4** *Let $\Sigma$ be a context free or literal-based combined action theory. Let $\phi$ be any ground, quantifier-free first-order formula without $K$ or any knowledge fluents. Then, there is a logically equivalent formula $\phi'$ such that*

$$\Sigma \models (\forall s).\mathbf{Knows}_{SL}(\phi', s) \equiv \mathbf{Knows}_{DP}(\phi', s).$$

**Proof** The $\phi'$ in question will be the conjunction of the non-tautologous prime implicates of $\phi$. Thus, $\models \phi \equiv \phi'$. Denote the prime implicates by $\pi_1, \pi_2, \ldots, \pi_k$ and let $s$ be any situation. Since:

$\Sigma \models \mathbf{Knows}_{SL}(\bigwedge_{i=1}^{k} \pi_i, s) \equiv \bigwedge_{i=1}^{k} \mathbf{Knows}_{SL}(\pi_i, s)$

(a property of $\mathbf{Knows}_{SL}$), the clausal form of the prime implicates allows us to apply Lemma 1 so that we have:

$\Sigma \models \bigwedge_{i=1}^{k} \mathbf{Knows}_{SL}(\pi_i, s) \equiv \bigwedge_{i=1}^{k} \mathbf{Knows}_{DP}(\pi_i, s)$.

Now, applying Definition 5 establishes the result:

$\Sigma \models \bigwedge_{i=1}^{k} \mathbf{Knows}_{DP}(\pi_i, s) \equiv \mathbf{Knows}_{DP}(\bigwedge_{i=1}^{k} \pi_i, s)$.

■

This theorem illustrates that our equivalence results can be extended to the class of ground, quantifier-free formulae. In particular, a sentence can be formulated in such a way that it is known in the SL sense (with possible worlds) iff it is known in the DP sense (with knowledge fluents). Furthermore, we believe that this equivalence can be extended to include formulae containing quantifiers. For instance, the techniques used in (Levesque, 1998) could be adopted to deal with such formulae, by restricting them to be in a normal form.

While Theorem 4 does not allow quantification in general, provided we ensure that the scope of quantifiers range only over situation independent formulae (i.e., "closed" situation independent formulae) we can consider a simple extension to our equivalence results:

**Corollary 1** *Let $\Sigma$ be a context free or literal-based combined action theory. Let $\phi$ be any first-order sentence, without $K$ or any knowledge fluents, whose quantifiers only range over situation independent formulae. Then, there is a logically equivalent formula $\phi'$ such that*

$$\Sigma \models (\forall s).\mathbf{Knows}_{SL}(\phi', s) \equiv \mathbf{Knows}_{DP}(\phi', s).$$

**Proof** Put $\phi$ into a conjunctive normal form, keeping any situation independent formula closed. Break apart the conjunctions into knowledge of the component parts. Since $\Sigma \models (\forall s).\mathbf{Knows}_{SL}(\psi, s) \equiv \psi$, when $\psi$ is a situation independent formula (a property of $\mathbf{Knows}_{SL}$), the result quickly follows from Theorem 4 and Definition 5. ■

### 4.4 DISJUNCTIVE KNOWLEDGE AND INITIAL SITUATIONS

Our definition of a combined action theory enforces a strong property on disjunctive knowledge, namely that disjunctions (both immediate and future) can be broken apart into knowledge of the individual disjuncts. Moreover, provided that this property holds in all initial situations, it will also hold in all subsequent situations, independent of the combined action theory. But what exactly does this property tell us about the structure of $S_0$ and other initial situations? Is such a strong property necessary?

We begin by considering a much less restrictive property about disjunctive knowledge:

**Definition 6** Let $\Sigma$ be a basic action theory. A situation $s$ is said to satisfy the *weak disjunctive knowledge property* if for all fluent literals $P_1, P_2, \ldots, P_k$ that are not complementary,

$$\Sigma \models (\forall \vec{x}).\mathbf{Knows}_{SL}(\bigvee_{i=1}^{k} P_i(\vec{x}, \mathit{now}), s) \equiv \bigvee_{i=1}^{k} \mathbf{Knows}_{SL}(P_i(\vec{x}, \mathit{now}), s).$$

With this weaker form of disjunctive knowledge we no longer require constraints on "future" disjunctions, just immediate ones. It turns out that for the class of context free theories such a property is sufficient to maintain our equivalence results.

**Theorem 5** *Let $\Sigma$ be defined as in Definition 3 with the (strong) disjunctive knowledge property replaced with the weak disjunctive knowledge property on initial situations. Then, (i) the weak disjunctive knowledge property holds for all situations, and (ii) the equivalence results of Sections 4.1 and 4.3 extend to $\Sigma$.*

**Proof** The proof of (i) is straight-forward by induction over situations, using the form of the translation in Definition 1 when we have context free successor state axioms, and the property that $\Sigma \models (\forall s).\mathbf{Knows}_{SL}(\psi, s) \equiv \psi$, when $\psi$ is situation independent. For (ii), the proofs carry over from Sections 4.1 and 4.3 with all references to Theorem 1 replaced with references to Theorem 5(i). ■

Thus, for context free theories at least we need only be concerned about immediate disjunctions in the initial situation (i.e., those that only mention *now* and no other action terms). This weaker notion of disjunctive knowledge, however, is not necessarily preserved if we consider non-context free theories, even literal-based ones.

**Example 3** Consider the following axioms:

$$\begin{aligned}
&F(do(a, s)) \equiv (a = A \wedge G(s)) \vee F(s),\\
&G(do(a, s)) \equiv G(s),\\
&(\exists s_1, s_2, s_3, s_4).K(s_1, S_0) \wedge K(s_2, S_0) \wedge K(s_3, S_0) \wedge\\
&\quad K(s_4, S_0) \wedge F(s_1) \wedge G(s_1) \wedge F(s_2) \wedge \neg G(s_2) \wedge\\
&\quad \neg F(s_3) \wedge G(s_3) \wedge \neg F(s_4) \wedge \neg G(s_4).
\end{aligned}$$

Nothing is known initially about $F$ and $G$. (The specification of initial situations means that the weak disjunctive knowledge property holds of $S_0$.) In the situation $S = do(A, S_0)$, however, we have that $\mathbf{Knows}_{SL}((F(now) \vee \neg G(now)), S)$ holds, but neither $\mathbf{Knows}_{SL}(F(now), S)$ nor $\mathbf{Knows}_{SL}(\neg G(now), S)$ hold.

Thus, by considering even slightly more complex successor state axioms the weaker notion of disjunctive knowledge can quickly fail. Since we require such a property hold in order to establish our equivalence results, this motivates the need for our stronger restriction.

What this property does *not* provide, however, is an efficient method of detecting all the necessary conditions that must hold of an initial situation. In constructing an SL theory one must potentially consider disjunctions that arise from *any* sequence of actions and make sure that the appropriate knowledge is encoded in the initial situations. For instance, in Example 3 we would require that $\mathbf{Knows}_{SL}(F(do(A, now)), S_0)$ or $\mathbf{Knows}_{SL}(\neg G(do(A, now)), S_0)$ hold of $S_0$.

### 4.5 NON-EQUIVALENCE OF SL AND DP THEORIES

While we have been able to correlate the SL and DP approaches for an expressive class of theories, the equivalence of SL and DP theories is not one-to-one. Clearly, there exist SL theories without equivalent DP formulations (e.g., Example 1). The converse is also true. Depending on the form of the successor knowledge state axioms, DP theories can be modelled so that knowledge fluents evolve independent of ordinary fluents. Consequently, we can construct DP theories that manipulate knowledge in a way that cannot be easily reproduced in a standard SL theory.

**Example 4** Consider the following axioms:

$$\begin{aligned}
&F(do(a, s)) \equiv F(s),\\
&KF(do(a, s)) \equiv KF(s) \wedge \neg(a = forget),\\
&K\neg F(do(a, s)) \equiv K\neg F(s) \wedge \neg(a = forget),\\
&KF(S_0).
\end{aligned}$$

In the situation $S = do(forget, S_0)$, both $\neg KF(S)$ and $\neg K\neg F(S)$ hold. Thus, *forget* produces a *knowledge reducing* effect without changing any ordinary fluents. Such an action cannot be modelled directly in a standard SL theory (see the theorems concerning *memory* in (Scherl and Levesque, 1993)).

To make our equivalence more encompassing, one possibility is to extend the SL theory. For instance, a richer representation that allows actions such as *forget* to be modelled at the possible world level could provide a closer correspondence to the DP theory. We are currently investigating such an approach as well as alternate theories that could subsume the SL approach altogether (see Section 6).

## 5 AN EXAMPLE

One of our main objectives has been to provide a means of translating certain SL theories into equivalent DP theories that avoid the use of possible worlds. We now illustrate our approach with an example from the UNIX domain that involves both ordinary and knowledge-producing actions.

Consider the following UNIX-style domain involving two fluents, *indir* and *readable*, and two actions, *ls* and *mv*. The fluent $indir(f, d, s)$ can be understood as "file $f$ is in directory $d$ in situation $s$." The fluent $readable(f, s)$ indicates that "file $f$ is readable in situation $s$." The action $mv(f, d', d)$ is an ordinary (physical) action that has the effect of moving file $f$ from directory $d'$ to directory $d$. The action $ls(d)$ is a knowledge-producing action that provides information about the files in directory $d$. We encode the successor state axioms in our SL theory as follows:

$$\begin{aligned}
&indir(f, d, do(a, s)) \equiv\\
&\quad ((\exists d').a = mv(f, d', d) \wedge d \neq d' \wedge indir(f, d', s)) \vee\\
&\quad indir(f, d, s) \wedge \neg((\exists d').a = mv(f, d, d') \wedge d \neq d'),
\end{aligned}$$

$$readable(f, do(a, s)) \equiv readable(f, s),$$

$$\begin{aligned}
&K(s'', do(a, s)) \equiv (\exists s').s'' = do(a, s') \wedge K(s', s) \wedge\\
&\quad (\exists d).(a = ls(d) \supset\\
&\qquad (\forall f)\ (indir(f, d, s) \equiv indir(f, d, s')) \wedge\\
&\qquad (\forall f)\ (indir(f, d, s) \supset\\
&\qquad\quad (readable(f, s) \equiv readable(f, s'))))).
\end{aligned}$$

Our successor state axiom for $K$ encodes two types of knowledge-producing effects for *ls*. First, it encodes a universal effect: *ls* senses the files $f$ that are in directory $d$ (i.e., all $f$ that satisfy $indir(f, d, s)$). It also encodes a type of conditional sensing effect: besides sensing the contents of the directory, *ls* also senses the readability *of the files that are in directory* $d$ (i.e., $readable(f, s)$ for all $f$ such that $indir(f, d, s)$ is true). Using Definitions 1 and 4, we can translate the SL axioms into corresponding DP axioms:

$$\begin{aligned}
&Kindir(f, d, do(a, s)) \equiv\\
&\quad ((\exists d').a = mv(f, d', d) \wedge d \neq d' \wedge Kindir(f, d', s)) \vee\\
&\quad (a = ls(d) \wedge indir(f, d, s)) \vee\\
&\quad Kindir(f, d, s) \wedge \neg((\exists d').a = mv(f, d, d') \wedge d \neq d'),\\
&K\neg indir(f, d, do(a, s)) \equiv\\
&\quad ((\exists d').a = mv(f, d, d') \wedge d \neq d') \vee\\
&\quad (a = ls(d) \wedge \neg indir(f, d, s)) \vee K\neg indir(f, d, s) \wedge\\
&\quad \neg((\exists d').a = mv(f, d', d) \wedge d \neq d' \wedge Kindir(f, d', s)).
\end{aligned}$$

$$\begin{aligned}
&Kreadable(f, do(a, s)) \equiv\\
&\quad ((\exists d).a = ls(d) \wedge indir(f, d, s) \wedge readable(f, s)) \vee\\
&\quad Kreadable(f, s),\\
&K\neg readable(f, do(a, s)) \equiv\\
&\quad ((\exists d).a = ls(d) \wedge indir(f, d, s) \wedge \neg readable(f, s)) \vee\\
&\quad K\neg readable(f, s).
\end{aligned}$$

In the translation, the knowledge-producing effects of *ls* are distributed from the $K$ successor state axiom into the appropriate DP successor knowledge state axioms: the universal effect into $K\mathit{indir}$, $K\neg\mathit{indir}$ and the conditional effect into $K\mathit{readable}$, $K\neg\mathit{readable}$. The explicit universal quantification in the $K$ axiom is now expressed implicitly in the knowledge successor state axioms. The successor state axiom for *indir* is converted to its knowledge fluent version and also included in the axioms for $K\mathit{indir}$, $K\neg\mathit{indir}$.

We must also ensure that we have an initial knowledge equivalence for fluent literals. For instance, suppose nothing is known initially about the location of a file *kr.tex*. We have the following SL and DP axioms:

$$(\forall d).\neg\mathbf{Knows}_{SL}(\mathit{indir}(\mathit{kr.tex}, d, \mathit{now}), S_0) \wedge \\ \neg\mathbf{Knows}_{SL}(\neg\mathit{indir}(\mathit{kr.tex}, d, \mathit{now}), S_0), \\ (\forall d).\neg K\mathit{indir}(\mathit{kr.tex}, d, S_0) \wedge \neg K\neg\mathit{indir}(\mathit{kr.tex}, d, S_0).$$

Using the DP theory we can now reason about knowledge change as updates to the knowledge fluents. For example, consider the situation $S_1 = \mathit{do}(\mathit{mv}(\mathit{kr.tex}, \mathit{tmp}, \mathit{papers}), S_0)$. By the successor state axiom for $K\neg\mathit{indir}$, it will be the case that $K\neg\mathit{indir}(\mathit{kr.tex}, \mathit{tmp}, S_1)$ holds. However, it will also be the case that $\neg K\mathit{indir}(\mathit{kr.tex}, \mathit{papers}, S_1)$ and $\neg K\neg\mathit{indir}(\mathit{kr.tex}, \mathit{papers}, S_1)$ hold since initially it is not known whether *kr.tex* is in directory *tmp*. If we then consider the situation $S_2 = \mathit{do}(\mathit{ls}(\mathit{papers}), S_1)$ then either $K\mathit{indir}(\mathit{kr.tex}, \mathit{papers}, S_2)$ or $K\neg\mathit{indir}(\mathit{kr.tex}, \mathit{papers}, S_2)$ will hold (i.e., the agent will know whether *kr.tex* is in directory *papers*), depending on whether *kr.tex* is actually in directory *papers* or not. If $K\mathit{indir}(\mathit{kr.tex}, \mathit{papers}, S_1)$ holds, then either $K\mathit{readable}(\mathit{kr.tex}, S_1)$ or $K\neg\mathit{readable}(\mathit{kr.tex}, S_1)$ will also hold (i.e., the agent will know whether *kr.tex* is readable).

Moreover, our equivalence results ensure that the DP knowledge fluents can also be understood in terms of the SL theory. For instance, by Theorem 3 we will have that $\mathbf{Knows}_{SL}(\neg\mathit{indir}(\mathit{kr.tex}, \mathit{tmp}, \mathit{now}), S_1)$ holds. Also, both $\neg\mathbf{Knows}_{SL}(\mathit{indir}(\mathit{kr.tex}, \mathit{papers}, \mathit{now}), S_1)$, and $\neg\mathbf{Knows}_{SL}(\neg\mathit{indir}(\mathit{kr.tex}, \mathit{papers}, \mathit{now}), S_1)$ will hold.

## 6 DISCUSSION

In this paper we provide a means of translating certain types of SL theories into corresponding DP theories that avoid the use of possible worlds. As a result, reasoning about knowledge change reduces to reasoning about ordinary fluent change. With $n$ atomic formulae, determining the truth of a formula reduces from checking $2^n$ possible worlds to checking the truth of $3n$ fluents in the worst case. Moreover, we can make use of standard tools such as regression for addressing issues like the projection problem (Demolombe and Pozos Parra, 2000; Reiter, 2001a). From a practical standpoint, we believe our approach will lead to more efficient implementations of systems for high-level agent control or planning. Furthermore, we believe that the tradeoffs in expressiveness do not detract from the advantages of modelling certain types of problems at the knowledge level instead of the possible world level. Indeed, recent results in knowledge-based planning (Petrick and Bacchus, 2002) lend support to the viability of such an approach.

Our results can also be extended in a number of ways. Even with our current restrictions, we are still able to model powerful (and interesting) types of sensing, such as actions with universal sensory effects or a form of conditional sensing that allows fluents to be sensed, contingent on the truth of other fluents. We are exploring extensions to our combined action theories to model more comprehensive classes of sensing. For instance, our strong restrictions on disjunctive knowledge should allow us to extend our sensing to more general formulae. Likewise, a much more expressive class of physical effects could be modelled in our representation by considering less restrictive forms of quantification in successor state axioms. Such an addition, however, will require a strengthening of our disjunctive knowledge restriction, in particular, to include knowledge of existentially quantified formulae.

We also seek to extend our knowledge equivalence results to formulae with unrestricted quantification. This would allow us equate knowledge of formulae containing $K$ or $KP$ (i.e., introspective formulae), currently restricted by our representation. The techniques of (Levesque, 1998), including the normal form proposed by Levesque, could be adapted for this purpose. We are also looking at the possibility of modelling knowledge reducing actions such as *forget* (see Section 4.5) in our combined action theories to take advantage of the flexibility of the DP approach and to extend our correspondence with it. An interesting discussion of some of the issues concerned with "forgetting" is presented in (Lin and Reiter, 1994).

We are also able to relax some of our assumptions. We have ignored any discussion of action preconditions, however, a simple extension to allow knowledge-based action preconditions, for instance, could be made. We could also drop our restriction that ordinary actions be distinct from knowledge-producing actions, allowing actions to have both physical and sensory effects. Finally, we are also investigating the addition of functional fluents to the representation. These and other related issues will be discussed further in (Petrick, 2003).

## References


Bacchus, F. and Petrick, R. (1998). Modeling an agent's incomplete knowledge during planning and execution. In Cohn, A. G., Schubert, L., and Shapiro, S. C., editors, *Proc. of KR-98*, pages 432–443. Morgan Kaufmann Publishers.

Baral, C. and Son, T. C. (1997). Approximate reasoning about actions in presence of sensing and incomplete information. In Małuszyński, J., editor, *Proceedings of the International Symposium on Logic Programming (ILPS-97)*, pages 387–404. MIT Press.

Demolombe, R. and Pozos Parra, M. P. (2000). A simple and tractable extension of situation calculus to epistemic logic. In *Proceedings of the Twelfth International Symposium on Methodologies for Intelligent Systems (ISMIS-2000)*, pages 515–524.

Etzioni, O., Golden, K., and Weld, D. (1997). Sound and efficient closed-world reasoning for planning. *Artificial Intelligence*, 89(1–2):113–148.

Fikes, R. E. and Nilsson, N. J. (1971). STRIPS: A new approach to the application of theorem proving to problem solving. *Artificial Intelligence*, 2:189–208.

Funge, J. (1998). Interval-valued epistemic fluents. In *AAAI 1998 Fall Symposium on Cognitive Robotics*, pages 23–25, Orlando, FL.

Hintikka, J. (1962). *Knowledge and Belief*. Cornell University Press, Ithaca, New York.

Levesque, H. J. (1998). A completeness result for reasoning with incomplete first-order knowledge bases. In Cohn, A. G., Schubert, L., and Shapiro, S. C., editors, *Proc. of KR-98*, pages 14–23. Morgan Kaufmann Publishers.

Lin, F. and Reiter, R. (1994). Forget it! In Greiner, R. and Subramanian, D., editors, *Working Notes of AAAI Fall Symposium on Relevance*, pages 154–159, Menlo Park, CA, USA. AAAI.

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

McCarthy, J. and Hayes, P. J. (1969). Some philosophical problems from the standpoint of artificial intelligence. *Machine Intelligence*, 4:463–502.

Moore, R. C. (1985). A formal theory of knowledge and action. In Hobbs, J. R. and Moore, R. C., editors, *Formal Theories of the Commonsense World*, pages 319–358. Ablex Publishing.

Peot, M. and Smith, D. (1992). Conditional nonlinear planning. In Hendler, J., editor, *Proceedings of the First International Conference on AI Planning Systems*, pages 189–197. Morgan Kaufmann.

Petrick, R. P. A. (2003). *Modelling knowledge and sensing for effective reasoning*. PhD thesis, Department of Computer Science, University of Toronto. Forthcoming.

Petrick, R. P. A. and Bacchus, F. (2002). A knowledge-based approach to planning with incomplete information and sensing. In *Proc. of AIPS-2002*. To appear.

Pryor, L. and Collins, G. (1996). Planning for contingencies: A decision-based approach. *Journal of Artificial Intelligence Research*, 4:287–339.

Reiter, R. (2001a). *Knowledge In Action: Logical Foundations for Specifying and Implementing Dynamical Systems*. MIT Press.

Reiter, R. (2001b). On knowledge-based programming with sensing in the situation calculus. *ACM Transactions on Computational Logic*, 2(4):433–457.

Scherl, R. B. and Levesque, H. J. (1993). The frame problem and knowledge-producing actions. In *Proc. of AAAI-93*, pages 689–695. AAAI Press/MIT Press.

Shapiro, S., Lespérance, Y., Levesque, H. J., and Pagnucco, M. (2000). Iterated belief change in the situation calculus. In Cohn, A. G., Giunchiglia, F., and Selman, B., editors, *Proc. of KR-2000*, pages 527–538. Morgan Kaufmann Publishers.

Thielscher, M. (2000). Representing the knowledge of a robot. In Cohn, A. G., Giunchiglia, F., and Selman, B., editors, *Proc. of KR-2000*, pages 109–120. Morgan Kaufmann Publishers.

Weld, D. S., Anderson, C. R., and Smith, D. E. (1998). Extending graphplan to handle uncertainty & sensing actions. In *Proc. of AAAI-98*, pages 897–904, Menlo Park. AAAI Press.