TPLP **23** (4): 715–729, 2023.
doi:10.1017/S1471068423000248 First published online 26 July 2023

Embracing Background Knowledge in the Analysis of Actual Causality: An Answer Set Programming Approach

MICHAEL GELFOND
Department of Computer Science, Texas Tech University, Lubbock, TX, USA
(*e-mail:* michael.gelfond@ttu.edu)

JORGE FANDINNO
Department of Computer Science, University of Nebraska at Omaha, Omaha, NE, USA
(*e-mail:* jfandinno@unomaha.edu)

EVGENII BALAI
Department of Computer Science, Texas Tech University, Lubbock, TX, USA
(*e-mail:* evgenii.balai@gmail.com)

submitted 3 June 2023; revised 18 June 2023; accepted 19 June 2023

Abstract

This paper presents a rich knowledge representation language aimed at formalizing causal knowledge. This language is used for accurately and directly formalizing common benchmark examples from the literature of actual causality. A definition of cause is presented and used to analyze the actual causes of changes with respect to sequences of actions representing those examples.

KEYWORDS: answer set programming, causality, knowledge representation

1 Introduction

This work is a part of larger research program, originated by John McCarthy and others in the late fifties. The program is aimed at the development of Knowledge Representation (KR) languages capable of clear and succinct formalization of commonsense knowledge. In this paper we concentrate on a long standing problem of giving a formal account of the notion of actual causality. Despite significant amount of work in this area the problem remains unsolved. We believe that the difficulty is related to insufficient attentions paid to relevant commonsense background knowledge. To analyze causal relations involved in a sequence of events happening even in comparatively simple domains, we need to be able to represent sophisticated causal laws, time, defaults and their exceptions, recursive definitions, and other non-trivial phenomena of natural language. To the best of our knowledge none of the KR-languages used in previous works are capable of representing all of these phenomena. We propose to remedy this problem by analyzing causality in the context of a new rich KR-language $\mathcal{W}$ based on the ideas from Answer Set Prolog (ASP), Theories of Action and Change (TAC), and Pearl's do-operator Pearl (2009). The language is used to define several causal relations capable

of accurate analysis of a number of examples which could not have been properly analyzed by the previous approaches. Special emphasis in our approach is given to accuracy and *elaboration tolerance* McCarthy (1998) of translations of English texts into theories of $\mathcal{W}$. This is facilitated by the well developed methodology of such translations in ASP and TAC. These issues were not typically addressed in work on causality, but they are essential from the standpoint of KR. We focus on the suitability of $\mathcal{W}$ for causal analysis, illustrated by its application to well-known benchmarks from the literature. The paper is organized as follows. In the next section we motivate the need for a richer KR-language by analyzing such benchmarks. After that, we introduce *causal theories* of $\mathcal{W}$ and a methodology for formalizing natural language stories. This is illustrated on these benchmarks. Special care is taken of obtaining accurate and direct translation from natural language sentences, and the elaboration tolerance of the representation. The later is obtained by a clear separation between background commonsense knowledge (formalized in a *background theory*) and the particular story (formalized as a sequence of events that we call *scenario*). Finally, we introduce our definition of cause and discuss several variations of the benchmark examples. This definition provides answers that match our intuition. Note that, since $\mathcal{W}$ is a powerful action language based on ASP it can also be used for reasoning about temporal prediction, planning, etc. Due to space limitation the paper does not demonstrate the full power of $\mathcal{W}$ and the full variety of its causal relations. This will be done in a longer version of the paper.

2 Motivating examples

In this section, we discuss two problematic benchmarks from the literature and provide their causal description based on KR perspective. We start by considering the *Suzy First* example Hall (2004) and extensively discussed in the literature. The following reading is by Halpern and Pearl Halpern and Pearl (2001).

Example 1
Suzy and Billy both pick up rocks and throw them at a bottle. Suzy's rock gets there first, shattering the bottle. *Since both throws are perfectly accurate, Billy's would have shattered the bottle had it not been preempted by Suzy's throw. Common sense suggests that Suzy's throw is the cause of the shattering, but Billy's is not.*

Time and actions are essential features of this example. The reasoning leading to Suzy's throw being regarded as the cause of the bottle directly points to the sentence "Suzy's rock gets there first, shattering the bottle." Had Billy's throw got there first, we would have concluded that Billy's throw was the cause. Despite the importance of time in this example, most approaches do not explicitly represent time. As a result, the fact that "Suzy's rock gets there first," which naturally is part of the particular scenario, is represented as part of background knowledge Halpern and Pearl (2001); Hopkins and Pearl (2003); Chockler and Halpern (2004); Hall (2004; 2007); Vennekens (2011); Halpern (2014; 2015); Halpern and Hitchcock (2010); Beckers and Vennekens (2016; 2018); Bochman (2018); Denecker *et al.* (2019); Beckers (2021). This means that a small change in the scenario such as replacing "Suzy's rock gets there first, shattering the bottle" by "Billy's rock gets there first, shattering the bottle" or "Suzy's rock gets there first, but her throw was too weak to shatter the bottle" requires a complete change

https://doi.org/10.1017/S1471068423000248 Published online by Cambridge University Press

of the formal model of the domain instead of a small change to the scenario. This is a problem of *elaboration tolerance*. Several approaches addressed the lack of representation of time by introducing features from the area of *Reasoning about Actions and Change*. Approaches in the context of the *Situation Calculus* Hopkins and Pearl (2007); Batusov and Soutchanski (2018) and *Logic Programming* Cabalar *et al.* (2014); Cabalar and Fandinno (2016); Fandinno (2016); LeBlanc *et al.* (2019) allow us to reason about actual causes with respect to different sequences of actions, where the order of these actions matter. For instance, Cabalar and Fandinno (2016) explicitly represent a variation of this example where "Suzy's rock gets there first" is replaced by "Suzy throws first." The model associated with this example is represented by the following rules:

$$broken(I) \leftarrow throw(A, I-1),\ not\ broken(I-1), \tag{1}$$

$$broken(I) \leftarrow broken(I-1),\ not\ \neg broken(I), \tag{2}$$

$$\neg broken(I) \leftarrow \neg broken(I-1),\ not\ broken(I). \tag{3}$$

This can be used together with facts: $\neg broken(0)$, $throw(suzy, 0)$, $throw(billy, 1)$ representing the particular scenario.We can represent an alternative story where "Billy throws first" by replacing the last two facts by $throw(billy, 0)$ and $throw(suzy, 1)$. Clearly, this constitutes a separation between model and scenario because we do not need to modify the rules that represent the model of the domain to accommodate the new scenario. We go a step further and show how to represent the fact that "Suzy's rock gets there first" independently of who throws first. The rock may get there first because Suzy throws first, because she was closer, etc. The reason why her rock gets there first is not stated in the example and it is unnecessary to determine the cause of the shattering. We are able to do that thanks to the introduction of *abstract time-steps* in our language, a feature missing in all previously discussed approaches. As a second example, consider the *Engineer* scenario Hall (2000).

Example 2
An engineer is standing by a switch in the railroad tracks. A train approaches in the distance. She flips the switch, so that the train travels down the right-hand track, instead of the left. Since the tracks reconverge up ahead, the train arrives at its destination all the same; let us further suppose that the time and manner of its arrival are exactly as they would have been, had she not flipped the switch.

It is commonly discussed whether flipping the switch should be (part of) the cause of the train arriving at its destination Halpern and Pearl (2001); Hall (2007); Halpern (2015); Cabalar and Fandinno (2016); Beckers and Vennekens (2017); Batusov and Soutchanski (2018). Normally these solutions are not elaboration tolerant. For instance, adding a neutral switch position or a third route that does not reconverge, requires a different model or leads to completely different answers.

3 Causal theory

This section introduces a simplified version of knowledge representation language $\mathcal{W}$, which is used for the analysis of basic causal relations. Formally, $\mathcal{W}$ is a *subset* of P-log Baral *et al.* (2009); Balai *et al.* (2019) expanded by a simple form of constraints with

https://doi.org/10.1017/S1471068423000248 Published online by Cambridge University Press

signature tailored toward reasoning about change. Theories of $\mathcal{W}$ are called *causal*. A causal theory consists of a *background theory* $\mathcal{T}$ representing *general knowledge* about the agent's domain and *domain scenario* $\mathcal{S}$ containing the record of deliberate actions performed by the agents. A sorted signature Σ of $\mathcal{T}$, referred to as *causal*, consists of sorts, object constants, and function symbols. Each object constant comes together with its sort; each function symbol – with sorts of its parameters and values. In addition to *domain specific sorts* and *predefined sorts* such as *Boolean*, *integer*, etc., a causal signature includes sorts for *time-steps*, *fluents*, *actions*, and *statics*. Fluents are divided into *inertial*, *transient* and *time-independent*. An inertial fluent can only change its value as a result of an action. Otherwise the value remains unchanged. The default value of a transient fluent is *undefined*. A time-independent fluent does not depend on time. But, different from a static, it may change its value after the scenario is expanded by new information. The value sort of actions is Boolean. Terms of Σ are defined as usual. Let e be a function symbol, $\bar{t}$ be a sequence of ground terms not containing time-steps, i be a time-step, and $y \in range(e)$. A ground atom of Σ is an expression of one of the forms

$$e(\bar{t}, i) = y \qquad e(\bar{t}, i) \neq y,^{1}$$

where e is an action, a fluent or a static. If e does not depend on time, i will be omitted.

If e is a Boolean fluent then $e(\bar{t}) = \top$ (resp. $e(\bar{t}) = \bot$) is sometimes written as $e(i)$ (resp. $\neg e(i)$). Atoms formed by actions are called *action atoms*. Similarly for statics, fluents, etc. Action atom $a(i)$ may be written as $occurs(a, i)$.

The main construct used to form background theories of $\mathcal{W}$ is *causal mechanism* (or *causal law*) – a rule of the form:

$$m: \ e(\bar{t}, I) = y \leftarrow body, \neg ab(m, I), \tag{4}$$

where e is a non-static, I ranges over time-steps, m is the unique name of this causal mechanism, *body* is a set of atoms of Σ and arithmetic atoms of the form $N \prec AE$ where N is a variable or a natural number and AE is an arithmetic function built from $+$, $-$, $\times$, etc., and $\prec$ is $=$, $>$, or $\geq$. Special Boolean function $ab(m, I)$ is used to capture exceptions to application of causal mechanism m at step I. As usual in logic programming we view causal mechanisms with variables as sets of their ground instances obtained by replacing variables by their possible values and evaluating the remaining arithmetic terms. If e in rule (4) is an action we refer to m as a *trigger*. A causal mechanism of the form (4) says that "at time-step I, *body* activates causal mechanism m which, unless otherwise specified, sets the value of e to y". To conform to this reading, we need to enforce a broadly shared *principle of causality: "the cause must precede its effect."* Our version of this principle is given by the following requirement: For every ground instance of causal mechanism such that i is a time-step occurring in its head and j is a time-step occurring in its body, the following two conditions are satisfied:

- $j < i$ if j occurs within an action atom; and
- $j \leq i$, otherwise.

[1] As in several existing formalisms Balduccini (2012) and Balai *et al.* (2019), $f(\bar{x}) \neq y$ holds if $f(\bar{x}) = z$ such that $z \neq y$.

https://doi.org/10.1017/S1471068423000248 Published online by Cambridge University Press

$$
\begin{array}{lcll}
m_1 & : & arrived(fork) & \leftarrow \quad approach(I),\ \neg ab(m_1, I) \\
m_2 & : & arrivTime(fork) = I & \leftarrow \quad arrived(fork),\ approach(I - time2fork),\ \neg ab(m_2, I) \\
m_3(P) & : & switch(I) = P & \leftarrow \quad flipTo(P, I-1),\ switch(I-1) \neq P,\ \neg ab(m_3(P), I) \\
m_4 & : & arrived(dest) & \leftarrow \quad arrivTime(fork) = I,\ switch(I) \neq neutral,\ \neg ab(m_4, I) \\
m_5 & : & arrivTime(dest) = I & \leftarrow \quad arrived(dest),\ arrivTime(fork) = I',\ switch(I') = P, \\
 & & & \quad\ I = I' + time2dest(P),\ \neg ab(m_5, I)
\end{array}
$$

Fig. 1. Causal mechanisms in the background theory representing the Engineer story.

A *scenario* of background theory T with signature Σ is a collection of static and arithmetic atoms together with *expressions* of the form:

- $\mathbf{init}(f = y)$ – the initial value of inertial fluent f is y;
- $\mathbf{do}(a, i)$ – an agent deliberately executes action a at time-step i;
- $\mathbf{do}(\neg a, i)$ – an agent deliberately refrains from executing action a at i;
- $\mathbf{obs}(f, y, i)$ – the value of f at time-step i is observed to be y;

We refer to these expressions as *extended atoms of* Σ; a set of extended atoms of the form $\mathbf{init}(f = y)$, $\mathbf{init}(g = z)\ldots$ will be written as $\mathbf{init}(f = y,\ g = z, \ldots)$. We assume that the sort for time-steps consists of all natural numbers and symbolic constants we refer to as *abstract time-steps*. Atoms, extended atoms and scenarios where all object constants of the sort for time-steps are natural numbers are called *concrete*; those that contain abstract time-steps are called *abstract*.

The story of *Suzy First* (Example 1) can be represented in $\mathcal{W}$ by a background theory $\mathcal{T}_{fst}$ which contains a sub-sort *throw* of actions, inertial fluent *broken*, statics *member*, *agent*, and *duration* and causal mechanism

$$
\begin{array}{lcll}
m_0(A): & broken(I) & \leftarrow & occurs(A, I - D), member(A, throw), \\
 & & & agent(A) = Ag, duration(A) = D, \\
 & & & \neg broken(I-1), \neg ab(m_0(A), I).
\end{array}
$$

The theory will be used together with an abstract scenario $\mathcal{S}_{suzy}$ which includes actions a_1 and a_2 of the sort *throw* and atoms

$$\mathbf{init}(\neg broken),\ \mathbf{do}(a_1, t_1),\ \mathbf{do}(a_2, t_2),\ t_1 + duration(a_1) < t_2 + duration(a_2),$$

where t_1 and t_2 are abstract time-steps. The last (arithmetic) atom represents the fact that Suzy's stone arrives first. Actions of $\mathcal{S}_{suzy}$ are described by statics

$$agent(a_1) = suzy \quad member(a_1, throw) \quad agent(a_2) = billy \quad member(a_2, throw), \tag{5}$$

and arithmetic atoms $duration(a_1) \geq 1$, $duration(a_2) \geq 1$. Here, and in other places $f(\bar{t}) \geq y$ is understood as a shorthand for $f(\bar{t}) = d$ and $d \geq y$, where d is a fresh abstract constant. Similarly for $>$, $=$ and $\neq$. To save space, we omit executability conditions for causal mechanisms. Note that causal mechanism $m_0(A)$ is a general commonsense law which is not specific to this particular story. This kind of general commonsense knowledge can be compiled into a background library and retrieved when necessary Inclezan (2016). The same applies to all the other causal mechanisms for variations of this example discussed in the paper. Note that we explicitly represent the temporal relation among

https://doi.org/10.1017/S1471068423000248 Published online by Cambridge University Press

time-steps and *make no further assumptions* about the causal relation among the rocks. The definition of cause introduced below is able to conclude that Suzy's rock is the cause of breaking the bottle. This is a distinguishing feature of our approach. Representing that Billy's stone arrives first is obtained simply by replacing the corresponding constraint in $\mathcal{S}_{suzy}$ by $t_1 + duration(a_1) > t_2 + duration(a_2)$.

The *Engineer* story (Example 2) can be represented by a background theory $\mathcal{T}_{eng}$ containing causal mechanisms in Figure 1. The arrival of the train is modeled by a time-independent fluent $arrived(point)$. Action *approach* of m_2 causes the train to arrive at the fork after the amount of time determined by static *time2fork* (note that since m_1 can fail to cause $arrived(fork)$, this atom cannot be removed from m_2).

The switch is controlled by action *flipTo* which takes one unit of time. This action can change the switch to any of its three positions: *neutral*, *left*, and *right*. Static $time2dest(track)$ determines the time it takes the train to traverse the distance between the fork and the train's destination depending on the *track* taken. When the switch is in the *neutral* position, the train does not arrive at its destination. Inertial fluent *switch* represents the position of the switch.

The times to travel between two points must obey the following constraints included in scenario $\mathcal{S}_{eng}$:

$$time2fork \geq 1 \quad time2dest(left) \geq 1 \quad time2dest(right) \geq 1.$$

The rest of the scenario $\mathcal{S}_{eng}$ consists of the following atoms

$$\mathbf{init}(switch = left) \quad \mathbf{do}(approach, t_3) \quad \mathbf{do}(flipTo(right), t_4),$$
$$time2dest(left) = time2dest(right)$$

We make no assumptions regarding the order in which actions *approach* and *flipTo* occur. We can easily modify the scenario to accommodate a variation of the story where traveling down the right-hand track is faster than over the left one by replacing the last arithmetic atom by $time2dest(left) > time2dest(right)$.

Definition 1 (*Causal Theory*)
A *causal theory* $\mathcal{T}(\mathcal{S})$ is a pair where $\mathcal{T}$ is a background theory and $\mathcal{S}$ is a scenario.

We identify each causal theory $\mathcal{T}(\mathcal{S})$ with the logic program that consists of causal mechanisms without their labels, all atoms in $\mathcal{S}$ as facts and the following general axioms:

$$def(f(\bar{X})) \leftarrow f(\bar{X}) = Y, \tag{6}$$

$$\leftarrow f(\bar{X}) \neq Y,\ not\ def(f(\bar{X})), \tag{7}$$

$$f(\bar{X}) \neq Y \leftarrow f(\bar{X}) = Z,\ Z \neq Y, \tag{8}$$

for every function symbol f,

$$\neg ab(m, I) \leftarrow\ not\ ab(m, I), \tag{9}$$

for every causal mechanism m,

$$f(0) = y \leftarrow \mathbf{init}(f = y), \tag{10}$$

$$f(I) = Y \leftarrow f(I-1) = Y,\ not\ f(I) \neq Y, \tag{11}$$
$$f(I) \neq Y \leftarrow f(I-1) \neq Y,\ not\ f(I) = Y, \tag{12}$$

https://doi.org/10.1017/S1471068423000248 Published online by Cambridge University Press

for every inertial fluent f,

$$a(I) \leftarrow \mathbf{do}(a, I) \qquad \neg a(I) \leftarrow \mathbf{do}(\neg a, I), \tag{13}$$

$$\neg a(I) \leftarrow \ not\ a(I), \tag{14}$$

$$a(I) \stackrel{+}{\leftarrow}, \tag{15}$$

$$ab(m, I) \leftarrow \mathbf{do}(a = v, I), \tag{16}$$

for every action a, Boolean value v and causal law m with head $a(I) = w$ and $v \neq w$.

$$\leftarrow obs(f(\bar{X}), Y, I), not\ f(\bar{X}, I) = Y. \tag{17}$$

Axioms (6), (7), and (8) reflect the reading of relation $\neq$. Axiom (9) ensures that causal mechanisms are defeasible. Axiom (10) ensures that fluents at the initial situation take the value described in the scenario. Axioms (11-12) are the *inertia axioms*, stating that inertial fluents normally keep their values. Axiom (13) ensures that the actions occur as described in the scenario. Axiom (14) states the close world assumption for actions. Axiom (15) is a cr-rule Balduccini and Gelfond (2003); Gelfond and Kahl (2014) which allows indirect exceptions to (14). Intuitively, it says that $a(I)$ *may be true, but such a possibility is very rare and, whenever possible, should be ignored.* Axiom (16) ensures that deliberate actions overrule the default behavior of contradicting causal mechanisms (See Example 3 below for more details). Axiom (17) ensures that observations are satisfied in the model. Note, that if $\mathcal{T}(\mathcal{S})$ contains occurrences of abstract time-steps then its grounding may still have occurrences of arithmetic operations. (If d is an abstract time-step then, say, $d + 1 > 5$ will be unchanged by the grounding). The standard definition of answer set is not applicable in this case. The following modification will be used to define the meaning of programs with abstract time-steps. Let γ be a mapping of abstract time-steps into the natural numbers and $\mathcal{T}(\mathcal{S})$ be a program not containing variables. By $\mathcal{T}(\mathcal{S})|_\gamma$ we denote the result of

(a) applying γ to abstract time-steps from $\mathcal{T}(\mathcal{S})$,
(b) replacing arithmetic expressions by their values,
(c) removing rules containing false arithmetic atoms.

Condition (c) is needed to avoid violation of principle of causality by useless rules. By an *answer set* of $\mathcal{T}(\mathcal{S})$ we mean an answer set of $\mathcal{T}(\mathcal{S})|_\gamma$ for some γ. If $\mathcal{T}(\mathcal{S})|_\gamma$ is consistent, that is, has an answer set then γ is called an *interpretation* of $\mathcal{T}(\mathcal{S})$. $\mathcal{T}(\mathcal{S})$ is called consistent if it has at least one interpretation. If $\mathcal{T}(\mathcal{S})$ is consistent and for each interpretation γ of $\mathcal{T}(\mathcal{S})$, $\mathcal{T}(\mathcal{S})|_\gamma$ has exactly one answer set then $\mathcal{T}(\mathcal{S})$ is called *deterministic. In this paper we limit ourselves to deterministic causal theories.* To illustrate our representation of triggers, parallel actions, and the defeasibility of causal laws, we introduce the following variation of *Suzy First* (Example 1).

Example 3
Suzy and Billy throw rocks by the order of a stronger girl. Suzy's rock gets there first.

The effects of orders are described by the causal mechanism:

$$\begin{array}{lllll} m_6(A,T,B): & occurs(A,I) & \leftarrow & member(B, order), & occurs(B,T), \\ & & & what(B) = A, & when(B) = I, \\ & & & I > T, & \neg ab(m_6(A,T,B), I). \end{array}$$

https://doi.org/10.1017/S1471068423000248 Published online by Cambridge University Press

The scenario $\mathcal{S}_{order}$ is obtained from $\mathcal{S}_{suzy}$ by adding new actions, b_1 and b_2, of the sort *order* described by statics

$$what(b_1) = a_1 \quad when(b_1) = t_1 \quad what(b_2) = a_2 \quad when(b_2) = t_2, \tag{18}$$

and new constraints $t_1 > 0,\ t_2 > 0$, and replacing its extended atoms by

$$\mathbf{init}(\neg broken),\ \mathbf{do}(b_1, 0),\ \mathbf{do}(b_2, 0).$$

For any interpretation γ, in the unique answer set of $\mathcal{T}(\mathcal{S}_{order})|_\gamma$ atom *broken* becomes true at time-step[2] $\gamma(t_1) + \gamma(duration(a_1))$. For the sake of simplicity, we assume that orders are given at time-step 0, but in general we would use two abstract time-steps. The example illustrates representation of triggers and parallel actions. To illustrate defeasibility, let us consider a scenario where both Suzy and Billy refuse to follow the order. This can be formalized as scenario $\mathcal{S}_{order2}$ obtained from $\mathcal{S}_{order}$ by adding the extended atoms $\mathbf{do}(\neg a_1, t_1)$ and $\mathbf{do}(\neg a_2, t_2)$. Due to axioms (16), causal mechanisms $m_6(a_1, 0, b_1)$ and $m_6(a_2, 0, b_2)$ do not fire, and *broken* never becomes true.

4 Cause of change

In this section, we describe our notion of cause of change. We start with scenarios not containing observations.

Definition 2
We say that a ground atom $e(\bar{t}, k) = y$ is a *change* in $\mathcal{T}(\mathcal{S})|_\gamma$ if the unique answer set M of $\mathcal{T}(\mathcal{S})|_\gamma$ satisfies $e(\bar{t}, k) = y$ and one of the following conditions holds:

- e is inertial and $e(\bar{t}, k-1)$ is either undefined in M or M satisfies $e(\bar{t}, k-1) = z$ with $z \neq y$;
- e is an action or a transient or time-independent fluent.

The definition of cause of change relies on the definition of *tight proof* that we introduce next. By $[P]_i$, we denote the sequence consisting of the first i elements of sequence P. By $atoms(P)$ we denote the atoms occurring in P.

Definition 3 (*Proof*)
A *proof* of a set $\mathcal{U}$ of ground atoms in $\mathcal{T}(\mathcal{S})|_\gamma$ is a sequence P of atoms in the unique answer set M of $\mathcal{T}(\mathcal{S})|_\gamma$ and rules of the ground logic program associated with $\mathcal{T}(\mathcal{S})|_\gamma$ satisfying the following conditions:

- P contains all the atoms in $\mathcal{U}$.
- Each element x_i of P is one of the following:
  - — a rule whose body is satisfied by the set
  $$atoms([P]_i) \cup \{\text{not } l : l \notin M\},\ \text{or}$$
  - — an axiom, that is, a **do**-atom or a static from $\mathcal{S}|_\gamma$, or
  - — the head of some rule from $[P]_i$.
- No proper subsequence[3] of P satisfies the above conditions.

[2] $duration(a_1)$ stands for d where $duration(a_1) = d$.
[3] A sequence obtained from P by removing some of its elements.

https://doi.org/10.1017/S1471068423000248 Published online by Cambridge University Press

$\mathbf{do}(approach, \gamma(t_3))$	$\mathbf{do}(approach, \gamma(t_3))$
m_1 at $I = \gamma(t_3)$	m_1 at $I = \gamma(t_3)$
$arrived(fork)$	$arrived(fork)$
m_2 at $I = n_1$	m_2 at $I = n_1$
$arrivTime(fork) = n_1$	$arrivTime(fork) = n_1$
$switch(0) \neq neutral$	$\mathbf{do}(flipTo(right), \gamma(t_4))$
$\ldots$ $(inertia\ rules)$	$m_3(right)$ at $I = \gamma(t_4) + 1$
$switch(n_1) \neq neutral$	$switch(\gamma(t_4) + 1) = right$
m_4 at $I = n_1$	$axiom$ (8)
$arrived(dest)$	$switch(\gamma(t_4) + 1) \neq neutral$
	$\ldots$ $(inertia\ rules)$
	$switch(n_1) \neq neutral$
	m_4 at $I = n_1$
	$arrived(dest)$

Fig. 2. Proofs P_1 and P_2 of $arrived(dest)$ in scenario $\mathcal{S}_{eng}$ with any interpretation γ satisfying condition $\gamma(t_3) + \gamma(time2fork) > \gamma(t_4)$. We use n_1 to denote the positive natural number $\gamma(t_3) + \gamma(time2fork)$. Only P_1 is a tight proof.

Let us consider the *Engineer* story (Example 2) and an interpretation γ of the abstract theory $\mathcal{T}_{eng}(\mathcal{S}_{eng})$, that is, a function mapping $time2fork$, $time2dest(left)$ and $time2dest(right)$ to natural numbers such that

$$\gamma(time2dest(left)) = \gamma(time2dest(right)).$$

For instance, an interpretation γ satisfying $\gamma(t_3) = 0$, $\gamma(t_4) = 1$, $\gamma(time2fork) = 3$ and

$$\gamma(time2dest(left)) = \gamma(time2dest(right)) = 5.$$

The unique answer set of $\mathcal{T}_{eng}(\mathcal{S}_{eng})|_\gamma$ contains, among others, atoms

$$switch(0) \neq neutral, \ldots, switch(3) \neq neutral,$$

$$arrivTime(fork) = 3, \quad arrived(dest).$$

Since $arrived$ is a time-independent fluent, we can conclude that $arrived(dest)$ is a change in this concrete causal theory. In general, we can check that, for any interpretation γ of the abstract theory $\mathcal{T}_{eng}(\mathcal{S}_{eng})$ where the switch is flipped before the train arrives to the fork, that is, satisfying $\gamma(t_4) < \gamma(t_3) + \gamma(time2fork)$, the unique answer set of $\mathcal{T}_{eng}(\mathcal{S}_{eng})|_\gamma$ contains atoms

$$switch(0) \neq neutral, \ldots, switch(n_1) \neq neutral,$$

$$arrivTime(fork) = n_1, \quad arrived(dest),$$

where $n_1 = \gamma(t_3) + \gamma(time2fork)$ is a natural number corresponding to the arrival time of the train to the switch. We can then conclude that $arrived(dest)$ is a change in this causal theory for any such interpretation γ. Figure 2 depicts (condensed versions) of the two proofs in this scenario for any such interpretation γ. In P_1, we reach the conclusion that the switch is not in the neutral position by inertia. In P_2, the same conclusion is the result of the flipping the switch to the *right*. Both are valid derivations of the change $arrived(dest)$. However, to infer the causes of an event we give preference to proofs using inertia over those using extra causal mechanisms. This idea is formalized in the following notion of *tight proof*.

https://doi.org/10.1017/S1471068423000248 Published online by Cambridge University Press

Definition 4 (*Tight proof*)
Let P_1 and P_2 be proofs of change $e(\bar{t},k)=y$ in $\mathcal{T}(\mathcal{S})|_\gamma$. P_1 is *(causally) tighter* than P_2 if every causal mechanism of P_1 belongs to P_2 but not vice-versa. Proof P of $e(\bar{t},k)=y$ in $\mathcal{T}(\mathcal{S})|_\gamma$ is *tight* if there is no proof of $e(\bar{t},k)=y$ in $\mathcal{T}(\mathcal{S})|_\gamma$ that is tighter than P.

Clearly, proof P_1 from our running example is tighter than P_2; causal mechanisms of P_1 are m_1, m_2 and m_4, while P_2 contains the additional causal mechanism $m_3(right)$.

Definition 5 (*Causal chain*)
Given a numeric time-step i and an atom $e(\bar{t},k)=y$ in $\mathcal{T}(\mathcal{S})|_\gamma$, a *causal chain* $Ch(i)$ from i to $e(\bar{t},k)=y$ is a sequence

$$a_1,\ldots,a_n,C_1,\ldots,C_m,e(\bar{t},k)=y,$$

of atoms and ground causal mechanisms of $\mathcal{T}(\mathcal{S})|_\gamma$ with $n\geq 1$ and $m\geq 0$ such that there is a tight proof P of $e(\bar{t},k)=y$ in $\mathcal{T}(\mathcal{S})|_\gamma$ satisfying the following conditions:

- a_1 is a do-atom from P with time-step i,
- $a_2,\ldots a_n$ are all other do-atoms from P with time-steps greater than or equal to i, and
- $C_1,\ldots,C_m$ are all causal mechanisms of P with time-steps greater than i.

Let us introduce some terminology. We say that $Ch(i)$ is *generated* from the proof P above. If $e(\bar{t},k)=y$ is a change, we say that causal chain from i to $e(\bar{t},k)=y$ in $\mathcal{T}(\mathcal{S})|_\gamma$ *leads to change* $e(\bar{t},k)=y$. A causal chain is *initiated* by the set of all its do-atoms. Two proofs of a set of ground atoms $\mathcal{U}$ are *equivalent* if they differ only by the order of their elements. Two chains are *equivalent* if they are generated from equivalent proofs.

Continuing with our running example, sequence

$$\mathbf{do}(approach,\gamma(t_3)),\ m_1,\ m_2,\ \ m_4,\ arrived(dest), \tag{19}$$

is a causal chain in this scenario that leads to change $arrived(dest)$, and it is generated by proof P_1 in Figure 2. However, sequence

$$\mathbf{do}(approach,\gamma(t_3)),\ \mathbf{do}(flipTo(right),\gamma(t_4)),\ m_1,\ m_2,\ m_3(right),\ m_4,\ arrived(dest),$$

corresponding to proof P_2 is not causal chain because P_2 is not a tight proof.

Definition 6 (*More informative causal chain*)
Given causal chains $Ch(i)$ and $Ch(j)$ to $e(\bar{t},k)=y$, we say that $Ch(i)$ *is more informative than* $Ch(j)$ if $i<j$ and $Ch(i)$ contains all elements of $Ch(j)$.

Definition 7 (*Candidate inflection point*)
A time-step i is called a *candidate inflection point of change* $e(\bar{t},k)=y$ *in* $\mathcal{T}(\mathcal{S})|_\gamma$ if it satisfies the following conditions:

(a) There is a causal chain from i to $e(\bar{t},k)=y$ in $\mathcal{T}(\mathcal{S}[i])|_\gamma$, and
(b) There is a causal chain from i to $e(\bar{t},k)=y$ in $\mathcal{T}(\mathcal{S})|_\gamma$

where $\mathcal{S}[i]$ is the scenario obtained from $\mathcal{S}$ by removing all **do**-atoms after i.

Definition 8 (*Inflection point*)
A candidate inflection point i is called an *inflection point of* $e(\bar{t},k)=y$ *in* $\mathcal{T}(\mathcal{S})|_\gamma$ if there is a causal chain $Ch(i)$ from i to $e(\bar{t},k)=y$ in $\mathcal{T}(\mathcal{S})|_\gamma$ such that there is no candidate

https://doi.org/10.1017/S1471068423000248 Published online by Cambridge University Press

inflection point j and causal chain $Ch(j)$ from j to $e(\bar{t},k) = y$ in $\mathcal{T}(\mathcal{S})|_\gamma$ which is more informative than $Ch(i)$.

Note that a scenario can have more than one inflection point (see Example 4).

Definition 9 (*Deliberate cause of change*)
A non-empty set α of do-atoms is called a *(deliberate) cause of change* $e(\bar{t},k) = y$ in $\mathcal{T}(\mathcal{S})|_\gamma$ if there is an inflection point i of $e(\bar{t},k) = y$ in $\mathcal{T}(\mathcal{S})|_\gamma$ such that α initiates a causal chain in $\mathcal{T}(\mathcal{S})|_\gamma$ from i to $e(\bar{t},k) = y$.

It is said to be *(deliberate) cause of change* $e(\bar{t},k) = y$ in $\mathcal{T}(\mathcal{S})$ if it is a cause of change $e(\gamma(k)) = y$ in $\mathcal{T}(\mathcal{S})|_\gamma$ for every interpretation γ of $\mathcal{T}(\mathcal{S})$.

Following with the Engineer example (Example 2), let us consider scenario $\mathcal{S}_{eng}$. Since this is an abstract scenario, to answer questions about the cause of change we have to consider all interpretations of this scenario. We proceed by cases. Let us first consider an interpretation γ satisfying condition $\gamma(t_4) < \gamma(t_3) + \gamma(\mathit{time2fork})$.

As we discussed above, (19) is the only causal chain leading to change $\mathit{arrived}(\mathit{dest})$. Furthermore, we can see that this is also a causal chain from $\gamma(t_3)$ to this change in $\mathcal{T}(\mathcal{S}[\gamma(t_3)])|_\gamma$. Therefore, time-step $\gamma(t_3)$ is the unique candidate inflection point of change $\mathit{arrived}(\mathit{dest})$ and, thus, it is the unique inflection point as well. As a result, singleton set $\{\mathbf{do}(\mathit{approach}, \gamma(t_3))\}$ is the unique cause of this change with respect to any such γ. Let us now consider an interpretation γ satisfying $\gamma(t_4) \geq \gamma(t_3) + \gamma(\mathit{time2fork})$. In this case $\mathit{arrived}(\mathit{dest})$ is still a change and P_1 is the only proof of this change. Hence, $\{\mathbf{do}(\mathit{approach}, \gamma(t_3))\}$ is also the unique cause of this change with respect to any such γ. Consequently, $\{\mathbf{do}(\mathit{approach}, t_3)\}$ is the unique cause of this change in this story.

Let us now consider *Suzy First* story (Example 1). The unique answer set of $\mathcal{T}_{fst}(\mathcal{S}_{suzy})|_\gamma$ contains atoms

$$\mathbf{do}(a_1, (\gamma(t_1)),\ \mathbf{do}(a_2, \gamma(t_2)),\ \neg \mathit{broken}(0),\ \ldots, \neg \mathit{broken}(n_4 - 1),\ \mathit{broken}(n_4),$$

with $n_4 = \gamma(t_1) + \gamma(\mathit{duration}(a_1))$ being a positive integer representing the arriving time-step of Suzy's rock. This means that $\mathit{broken}(n_4)$ is a change. There is only one causal chain leading to this change:

$$\mathbf{do}(a_1, \gamma(t_1)),\ m_0(a_1),\ \mathit{broken}(n_4), \tag{20}$$

and the only inflection point is $\gamma(t_1)$. As a result, Suzy's throw, $\{\mathbf{do}(a_1, t_1)\}$ is the only cause of this change. Note that the order in which Suzy and Billy throw is irrelevant (as long as the constraint $t_1 + \mathit{duration}(a_1) < t_2 + \mathit{duration}(a_2)$ is satisfied): the reason for Suzy's rock to get first may be because she throws first or because her rock is faster or any other reason. It is easy to check that, if we consider a scenario where Billy's rock gets first – formally a scenario $\mathcal{S}_{billy}$ obtained from $\mathcal{S}_{suzy}$ by replacing constraint $t_1 + \mathit{duration}(a_1) < t_2 + \mathit{duration}(a_2)$ by $t_1 + \mathit{duration}(a_1) > t_2 + \mathit{duration}(a_2)$ – then Billy's throw, $\{\mathbf{do}(a_2, t_2)\}$, is the only cause of this change.
In the following variation of *Suzy First* story *broken* has two inflection points.

Example 4
Suzy and Billy throw rocks at a bottle, but this time both rocks arrive at the same time.

This story can be formalized by a scenario $\mathcal{S}_{same}$ obtained from $\mathcal{S}_{suzy}$ by replacing $t_1 + \mathit{duration}(a_1) < t_2 + \mathit{duration}(a_2)$ by $t_1 + \mathit{duration}(a_1) = t_2 + \mathit{duration}(a_2)$

https://doi.org/10.1017/S1471068423000248 Published online by Cambridge University Press

For any interpretation γ of this scenario, we have change $broken(n_5)$ with $n_5 = \gamma(t_1) + duration(a_1) = \gamma(t_2) + duration(a_2)$ and two causal chains leading to this change:

$$\mathbf{do}(a_1, \gamma(t_1)),\ m_0(a_1),\ broken(n_5),$$

$$\mathbf{do}(a_2, \gamma(t_2)),\ m_0(a_2),\ broken(n_5).$$

Both $\gamma(t_1)$ and $\gamma(t_2)$ are inflection points. They may be the same inflection point or different ones depending of the interpretation γ.

In all cases, both $\{\mathbf{do}(a_1, \gamma(t_1))\}$ and $\{\mathbf{do}(a_2, \gamma(t_2))\}$ are causes of change $broken(n_5)$.

Let us now consider the variation of this story introduced in Example 3, where Suzy and Billy throw by the order of a stronger girl. As we discuss above, *broken* becomes true at time-step $t_1 + duration(a_1)$. In other words $broken(n_4)$ is a change in scenario $\mathcal{S}_{order}$.

In this scenario there is only one causal chain leading to this change:

$$\mathbf{do}(b_1, 0),\ m_6(a_1, t_1, b_1),\ m_0(a_1),\ broken(n_4),$$

and, thus, $\{\mathbf{do}(b_1, 0)\}$ is the only cause of $broken(n_4)$.

Note that our notion of cause is different from the notion of immediate or direct cause. The immediate cause of breaking the bottle is the throw of the rock, but the deliberate cause is the order. We also discussed the scenario where both Suzy and Billy refuse to follow the order and, thus, *broken* never happens. Therefore, there was no cause. Next let us consider a story where Suzy refuses to throw but Billy follows the order. This can be formalized by scenario $\mathcal{S}_{order3}$ obtained from $\mathcal{S}_{order}$ by adding extended atom $\mathbf{do}(\neg a_1, t_1)$. In this case the change happens later. That is, $broken(n_6)$ is a change with $n_6 = \gamma(t_2) + \gamma(duration(a_2))$. The only causal chain leading to this change is

$$\mathbf{do}(b_2, 0),\ m_6(a_2, t_1, b_2),\ m_0(a_2),\ broken(n_6),$$

and, thus, $\{\mathbf{do}(b_2, 0)\}$ is the only cause of $broken(n_6)$. Next example illustrates our treatment of preconditions of a cause.

Example 5

As in Example 1 Suzy picks up a rock and throws it at the bottle. However, this time we assume that she is accurate only if she aims first. Otherwise, her rock misses. Suzy aims before throwing and hits the bottle. Billy just looks at his colleague's performance.

The story is formalized by causal theory $\mathcal{T}_{aim}$:

$$\begin{array}{lll} m_0'(A): & broken(I) & \leftarrow \ occurs(A, I-D), member(A, throw), \\ & & \quad\ agent(A) = Ag, duration(A) = D, \\ & & \quad\ aimed(Ag, I-D), \neg broken(I-1), \neg ab(m_0'(A), I), \end{array}$$

$$\begin{array}{lll} m_7(A): & aimed(Ag, I) & \leftarrow \ occurs(A, I-D), member(A, aim), \\ & & \quad\ agent(A) = Ag, duration(A) = D, \neg ab(m_7(A), I), \end{array}$$

where *aimed* is an inertial fluent and scenario $\mathcal{S}_{aim}$

$$\begin{array}{lll} agent(a_1) = suzy & member(a_1, throw) & duration(a_1) \geq 1 \\ agent(c) = suzy & member(c, aim) & duration(c) \geq 1 \\ \mathbf{do}(c, t_5),\ \mathbf{do}(a_1, t_1),\ t_5 + duration(c) < t_1. & & \end{array} \tag{21}$$

The inflection point in $\mathcal{S}_{aim}$ is t_1 and the only deliberate cause of $broken(n_4)$ is $\{\mathbf{do}(a_1, t_1)\}$. Action $\mathbf{do}(c, t_5)$ is necessary for shattering the bottle, because it is

https://doi.org/10.1017/S1471068423000248 Published online by Cambridge University Press

required by one of the preconditions of $m_0'(a_1)$. However, it is not a deliberate cause because at the time of its occurrence the shattering could not be predicted (see condition (a) in Definition 7).

Definition 9 can be used to define the notion of *causal explanation* of unexpected observations: $T(S)$ is called *strongly consistent* if $T^{reg}(S)$, obtained from $T(S)$ by dropping cr-rules, is consistent. If $T(S)$ is strongly consistent and $T(S \cup \{\mathbf{obs}(f, y, i)\})$ is not we say that $\mathbf{obs}(f, y, i)$ is *unexpected*. We assume that every abductive support[4] U of this theory has exactly one answer set. By a cause of atom $f(i) = y$ we mean a cause of the last change of f to y which precedes $i + 1$ (note that for actions and time-independent fluent f, $f(i) = y$ is a change). By *causal explanation* of $\mathbf{obs}(f, y, i)$ we mean a cause of $f(i) = y$ in $T(S_U)$ where S_U is obtained from S by adding $\mathbf{do}(a, i)$ for every rule $a(i) \stackrel{+}{\leftarrow}$ from U for some abductive support U. For example, consider a scenario S of $\mathcal{T}_{fst}$ consisting of $\mathbf{init}(\neg broken), \mathbf{obs}(broken, true, 2)$, actions a_1 and a_2 from $\mathcal{S}_{suzy}$ with durations 2 and 4 respectively. The program has one abductive support, $a_1(0) \stackrel{+}{\leftarrow}$ and hence $\mathbf{do}(a_1, 0)$ explains the unexpected observation. If *broken* were observed at 3 we'd have two explanations: $\mathbf{do}(a_1, 0)$ and $\mathbf{do}(a_1, 1)$. This can be compactly represented using a do-atom $\mathbf{do}(a_1, t)$ where t is an abstract time-step satisfying $0 \leq t < 2$.

5 Conclusions

The paper describes a new approach for representing causal knowledge, and its use for causal analysis. The approach emphasizes the separation between background theory and scenario. The first contains general knowledge that may be shared by different stories and the latter contains the information specific to the considered story. This, together with the use of abstract constants, provides a higher degree of *elaboration tolerance* than other approaches to causal analysis. We also propose the use of a rich KR-language that is able to represent sophisticated causal laws, time, defaults and their exceptions, recursive definitions, and other non-trivial phenomena of natural language. As a result, we can obtain accurate and direct formalizations of natural language sentences that, we believe, is essential for causal analysis. We have illustrated this with common challenging examples from the literature on actual causality. Causal analysis is realized over a formal representation rather than over the natural language statements. However, our intuitions are usually more clear with respect to the natural language statements than with respect to the formal representation. The closer the formal representation is to the natural language statements of a story, the better we can use our intuition to guide us towards a formal analysis of actual causality. A preliminary version of this paper was presented at a workshop Gelfond and Balai (2020). We substantially extend that version and correct mistakes that were discovered after its presentation. This has led us to change the definition of inflection point, to introduce the notion of tight proof and abstract time-steps, etc. In the future, this work should be expanded to consider other types of causal relations. Some, like prevention, are not included due to space limitation. Others

[4] *Abductive support* of a program Π is a minimal collection of cr-rules of Π which, if turned into regular rules and added to the regular part of Π, produce a consistent program Π'. Answer set of Π is then defined as an answer set of Π'.

https://doi.org/10.1017/S1471068423000248 Published online by Cambridge University Press

require further work. In particular we plan to expand $\mathcal{W}$ by probabilistic constructs of P-log and use it to study probabilistic causal relations. Finally, we plan to investigate mathematical properties of causal theories and algorithms for effectively computing the causes of various causal relations and their implementations. The notion of tight proof is closely related to the notion of causal justifications for answer set programs Cabalar *et al.* (2014). This may open the door to use `xclingo` Cabalar *et al.* (2020) as the first-step of a new system for computing causes according to our definition.

References

Balai, E., Gelfond, M. and Zhang, Y. 2019. P-log: refinement and a new coherency condition. *Annals of Mathematics and Artificial Intelligence 86,* 1-3, 149–192.

Balduccini, M. 2012. Answer set solving and non-herbrand functions. In *Proceedings of the 14th International Workshop on Non-Monotonic Reasoning (NMR'2012)(Jun 2012)*.

Balduccini, M. and Gelfond, M. 2003. Logic programs with consistency-restoring rules. In *International Symposium on Logical Formalization of Commonsense Reasoning, AAAI 2003 Spring Symposium Series*, Vol. 102.

Baral, C., Gelfond, M. and Rushton, J. N. 2009. Probabilistic reasoning with answer sets. *Theory and Practice of Logic Programming 9,* 1, 57–144.

Batusov, V. and Soutchanski, M. 2018. Situation calculus semantics for actual causality. In *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 32.

Beckers, S. 2021. The counterfactual ness definition of causation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 6210–6217.

Beckers, S. and Vennekens, J. 2016. A general framework for defining and extending actual causation using cp-logic. *International Journal of Approximate Reasoning 77*, 105–126.

Beckers, S. and Vennekens, J. 2017. The transitivity and asymmetry of actual causation. *Ergo: An Open Access Journal of Philosophy 4*.

Beckers, S. and Vennekens, J. 2018. A principled approach to defining actual causation. *Synth. 195,* 2, 835–862.

Bochman, A. 2018. Actual causality in a logical setting. In *IJCAI*. ijcai.org, 1730–1736.

Cabalar, P. and Fandinno, J. 2016. Enablers and inhibitors in causal justifications of logic programs. *CoRR abs/1602.06897*.

Cabalar, P., Fandinno, J. and Fink, M. 2014. Causal graph justifications of logic programs. *Theory Pract. Log. Program. 14,* 4–5, 603–618.

Cabalar, P., Fandinno, J., and Muñiz, B. 2020. A system for explainable answer set programming. In *ICLP Technical Communications*. EPTCS, vol. 325. 124–136.

Chockler, H. and Halpern, J. Y. 2004. Responsibility and blame: A structural-model approach. *Journal of Artificial Intelligence Research 22*, 93–115.

Denecker, M., Bogaerts, B., and Vennekens, J. 2019. Explaining actual causation in terms of possible causal processes. In *European Conference on Logics in Artificial Intelligence*. Springer, 214–230.

Fandinno, J. 2016. Towards deriving conclusions from cause-effect relations. *Fundamenta Informaticae 147,* 1, 93–131.

Gelfond, M. and Balai, E. 2020. Causal analysis of events occurring in trajectories of dynamic domains. In *ICLP Workshops*. CEUR Workshop Proceedings, vol. 2678.

Gelfond, M. and Kahl, Y. 2014. *Knowledge Representation, Reasoning, and the Design of Intelligent Agents: The Answer-Set Programming Approach*. Cambridge University Press.

Hall, N. 2000. Causation and the price of transitivity. *Journal of Philosophy 97,* 4, 198–222.

https://doi.org/10.1017/S1471068423000248 Published online by Cambridge University Press

Hall, N. 2004. Two concepts of causation. In *Causation and Counterfactuals*, J. Collins, N. Hall, and L. A. Paul, Eds. Cambridge, MA: MIT Press, 225–276.

Hall, N. 2007. Structural equations and causation. *Philosophical Studies 132,* 1, 109–136.

Halpern, J. and Hitchcock, C. 2010. Actual causation and the art of modeling. 383–406.

Halpern, J. Y. 2014. Appropriate causal models and stability of causation. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourteenth International Conference, KR 2014, Vienna, Austria, July 20–24, 2014*. AAAI Press.

Halpern, J. Y. 2015. A modification of the Halpern-Pearl definition of causality. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015*, Q. Yang and M. Wooldridge, Eds. AAAI Press, 3022–3033.

Halpern, J. Y. and Pearl, J. 2001. Causes and explanations: A structural-model approach: Part 1: Causes. In *Proceedings of the Seventeenth Conference in Uncertainty in Artificial Intelligence, UAI 2001, University of Washington, Seattle, Washington, USA, August 2-5, 2001*. Morgan Kaufmann, 194–202.

Hopkins, M. and Pearl, J. 2003. Clarifying the usage of structural models for commonsense causal reasoning. In *Proceedings of the AAAI Spring Symposium on Logical Formalizations of Commonsense Reasoning*, 83–89.

Hopkins, M. and Pearl, J. 2007. Causality and counterfactuals in the situation calculus. *Journal of Logic and Computation 17,* 5, 939–953.

Inclezan, D. 2016. Corealmlib: An ALM library translated from the component library. *Theory and Practice of Logic Programming 16,* 5–6, 800–816.

LeBlanc, E. C., Balduccini, M., and Vennekens, J. 2019. Explaining actual causation via reasoning about actions and change. In *JELIA*. Lecture Notes in Computer Science, vol. 11468. Springer, 231–246.

McCarthy, J. 1998. Elaboration tolerance.

Pearl, J. 2009. *Causality*. Cambridge university press.

Vennekens, J. 2011. Actual causation in CP-logic. *TPLP 11,* 4–5, 647–662.

https://doi.org/10.1017/S1471068423000248 Published online by Cambridge University Press