Encoding probabilistic causal models in probabilistic action language PAL

Chitta Baral and Nam Tran
Department of Computer Science and Engineering
Arizona State University
Tempe, Arizona 85287
{*chitta,namtran*}*@asu.edu*

Abstract

Pearl's probabilistic causal model has been used in many domains to reason about causality. Pearl's treatment of actions is very different from the way actions are represented (explicitly) and their impact is reasoned in most other papers in the literature. In this paper we show how to encode Pearl's probabilistic causal model in the action language PAL thus relating this two distinct approaches to reason about actions.

1 Introduction and motivation

Normally an action when executed in a world changes the state of the world. Reasoning about actions is important in several 'intelligent' tasks such as planning, hypothetical reasoning, control generation and verification (for dynamical systems), and diagnosis. Often the effect of an action on the world is not deterministic but rather has an uncertainty associated with it. In recent years there have been several proposals (for example, [Pearl, 1999; 2000; Baral *et al.*, 2002; Boutilier and Goldszmidt, 1996; Reiter, 2001; Littman, 1997]) to represent and reason with such actions. In all these proposals, except in [Pearl, 1999; 2000; 1995], actions have explicit names and their effects on the world are described by various means. In [Pearl, 1999; 2000] the dynamics of the world is described through relationships between fluents or variables (which denote properties of objects in the world) that are expressed through functional relationships between them. In addition probabilities are associated with a set of fluents referred to as exogenous (or unknown) variables. Together they are referred to as *probabilistic causal models*. The effect of actions are then formulated as "local surgery"on these models.

In this paper our goal is to study the relationship between reasoning about actions in probabilistic causal models and similar reasoning in the action description language PAL [Baral *et al.*, 2002] as a representative of the approaches where actions are named and have effects associated with it. The motivation behind studying this relationship is to objectively compare the expressiveness of these two formalisms vis-a-vis each other. We pick *probabilistic causal models* as it is the most recent representative of its class and the award winning book [Pearl, 2000] is based on it. We pick PAL as a representative of the high level action description languages [Gelfond and Lifschitz, 1993; McCain and Turner, 1995; Gelfond and Lifschitz, 1998] as among the action theories where actions have distinct names, this is the most similar to *probabilistic causal models* in the way it handles uncertainty. Also, most other action formalisms that incorporate uncertainty are limited in the sense that they are solely geared towards planning.

The rest of this paper is organized as follows. In Section 2 we give a brief overview of PAL. In Section 3 we give a brief overview of probabilistic causal models (PCM). We then (in Section 4) present an encoding of PCM in PAL. We then state and prove a result that shows the equivalence of various reasonings in PCM and the corresponding reasonings in the encoded PAL. We illustrate the encoding and the logic behind the proof through an example. Finally, we conclude and discuss some implications of this result and future works in Section 6.

2 The Language PAL: a brief overview

The alphabet of the language PAL consists of four non-empty disjoint sets of symbols $\mathbf{F}$, $\mathbf{U}_I$, $\mathbf{U}_N$ and $\mathbf{A}$. They are called the set of fluents, the set of *inertial* unknown variables, the set of *non-inertial* unknown variables and the set of actions. The unknown variables are assumed to be independent of each other. A *fluent literal* is a fluent or a fluent preceded by $\neg$. An unknown variable literal is an unknown variable or an unknown variable preceded by $\neg$. A *literal* is either a fluent literal or an unknown variable literal. A *formula* is a propositional formula constructed from literals.

A state s is an interpretation of fluents and unknown variables that satisfy certain conditions (to be mentioned while discussing semantics); For a state s, the sub-interpretations of s restricted to fluents, inertial unknown variables, and non-inertial unknown variables are denoted by s_F, s_I, and s_N respectively. PAL has four components: a domain description language PAL_D, a language PAL_P to express unconditional probabilities about the unknown variables, a language PAL_O to specify observations, and a query language.

2.1 PAL_D: The domain description language

A *domain description* is a collection of propositions of the following forms:

$$a \textbf{ causes } \psi \textbf{ if } \varphi \quad (2.1)$$
$$\theta \textbf{ causes } \psi \quad (2.2)$$
$$\textbf{impossible } a \textbf{ if } \varphi \quad (2.3)$$

where a is an action, ψ is a *fluent* formula, θ is a formula of fluents and *inertial* unknown variables, and φ is a formula of fluents and *unknown* variables.

Propositions of the form (2.1) describe the direct effects of actions on the world and are called *dynamic causal laws*. Propositions of the form (2.2), called *static causal laws*, describe causal relation between fluents and unknown variables in a world. Propositions of the form (2.3), called *executability conditions*, state when actions are not executable.

Semantics: Characterizing the transition function: Let $\mathcal{D}$ be a domain description in the language of PAL_D. An *interpretation* I of the fluents and unknown variables in PAL_D is a maximal consistent set of literals of PAL_D. A literal l is said to be true (resp. false) in I iff $l \in I$ (resp. $\neg l \in I$). The truth value of a formula in I is defined recursively over the propositional connective in the usual way. For example, $f \wedge q$ is true in I iff f is true in I and q is true in I. The formula ψ is said to hold in I (or I satisfies ψ), denoted by $I \models \psi$, if ψ is true in I.

A set of formulas from PAL_D is *logically closed* if it is closed under propositional logic (w.r.t. PAL_D).

Let V be a set of formulas and K be a set of static causal laws of the form θ **causes** ψ. V is said to be closed under K if for every rule θ **causes** ψ in K, if θ belongs to V then so does ψ. $Cn_K(V)$ denotes the least logically closed set of formulas from PAL_D that contains V and is also closed under K.

A *state* s of $\mathcal{D}$ is an interpretation that is closed under the set of static causal laws of $\mathcal{D}$.

An action a is *prohibited (not executable)* in a state s if there exists in $\mathcal{D}$ an executability condition of the form **impossible** a **if** φ such that φ holds in s.

The *effect of an action a* in a state s is the set of formulas $E_a(\mathsf{s}) = \{\psi \mid \mathcal{D}$ contains a law a **causes** ψ **if** φ and φ holds in $\mathsf{s}\}$.

Let $\mathcal{S}$ be the set of the states of $\mathcal{D}$. A transition function Φ is a function from $\mathcal{A} \times \mathcal{S}$ to $2^{\mathcal{S}}$. If a is not prohibited (i.e.,executable) in s, then

$$\Phi(a, \mathsf{s}) = \{\ \mathsf{s}' \mid \mathsf{s}'_{F,I} = Cn_R((\mathsf{s}_{F,I} \cap \mathsf{s}'_{F,I}) \cup E_a(\mathsf{s}))\ \}; \quad (2.4)$$

If a is prohibited (i.e., not executable) in s, then $\Phi(a, \mathsf{s})$ is $\emptyset$. Every domain description $\mathcal{D}$ in a language PAL_D has a unique transition function Φ.

An extended transition function $\hat{\Phi}$ expresses the state transition due to a sequence of actions.

Definition 1 $\hat{\Phi}([a], \mathsf{s}) = \Phi(a, \mathsf{s})$;
$\hat{\Phi}([a_1, \ldots, a_n], \mathsf{s}) = \bigcup_{\mathsf{s}' \in \hat{\Phi}(a_1, \mathsf{s})} \hat{\Phi}([a_2, \ldots, a_n], \mathsf{s}')$ □

Definition 2 Given a domain description $\mathcal{D}$, and a state s, $\mathsf{s} \models_{\mathcal{D}} \varphi$ **after** $a_1, \ldots, a_n$,
if φ is true in all states in $\Phi([a_1, \ldots, a_n], \mathsf{s})$.
(Often when it is clear from the context $\models$ is written instead of $\models_{\mathcal{D}}$.) □

2.2 PAL_P: Probabilities of unknown variables

A probability description $\mathcal{P}$ of the unknown variables is a collection of propositions of the following form:

$$\textbf{probability of } u \textbf{ is } p \quad (2.5)$$

where u is an unknown variable, and $p \in [0, 1]$.

Semantics: Each proposition above directly gives us the probability distribution of the corresponding unknown variable as: $P(u) = p$.

For any state s, s_u denotes the interpretation of the unknown variables of s, that is, $\mathsf{s}_u = \mathsf{s}_I \cup \mathsf{s}_N$. The unconditional probability of the various states is defined as:

$$P(\mathsf{s}) = \frac{P(\mathsf{s}_u)}{|\{\mathsf{s}' | \mathsf{s}'_u = \mathsf{s}_u)\}|} \quad (2.6)$$

2.3 PAL_Q: The Query language

A query is of the form:

$$\textbf{probability of } [\varphi \textbf{ after } a_1, \ldots, a_n] \textbf{ is } p \quad (2.7)$$

where φ is a formula of fluents and unknown variables, a_i's are actions, and $p \in [0, 1]$.

Semantics – Entailment of Queries in PAL_Q: The entailment is defined in several steps. First the transitional probability between states due to a single action is defined as follows.

$$P_{[a]}(\mathsf{s}'|\mathsf{s}) = P_a(\mathsf{s}'|\mathsf{s}) = \begin{cases} \frac{2^{|U_N|}}{|\Phi(a,\mathsf{s})|} P(\mathsf{s}'_N) \text{ if } \mathsf{s}' \in \Phi(a, \mathsf{s}) \\ 0, \text{ otherwise.} \end{cases}$$

The (probabilistic) correctness of a single action plan given that we are in a particular state s is defined as follows.

$$P(\varphi \textbf{ after } a|\mathsf{s}) = \sum_{\mathsf{s}' \in \Phi(a,\mathsf{s}) \wedge \mathsf{s}' \models \varphi} P_a(\mathsf{s}'|\mathsf{s})$$

Next the transitional probability due to a sequence of actions, is recursively defined starting with the base case.

$$P_{[\,]}(\mathsf{s}'|\mathsf{s}) = 1 \text{ if } \mathsf{s} = \mathsf{s}';\ \text{otherwise it is } 0.$$
$$P_{[a_1,\ldots a_n]}(\mathsf{s}'|\mathsf{s}) = \sum_{\mathsf{s}''} P_{[a_1,\ldots,a_{n-1}]}(\mathsf{s}''|\mathsf{s}) P_{a_n}(\mathsf{s}'|\mathsf{s}'')$$

Finally, the (probabilistic) correctness of a (multi-action) plan given that we are in a particular state s is defined as follows.

$$P(\varphi \textbf{ after } \alpha|\mathsf{s}) = \sum_{\mathsf{s}' \in \hat{\Phi}([\alpha],\mathsf{s}) \wedge \mathsf{s}' \models \varphi} P_{[\alpha]}(\mathsf{s}'|\mathsf{s}) \quad (2.8)$$

2.4 PAL_O: The observation language

An observation description $\mathcal{O}$ is a collection of proposition of the following form: ψ **obs_after** $a_1, \ldots, a_n$,
where ψ is a fluent formula, and a_i's are actions. When, $n = 0$, it is simply written as **initially** ψ. The probability $P(\varphi$ **obs_after** $\alpha|\mathsf{s})$ is computed by the right hand side of (2.8).

Semantics – assimilating observations in PAL_O: Using the Bayes' rule, the conditional probability of a state given some observations is given as follows.

$$P(\mathsf{s}_i|\mathcal{O}) = \begin{cases} \frac{P(\mathcal{O}|\mathsf{s}_i)P(\mathsf{s}_i)}{\sum_{\mathsf{s}_j} P(\mathcal{O}|\mathsf{s}_j)P(\mathsf{s}_j)} \text{ if } \sum_{\mathsf{s}_j} P(\mathcal{O}|\mathsf{s}_j)P(\mathsf{s}_j) \neq 0 \\ 0, \text{ otherwise .} \end{cases}$$

2.5 Queries with observation assimilation

The (probabilistic) correctness of a (multi-action) plan given only some observations is defined by:

$$P(\varphi \textbf{ after } \alpha|\mathcal{O}) = \sum_{\mathsf{s}} P(\mathsf{s}|\mathcal{O}) \times P(\varphi \textbf{ after } \alpha|\mathsf{s}) \quad (2.9)$$

A PAL action theory consists of a domain description, a probability description of the unknown variables, and an observation description. Using the above formula, the entailment from an action theory to queries is defined as follows:

Definition 3 $\mathcal{D} \cup \mathcal{P} \cup \mathcal{O} \models_A$ **probability of** $[\varphi$ **after** $a_1, \ldots, a_n]$ **is** p iff $P(\varphi$ **after** $a_1, \ldots, a_n|\mathcal{O}) = p$ □

(Often we write the entailment in the shorter form as $\mathcal{D} \cup \mathcal{P} \cup \mathcal{O} \models_A P(\varphi$ **after** $a_1, \ldots, a_n|\mathcal{O}) = p$).

3 Probabilistic causal models (PCMs)

Definition 4 [Pearl, 2000]

A *causal model* is a triple $M = \langle U, V, F\rangle$ where:

(i) U is a set of *background* variables, (also called *exogenous*), that are determined by factors outside the model.

(ii) V is a set $\{V_1, V_2, \ldots V_n\}$ of variables, called *endogenous*, that are determined by variables in the model - that is, variables in $U \cup V$; and

(iii) F is a set of functions $\{f_1, f_2, \ldots f_n\}$ such that each f_i is a mapping from (the respective domains of) $U \cup (V \setminus V_i)$ to V_i and such that the entire set F forms a mapping from U to V. In other words, each f_i tells us the value of V_i given the values of all other variables in $U \cup V$, and the entire set F has a unique solution $V(u)$. Symbolically, the set of equations F can be represented by writing

$$v_i = f_i(pa_i, u_i)\ i = 1, \ldots, n$$

where pa_i is any realization of the unique minimal set of variables PA_i in $V \setminus V_i$ (connoting *parents*) sufficient for representing f_i. Likewise, $U_i \subseteq U$ stands for the unique minimal set of variables in U sufficient for representing f_i.

Definition 5 [Pearl, 2000]

Let M be a causal model, X be a set of variables in V, and x be a particular realization of X. A *submodel* M_x of M is the causal model $M_x = \langle U, V, F_x\rangle$ where $F_x = \{f_i : V_i \notin X\} \cup \{X = x\}$. □

Submodels are useful for representing the effect of local actions and hypothetical changes. M_x represents the model that results from a minimal change to make $X = x$ hold true under any u.

Definition 6 [Pearl, 2000]

A *probabilistic causal* model (PCM) is a pair $\langle M, P(u)\rangle$ where M is a causal model and $P(u)$ is a probability function defined over the domain of U. □

Because the set of functional equations forms a mapping from U to V, the probability distribution $P(u)$ also determines a probability distribution over the endogenous variables. Hence, given any subsets X and E of $U \cup V$, the conditional probability $P(X = x|E = e)$ is well-defined by $\langle M, P(u)\rangle$.

Definition 7 *(Probabilistic queries and their entailment in PCM)*

(i) Given a probabilistic causal model $\mathcal{M} = \langle M, P(u)\rangle$, the probability of x given an observation e is the conditional probability $P(x|e)$. If $P(x|e) = p$, we write $\mathcal{M} \models_C P(x|e) = p$.

(ii) Given a probabilistic causal model $\mathcal{M} = \langle M, P(u)\rangle$, the probability of x given an intervention $do(y)$, denoted by $P(x|do(y))$, is the probability of x computed w.r.t the submodel $\mathcal{M} = \langle M_y, P(u)\rangle$. If $P(x|do(y)) = p$, we write $\mathcal{M} \models_C P(x|do(y)) = p$.

(iii) Given a probabilistic causal model $\mathcal{M} = \langle M, P(u)\rangle$, the probability of x given observation e and intervention $do(y)$, denoted by $P(x|e, do(y))$, is the probability $P(x|do(y))$ that is computed w.r.t the modified causal model $\mathcal{M}' = \langle M_y, P(u|e)\rangle$, where $P(u|e)$ is computed w.r.t the model $\mathcal{M}$. If $P(x|e, do(y)) = p$ we write $\mathcal{M} \models_C P(x|e, do(y)) = p$. □

From the above definition it follows that $\langle M, P(u)\rangle \models_C P(x|do(y)) = p$ iff $\langle M_y, P(u)\rangle \models_C P(x) = p$; and $\langle M, P(u)\rangle \models_C P(x|e, do(y))$ iff $\langle M_y, P(u|e)\rangle \models_C P(x) = p$.

4 Encoding PCM by PAL action theories

In this section we give an encoding of PCM in PAL, illustrate the encoding with an example, and show the correspondence between query entailment in PCM and query entailment of the corresponding encoding in PAL.

Definition 8 Given a PCM $\mathcal{M} = \langle M, P(u)\rangle$ and assuming that the functions $f_i(pa_i, U_i)$ in M are logical functions, we construct a PAL action theory $D(\mathcal{M})$ as follows:

- There are no non-inertial unknown variables.
- The inertial unknown variables are the exogenous variables in M with the same probability distributions:

  probability of u **is** $P(u)$, for every unknown variable u.

- The endogenous variables in M are fluents in $D(\mathcal{M})$. Moreover, for every fluent v_i, there is an additional fluent $ab(v_i)$ in $D(\mathcal{M})$.
- For each functional equation of the form $v_i = f_i(pa_i, U_i)$ in M the following static causal rule is in $D(\mathcal{M})$: $\neg ab(v_i)$ **causes** $v_i \Leftrightarrow f_i(pa_i, U_i)$.
- For every fluent v_i, $D(\mathcal{M})$ has actions '$make(v_i)$', '$make(\neg v_i)$' with the following effects:

  $make(v_i)$ **causes** $\{ab(v_i), v_i\}$
  $make(\neg v_i)$ **causes** $\{ab(v_i), \neg v_i\}$. □

We now consider the *firing squad* example from [Pearl, 1999], and show its encoding in PAL. (In the following sections, we denote the indicator function by χ, that is, $\chi(X) = 1$ if X is true and $\chi(X) = 0$ if X is false.)

Example 1 The probabilistic causal model of the firing squad example has two exogenous variables U, and W and endogenous variables A, B, C, and D. These variables stand for the following propositions:

$$\begin{array}{rcl} U & = & \text{court orders the execution;} \\ C & = & \text{captain gives a signal;} \\ A & = & \text{rifle A shoots;} \\ B & = & \text{rifle B shoots;} \\ D & = & \text{the prisoner dies;} \\ W & = & \text{rifle A pulls the trigger out of nervousness.} \end{array}$$

The causal relationships between the variables are described by the following functional equations:
$C = U; \quad A = C \vee W; \quad B = C; \quad D = A \vee B.$
In [Pearl, 1999] the goal is to compute the probability $P(\neg D|D, do(\neg A))$, which expresses the probability that the prisoner would be alive if A had not shot, given that the prisoner is in fact dead. There it is shown that:

$$P(u, w|D) = \begin{cases} \frac{P(u,w)}{1-(1-p)(1-q)} & \text{if } u = U \text{ or } w = W, \\ 0 & \text{if } u = \neg U \text{ and } w = \neg W. \end{cases}$$
$$P(\neg D|D, do(\neg A)) = \frac{q(1-p)}{1-(1-p)(1-q)} \tag{4.10}$$

We now construct PAL encoding of the above. The action theory contains inertial unknown variables U and W, fluents A, B, C, D, $ab(A), ab(B), ab(C)$, and $ab(D)$. Translated into PAL, the functional equations become the following static causal laws:

$$\begin{array}{lll} \neg ab(C) & \textbf{causes} & C \Leftrightarrow U \\ \neg ab(A) & \textbf{causes} & A \Leftrightarrow C \vee W \\ \neg ab(B) & \textbf{causes} & B \Leftrightarrow C \\ \neg ab(D) & \textbf{causes} & D \Leftrightarrow A \vee B \end{array} \tag{4.11}$$

We now relate the probabilities computed with respect to the PCM of the above example with its PAL encoding.

Proposition 1 Let $\mathcal{M} = \langle M, P(u)\rangle$ be the PCM of firing squad example given above, and let us denote its encoding in PAL by $D(\mathcal{M})$. Let $init_{\neg ab} = \{$**initially** $\neg ab(A) \wedge \neg ab(B) \wedge \neg ab(C) \wedge \neg ab(D)\}$. For any u and w literals of U and W:

$P($**initially** $\{u, w\}|init_{\neg ab},$ **initially** $D)$

$$= \begin{cases} \frac{P(u,w)}{1-(1-p)(1-q)} & \text{if } u = U \text{ or } w = W, \\ 0 & \text{if } u = \neg U \text{ and } w = \neg W. \end{cases} \tag{4.12}$$

$P(\neg D$ **after** $make(\neg A)|init_{\neg ab},$ **initially** $D)$

$$= \frac{q(1-p)}{1-(1-p)(1-q)} \tag{4.13}$$

Proof (sketch):
We first evaluate $P($**initially** $\{u, w\}|init_{\neg ab},$ **initially** $D)$. For that we use the Bayes' rule:

$P($**initially** $\{u, w\}|init_{\neg ab},$ **initially** $D)$

$$= \frac{P(\textbf{initially } \{u, w\}, \textbf{initially } D|init_{\neg ab})}{P(\textbf{initially } D|init_{\neg ab})} \tag{4.14}$$

Because of the static causal laws (4.11), given $init_{\neg ab}$, the variables U, W and the fluents A, B, C, D in the initial state satisfy:

$$\begin{array}{lcllcl} C & \Leftrightarrow & U, & A & \Leftrightarrow & C \vee W, \\ B & \Leftrightarrow & C, \text{ and} & D & \Leftrightarrow & A \vee B \end{array} \tag{4.15}$$

It follows from (4.15) that $\neg D \Leftrightarrow \neg U \wedge \neg W$ (in the initial state). Hence, $P($**initially** $\neg D|init_{\neg ab})$

$$\begin{array}{l} = P(\textbf{initially } \neg U, \textbf{initially } \neg W|init_{\neg ab}) \\ = P(U = 0, W = 0) = (1-p)(1-q). \end{array} \tag{4.16}$$

Therefore, we have: $P($**initially** $D|init_{\neg ab})$

$$\begin{array}{l} = 1 - P(\textbf{initially } \neg D|init_{\neg ab}) \\ = 1 - (1-p)(1-q). \end{array} \tag{4.17}$$

Because $\neg D \Leftrightarrow \neg U \wedge \neg W$ in the initial state, we also have that: **initially** $D \Leftrightarrow$ **initially** $U \vee$ **initially** W. Consequently, if $u = U$ or $w = W$ then: $P($**initially** $\{u, w\},$ **initially** $D|init_{\neg ab})$

$$= P(\textbf{initially } \{u, w\}|init_{\neg ab}) \quad = P(u, w).$$

Otherwise, if $u = \neg U$ and $w = \neg W$ then:

$$P(\textbf{initially } \{u, w\}, \textbf{initially } D|init_{\neg ab}) = 0.$$

Finally, we have that:

$P($**initially** $\{u, w\},$ **initially** $D|init_{\neg ab})$

$$= \begin{cases} P(u, w) & \text{if } u = U \text{ or } w = W \\ 0 & \text{if } u = \neg U \text{ and } w = \neg W. \end{cases} \tag{4.18}$$

It is easy to see that (4.12) follows from (4.14), (4.17) and (4.18).

For proving (4.13), we use the formula: $P(\neg D$ **after** $make(\neg A)|init_{\neg ab},$ **initially** $D) =$

$$\sum_{\mathsf{s}} P(\neg D \textbf{ after } make(\neg A)|\mathsf{s}) P(\mathsf{s}|init_{\neg ab}, \textbf{initially } D) \tag{4.19}$$

Observe that $P(\mathsf{s}|init_{\neg ab},$ **initially** $D) = 0$ if $init_{\neg ab}$ does not hold in s (that is, $\mathsf{s} \not\models init_{\neg ab}$). Hence, the right hand side depends only on the terms containing s such that $\mathsf{s} \models init_{\neg ab}$. In the following we will consider only s such that $\mathsf{s} \models init_{\neg ab}$. Let us assume that u and w are the literals of U and W that hold in the initial state s. The variables and fluents in the initial state s satisfy the functions in (4.15). Consequently, the variables and fluents in s are uniquely determined by the values of U and W, that is, by u and w. So s is also uniquely determined by u and w. Thus we have:

$P(\mathsf{s}|init_{\neg ab},$ **initially** $D) =$

$$P(\textbf{initially } \{u, w\}|init_{\neg ab}, \textbf{initially } D). \tag{4.20}$$

Assume that we reach the state s' by executing $make(\neg A)$ in the initial state s. The action causes $ab(A)$ and $\neg A$ to be true. Then it follows from the static causal laws (4.11) that in the

state s': $C \Leftrightarrow U$, $B \Leftrightarrow C$ and $D \Leftrightarrow B$. Therefore, in the state s', we have that $D \Leftrightarrow U$. Because U is an inertial unknown variable, its values are the same in s and s'. Consequently,

$$\begin{array}{l}\neg D \textbf{ after } make(\neg A) \Leftrightarrow \neg D \in \mathsf{s}' \Leftrightarrow \neg U \in \mathsf{s}' \\ \Leftrightarrow \neg U \in \mathsf{s} \Leftrightarrow \textbf{initially } \neg U\end{array} \quad (4.21)$$

Since s uniquely depends on u, w:

$$P(\textbf{initially } \neg U|\mathsf{s}) = P(\textbf{initially } \neg U|u, w)$$

Thus we have: $P(\neg D \textbf{ after } make(\neg A)|\mathsf{s})$

$$= P(\textbf{initially } \neg U|\mathsf{s}) = \chi(u = \neg U) \quad (4.22)$$

From (4.19), (4.20) and (4.22), we have that:

$$\begin{array}{l}P(\neg D \textbf{ after } make(\neg A)|init_{\neg ab}, \textbf{initially } D) \\ = \sum_{u,w} \chi(u = \neg U) P(\textbf{initially } \{u, w\}|init_{\neg ab}, \textbf{initially } D)\end{array}$$

We know that $\chi(u = \neg U) \neq 0$ only if $u = \neg U$. Furthermore, because of (4.12), if $u = \neg U$ then $P(\textbf{initially } \{u, w\}|init_{\neg ab}, \textbf{initially } D) \neq 0$ only if $w = W$. So the only possible positive term in the above sum corresponds to the pair $u = \neg U, w = W$. Then:

$$\begin{array}{l}P(\neg D \textbf{ after } make(\neg A)|init_{\neg ab}, \textbf{initially } D) \\ = P(\textbf{initially } \{\neg U, W\}|init_{\neg ab}, \textbf{initially } D) \\ = \frac{P(\neg U, W)}{1-(1-p)(1-q)} = \frac{q(1-p)}{1-(1-p)(1-q)}.\end{array}$$

5 Relating PCM and its encoding in PAL: the main result

We now generalize Proposition 1 to all PCMs and their encoding in PAL. The proof of Proposition 1 will now serve as a road map to the proof of the following general result.

Theorem 5.1 *Given a probabilistic causal model $\mathcal{M} = \langle M, P(u)\rangle$, let $D(\mathcal{M})$ be its respectively constructed PAL action theory. Let $init_{\neg ab} = \{\textbf{initially } \neg ab(v)|v \in V\}$. Let u be a subset of background variable, v and w be subsets of endogenous variables. Then we have the following relations between entailments in PCM and PAL:*

- $\mathcal{M} \models_C P(u|w) = p$ if and only if $D(\mathcal{M}) \models_A$
$$P(\textbf{initially } u|init_{\neg ab}, \textbf{initially } w) = p \quad (5.23)$$
- $\mathcal{M} \models_C P(w|do(v)) = p$ if and only if $D(\mathcal{M}) \models_A$
$$P(w \textbf{ after } do(v)|init_{\neg ab}) = p \quad (5.24)$$
- $\mathcal{M} \models_C P(\neg w|w, do(\neg v)) = p$ iff $D(\mathcal{M}) \models_A$
$$P(\neg w \textbf{ after } make(\neg v)\ |init_{\neg ab}, \textbf{initially } w) = p \quad (5.25)$$

Note that, since the action theory $D(\mathcal{M})$ does not have non-inertial unknown variables $\mathsf{s} = \mathsf{s}_I \cup \mathsf{s}_F$, for every state s. We will need some lemmas for the main proof. These lemmas are stated in the context of the causal model $\mathcal{M}$ and the action theory $D(\mathcal{M})$ given in Theorem 5.1.

Lemma 5.2 *Let u be a subset of exogenous variables and w be a subset of endogenous variables.*

$$P(u|w) = P(\textbf{initially } u|init_{\neg ab}, \textbf{initially } w) \quad (5.26)$$

Proof: Using the Bayes' rule we have:

$$P(\textbf{initially } u|init_{\neg ab}, \textbf{initially } w) = \frac{P(\textbf{initially } u, \textbf{initially } w|init_{\neg ab})}{P(\textbf{initially } w|init_{\neg ab})} \quad (5.27)$$

Because $P(u|w) = \frac{P(u,w)}{P(w)}$, the proof is completed once we prove that:

$$P(u, w) = P(\textbf{initially } u, \textbf{initially } w|init_{\neg ab}) \quad (5.28)$$
$$P(w) = P(\textbf{initially } w|init_{\neg ab}) \quad (5.29)$$

First let us prove (5.28). (5.29) can be shown in a similar manner. Let $u_1, \ldots u_n$ be the inertial unknown variables. We will use $u_{1:n}$ as the shorthand for $\{u_1, \ldots u_n\}$. We know that:

$$\begin{array}{l}P(\textbf{initially } u, \textbf{initially } w|init_{\neg ab}) = \\ \sum_{u_{1:n}} P(\textbf{initially } u, \textbf{initially } w|init_{\neg ab}, \textbf{initially } u_{1:n}) \times \\ \quad P(\textbf{initially } u_{1:n}|init_{\neg ab}).\end{array} \quad (5.30)$$

Since the unknown variables are independent from the fluents, **initially** $u_{1:n}$ is independent from $init_{\neg ab}$. Then we have that:

$$P(\textbf{initially } u_{1:n}|init_{\neg ab}) = P(\textbf{initially } u_{1:n}) = P(u_{1:n}). \quad (5.31)$$

When $init_{\neg ab}$ is true, by the static causal laws, the variables and fluents satisfy the same set of equations as that of the functional equations. Because the set of functional equations forms a mapping from U to V, w is uniquely determined by $u_1, \ldots, u_n$. Therefore, $P(\textbf{initially } u, \textbf{initially } w|init_{\neg ab}, \textbf{initially } u_{1:n})$

$$= P(u, w|u_{1:n}). \quad (5.32)$$

Then (5.28) follows from (5.30), (5.31) and (5.32), because:

$$\begin{array}{l}P(\textbf{initially } u, \textbf{initially } w|init_{\neg ab}) \\ = \sum_{u_1,\ldots,u_n} P(u, w|u_1, \ldots, u_n) P(u_1, \ldots, u_n) \\ = P(u, w).\end{array}$$

We can similarly prove (5.29).

Lemma 5.3 *Let v and w be subsets of endogenous variables. If s is an initial state such that $\mathsf{s} \models init_{\neg ab}$ then*

(i) *s is uniquely determined by its inertial unknown variable subset s_I,*

(ii) *probability of an observation depends only on the unknown variables: $P(\neg w \textbf{ after } make(\neg v)|\mathsf{s}) =$*
$$P(\neg w \textbf{ after } make(\neg v)|init_{\neg ab}, \textbf{initially } \mathsf{s}_I) \quad (5.33)$$

(iii) *given an evidence, probability of s depends only on the unknown variables:*
$$P(\mathsf{s}|init_{\neg ab}, \textbf{initially } w) = P(\mathsf{s}_I|w, do(\neg v)). \quad (5.34)$$

Proof: The proofs of (i) and (ii) are straightforward.

(iii) When $\mathsf{s} \models init_{\neg ab}$, the unknown variables determine what state we are in. That is, conditioning on $init_{\neg ab}$, the initial state s is determined by s_I . Consequently,

$$\begin{array}{l}P(\mathsf{s}|init_{\neg ab}, \textbf{initially } w) \\ = P(\textbf{initially } \mathsf{s}_I|init_{\neg ab}, \textbf{initially } w).\end{array} \quad (5.35)$$

Moreover, since the unknown variables are independent of each other:

$$P(\textbf{initially } \mathsf{s}_I | init_{\neg ab}, \textbf{initially } w) = \prod_u P(\textbf{initially } u | init_{\neg ab}, \textbf{initially } w). \quad (5.36)$$

By (5.26), $P(\textbf{initially } u | init_{\neg ab}, \textbf{initially } w) = P(u|w)$. Because the intervention $do(\neg v)$ does not effect the probability distributions of the exogenous variables: $P(u|w) = P(u|w, do(\neg v))$. Therefore,

$$P(\textbf{initially } \mathsf{s}_I | init_{\neg ab}, \textbf{initially } w) = \prod_u P(u|w, do(\neg v)) = P(\mathsf{s}_I | w, do(\neg v)). \quad (5.37)$$

Lemma 5.4 *Let $u_1, \ldots, u_n$ be the exogenous variables in the causal model. Let U be some realization of these variables. Let v and w be subsets of the endogenous variables. Assume that* s *is an initial state such that $s \models init_{\neg ab} \cup \{$***initially** $w\}$ *and $s_I = U$. Then*

$$P(\neg w \textbf{ after } make(\neg v) | \mathsf{s}) = P(\neg w | w, do(\neg v), U). \quad (5.38)$$

Proof: Let $\mathcal{F}$ be the set of the functional equations in the causal model $\mathcal{M}$ and $\mathcal{F}'$ be the set of functional equations in the modified submodel $\mathcal{M}_{\neg v}$.

Because the sets $\mathcal{F}$ and $\mathcal{F}'$ form mappings from U to V, given the realization U of the background variables, there exist unique realizations V and V' of the endogenous variables, such that $U \cup V$ is the (unique) solution of $\mathcal{F}$ and $U \cup V'$ is the (unique) solution of $\mathcal{F}'$.

Let $A = \{\neg ab(v_i) | v_i \text{ is an endogenous variable } \}$. Then $A \subseteq \mathsf{s}$, because $\mathsf{s} \models init_{\neg ab}$. Because of the static causal laws, the literals (of the unknown variables and the fluents) in the state s should satisfy $\mathcal{F}$. Since $\mathsf{s}_I = U$, it follows that $\mathsf{s} = A \cup U \cup V$. Moreover, since $\mathsf{s} \models$ **initially** w we have $U \cup V \Rightarrow w$. Now, $P(U|w, do(\neg v)) = P(U|w) \neq 0$. Therefore,

$$P(\neg w | w, do(\neg v), U) = \frac{P(\neg w, U | w, do(\neg v))}{P(U|w, do(\neg v))} = \frac{\chi(U \cup V' \Rightarrow \neg w) P(U|w, do(\neg v))}{P(U|w, do(\neg v))} = \chi(U \cup V' \Rightarrow \neg w).$$

It can be shown that there exists a unique state s' such that:

$$\Phi(make(\neg v), \mathsf{s}) = \{\mathsf{s}'\}. \quad (5.39)$$

It follows directly from (5.39) that $P(\neg w \textbf{ after } make(\neg v) | \mathsf{s}) = \chi(\mathsf{s}' \models \neg w)$. Finally, the proof will be completed by showing that

$$\chi(U \cup V' \Rightarrow \neg w) = \chi(\mathsf{s}' \models \neg w). \quad (5.40)$$

Proof of Theorem 5.1. (5.23) is proved in Lemma 5.2. To prove (5.25), we apply formula (2.9) and the lemmas:

$$p = P(\neg w \textbf{ after } make(\neg v) | init_{\neg ab}, \textbf{initially } w) = \sum_{\mathsf{s}} P(\neg w \textbf{ after } make(\neg v) | \mathsf{s}) P(\mathsf{s} | init_{\neg ab}, \textbf{initially } w)$$

Because $P(\mathsf{s}|init_{\neg ab}, \textbf{initially } w) = 0$ if $\mathsf{s} \not\models init_{\neg ab}$ or $\mathsf{s} \not\models$ **initially** w we have:

$$p = \sum_{\substack{\mathsf{s} \models init_{\neg ab} \\ \mathsf{s} \models \textbf{initially } w}} P(\neg w \textbf{ after } make(\neg v) | \mathsf{s}) P(\mathsf{s} | init_{\neg ab}, \textbf{initially } w)$$

Using Lemma 5.4 and (5.34) we have:

$$p = \sum_{\substack{\mathsf{s} \models init_{\neg ab} \\ \mathsf{s} \models \textbf{initially } w}} P(\neg w | w, do(\neg v), s_I) P(\mathsf{s}_I | w, do(\neg v))$$

Because s_I has range 2^U if $\mathsf{s} \models init_{\neg ab}$ we have:

$$p = \sum_{U: U \Rightarrow w} P(\neg w | w, do(\neg v), U) P(U | w, do(\neg v)) = P(\neg w | w, do(\neg v)). \quad \square$$

6 Conclusion and future work

In this paper we have shown how to encode reasoning in probabilistic causal models in the action description language PAL. The main observation is that functional equations of the form $v_i = f_i(pa_i, U_i)$ need to be encoded as $\neg ab(v_i)$ **causes** $v_i = f_i(pa_i, U_i)$ instead of the straightforward encoding $true$ **causes** $v_i = f_i(pa_i, U_i)$. This is because an action that directly changes the value of v_i makes the equation $v_i = f_i(pa_i, U_i)$ unusable. This is achieved in the PAL encoding by making $ab(v_i)$ true.

One important aspect of the PAL encoding is that as the world progresses if we want to reactivate a previously inactivated functional equation $v_i = f_i(pa_i, U_i)$ all we need to do is make $ab(v_i)$ false. (Reactivation need to be done with care though.) In probabilistic causal models once a functional equation is inactivated it can no longer be activated. (Note that the differences between PAL and probabilistic causal models are discussed in [Baral *et al.*, 2002] and we do not repeat them here as the purpose here is to relate them.)

In the definition of causal models, the set of functional equation F is assumed to be a mapping from U to V. While it is not clear how to do counterfactuals in the framework of [Pearl, 2000] without the assumption, it seems straightforward to do so in the PAL framework. We will further investigate this difference. We also plan to test PAL formalism in real world applications and to develop algorithms for learning PAL theories.

References

[Baral *et al.*, 2002] C. Baral, N. Tran, and L. Tuan. Reasoning about actions in a probabilistic setting. In *Proc. of AAAI'2002*, pages 507–512, 2002.

[Boutilier and Goldszmidt, 1996] C. Boutilier and M. Goldszmidt. The frame problem and bayesian network action representations. In *Proc. of CSCSI-96*, May 1996.

[Gelfond and Lifschitz, 1993] M. Gelfond and V. Lifschitz. Representing actions and change by logic programs. *Journal of Logic Programming*, 17(2,3,4):301–323, 1993.

[Gelfond and Lifschitz, 1998] M. Gelfond and V. Lifschitz. Action languages. *Electronic Transactions on AI*, 3(16), 1998.

[Littman, 1997] M. Littman. Probabilistic propositional planning: representations and complexity. In *AAAI 97*, pages 748–754, 1997.

[McCain and Turner, 1995] N. McCain and H. Turner. A causal theory of ramifications and qualifications. In *Proc. of IJCAI 95*, pages 1978–1984, 1995.

[Pearl, 1995] J. Pearl. Action as a local surgery. In C. Boutilier and M. Goldszmidt, editors, *Proc. of 1995 AAAI symposium on Extending theories of action: formal theory and practical applications*, pages 157–162, 1995.

[Pearl, 1999] J. Pearl. Reasoning with cause and effect. In *IJCAI 99*, pages 1437–1449, 1999.

[Pearl, 2000] J. Pearl. *Causality*. Cambridge University Press, 2000.

[Reiter, 2001] R. Reiter. *Knowledge in action: logical foundation for describing and implementing dynamical systems*. MIT press, 2001.