Credulous Acceptability, Poison Games and Modal Logic

Davide Grossi and Simon Rey

ABSTRACT

The Poison Game is a two-player game played on a graph in which one player can influence which edges the other player is able to traverse. It operationalizes the notion of existence of credulously admissible sets in an argumentation framework or, in graph-theoretic terminology, the existence of non-trivial semi-kernels. We develop a modal logic (poison modal logic, PML) tailored to represent winning positions in such a game, thereby identifying the precise modal reasoning that underlies the notion of credulous admissibility in argumentation. We study model-theoretic and decidability properties of PML, and position it with respect to recently studied logics at the cross-road of modal logic, argumentation, and graph games.

KEYWORDS

Modal logic; Dynamic logic; Memory logic; Poison Game; Games on graphs; Argumentation theory; Credulous acceptability

1 INTRODUCTION

In abstract argumentation theory [8, 9], an argumentation framework (or *attack graph*) [18] is a directed graph $(A, \rightarrow)$ where A is a set of nodes (or *arguments*) and $\rightarrow \subseteq A^2$ is a set of directed edges (or *attacks*). For $x, y \in A$ such that $x \rightarrow y$ we say that x attacks y. An *admissible* set [18], of a given attack graph, is a set $X \subseteq A$ such that: (a) no two nodes in X attack one another; and (b) for each node $y \in A \backslash X$ attacking a node in X, there exists a node $z \in X$ attacking y. That is, X is internally coherent, and can counterattack any attack moved to any of its arguments. Such sets are also called *semi-kernels* in the theory of directed graphs [20]. More precisely, if X is an admissible set of $(A, \rightarrow)$, then it is a semi-kernel of the directed graph obtained by inverting the attack relation $\rightarrow$ (i.e., the 'being attacked' graph), and vice versa. These sets form the basis of all main argumentation semantics first developed in [18] and they are central to the influential graph-theoretic systematization of logic programming and default reasoning pursued in [15, 16], where they have been proven to correspond to the so-called partial stable models of logic programming [30].

Contribution. Given the importance of admissible sets in argumentation theory, one of the key reasoning tasks in abstract argumentation consists in deciding whether any given argumentation framework contains non-empty admissible sets. In the terminology of argumentation, this amounts to deciding whether the framework contains any *credulously admissible* arguments. The property corresponds in turn to the existence of non-trivial semi-kernels in the inverted attack graph. Credulous acceptability is a benchmark semantics for the evaluation of arguments in abstract argumentation [10]. Interestingly, the notion has an elegant operationalization in the form of two-player games, called *Poison Game* in the graph theory literature [17], and *game for credulous acceptance* in the argumentation theory literature [28, 41]. The poison game is the starting point of the paper. Inspired by it we define a new modal logic, called *poison modal logic* (PML), whose operators capture the strategic abilities of players in the Poison Game, and are therefore fit to capture the modal reasoning involved in the notion of credulous admissibility. This answers, at least in part, a research question left open in [21]. The paper studies PML by: defining a suitable notion of bisimulation for it, which in turn answers another open question [19] concerning the logic of credulous admissibility, namely a notion of structural equivalence tailored for it; establishing a first-order characterization result for PML in the tradition of [35]; proving the undecidability of satisfiability in a multi-modal variant of PML; and exploring its links with hybrid [13] and memory logics [3, 4]. More broadly we see the paper as a contribution to bridging, in a systematic way, concepts from abstract argumentation theory [18], games on graphs [11] and modal logic [12].

Related work. The paper is a natural continuation of the line of work interfacing abstract argumentation and modal logic [19, 21–23, 32, 33], which focuses on the modal logic characterization of key argumentation-theoretic notions, and their analysis through model and proof-theoretic tools. A first bimodal dynamic logic tailored to the poison game was introduced in [27], where two modalities are used to keep track of which parts of the underlying graph are accessible to each player. Our approach is somewhat simpler and based on the combination of one classical and one dynamic modality. PML is also directly related to so-called memory logics, extensively studied in the last decade [3, 4]. In fact PML can be thought of as a modal logic with two operators: a standard one, and one which 'memorizes' the states which are reached by traversing an edge of the underlying frame. The paper relates also to the research program investigating the modal logic theory of graph games, sparked by the recent work on sabotage modal logic (SML, [1, 2, 7, 26, 31, 36]). SML was tailored to capture the logic behind winning strategies in a specific two-player, perfect-information, zero-sum game played on graphs, known as the sabotage game [36]. Like SML, PML sits at the intersection of two well-established lines of research in modal logic: dynamic epistemic logic [40] and the logical dynamics tradition it generated, which is broadly concerned with the study of operators interpreted on transformations of semantics structures [2, 6, 25, 37]; and game logics concerning the logical analysis of games [38, 39].

Outline. The article is organized as follow. First we set the ground in Section 2 by showing how the standard modal language can already capture the key logic behind statements of this type: "set X is a semi-kernel" (of a given directed graph), and by introducing the Poison Game. Section 3 introduces PML and establishes some basic facts. Section 4 gives a translation into First Order Logic (FOL) which is invariant for the poison bisimulation as defined in Section 4.2. Decidability is addressed in Section 5. Section 6 explores links between PML and other logical frameworks and concludes.

Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), N. Agmon, M. E. Taylor, E. Elkind, M. Veloso (eds.), May 2019, Montreal, Canada

https://doi.org/doi

2 PRELIMINARIES

We start by providing some preliminaries on existing bridges between modal logic and abstract argumentation, and a concise presentation of the Poison Game.

2.1 Modal Logic and Credulous Admissibility

As hinted in the introduction, and following [21] we study attack graphs $(A, \rightarrow)$ through their inversions (the 'being attacked' graphs) $(A, \rightarrow^{-1})$ which we view as Kripke frames [12] (W, R) where, $W = A$, and $R = \rightarrow^{-1}$. So, writing wRw' stands for argument w is attacked by argument w'.[1] In this view, a Kripke model $M = (W, R, V)$, where V is a valuation function $V : \mathbf{P} \rightarrow 2^W$, can be thought of an argumentation framework where propositional labels in $\mathbf{P}$ are assigned to sets of arguments. The standard modal language,

$$\mathcal{L} : \varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Diamond\varphi,$$

becomes therefore a language in which it is possible to express properties of argumentation frameworks. Through the standard modal semantic clause

$$(M, w) \models \Diamond\varphi \Leftrightarrow \exists w' \in W, wRw', (W, R, V), b \models \varphi, \qquad (1)$$

formulas $\Diamond\varphi$ are statements of the type "the current argument is attacked by an argument in the set of arguments denoted by φ". Specifically, as shown in [21] a number of key argumentation-theoretic properties are expressible in the standard modal language extended with the universal modality $[U]$ (known as logic K^U [12]). In particular, formula

$$[U](p \rightarrow \neg\Diamond p) \wedge [U](p \rightarrow \Box\Diamond p) \qquad (2)$$

expresses the property "the set denoted by p under function V is admissible" (in the underlying argumentation framework) [21]. So a well-studied logic such as K^U suffices to express that a given set of arguments is admissible. However, the existence of credulously admissible arguments has an obvious second-order flavor and the question of whether it could be modally expressed without resorting to second-order modal logic remained an open question in [21].

2.2 The Poison Game

The Poison Game was introduced in [17] to characterize the existence of non-empty semi-kernels in directed graphs. A very similar game was later independently introduced in [41] to characterize credulous admissibility of arguments in argumentation frameworks.[2] Our presentation of the game follows that of [17].

The Poison Game is a two-player ($\mathbb{P}$, the proponent, and $\mathbb{O}$, the opponent), win-lose, perfect-information game [29] played on a directed graph (W, R). The game starts by $\mathbb{P}$ selecting a node $w \in W$. After this initial choice, $\mathbb{O}$ selects a successor of the node picked by $\mathbb{P}$, $\mathbb{P}$ then selects a successor of the node picked by $\mathbb{O}$ and so on. However, while $\mathbb{O}$ can choose any successor of the current node, $\mathbb{P}$ can only select successors which have not yet been visited—*poisoned*—by $\mathbb{O}$. $\mathbb{O}$ wins if and only if $\mathbb{P}$ ends up in a position with no available successors. In all other cases the game is won by $\mathbb{P}$.

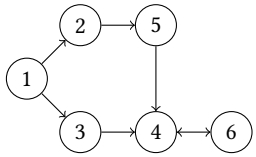

Figure 1: Graph of Examples 2.1 and 2.3. The graph is (the inversion of) a framework discussed in [41].

Example 2.1. A possible run of the Poison Game on the graph depicted in Figure 1 is: $\mathbb{P}$ starts by selecting node 1, then $\mathbb{O}$ moves to and poisons node 3, $\mathbb{P}$ answers by moving to 4, $\mathbb{O}$ in returns moves to 6 and from there the game will repeat the two last moves indefinitely. Player $\mathbb{P}$ therefore wins the game.

What makes this game interesting is that the existence of a winning strategy for $\mathbb{P}$, if (W, R) is finite[3], is equivalent to the existence of a (non-empty) semi-kernel in the graph or, in the argumentation terminology, the existence of credulously admissible arguments in the inverted graph (W, R^{-1}).

Theorem 2.2 (Duchet and Meyniel [17]). *Let (W, R) be a finite directed graph. There exists a non-empty semi-kernel in (W, R) if and only if $\mathbb{P}$ has a winning strategy in the Poison Game for (W, R).*

Sketch of proof. Left-to-right If a non-empty semi-kernel $X \subset W$ exists, then $\mathbb{P}$ can win the game simply by picking the initial node in X and then responding to each move of $\mathbb{P}$ with a successor in X, which is guaranteed to exist since X is a semi-kernel. Right-to-left If $\mathbb{P}$ has a winning strategy, she can play indefinitely no matter what $\mathbb{O}$ does. As W is finite, this means that $\mathbb{P}$ visits a finite set of states infinitely often. Call such set X. It suffices to show that X is indeed a semi-kernel. Clearly no state in X has a successor in X (i.e., X is independent), as otherwise such successor would have been poisoned by $\mathbb{O}$. Moreover for each state $x \in X$, for any successor $y \in W \backslash X$ of x that can be selected by $\mathbb{O}$, there exists a successor z of y that can be selected by $\mathbb{P}$ infinitely often, therefore belonging to X. □

Example 2.3. In the graph of Figure 1: $\{4\}$ and $\{6\}$ are two semi-kernels. $\mathbb{P}$ has several winning strategies in the game played on this graph. She can choose 1 and force the infinite run described in Example 2.1. Alternatively she could simply choose 4 or 6 and again force an infinite run.

3 POISON MODAL LOGIC (PML)

This section introduces the syntax and semantics of PML, discusses some of its validities and some properties it is able to express. The language we propose is directly motivated by the Poison Game: the standard modality $\Diamond$ tracks moves of $\mathbb{P}$ selecting successors of a current state, the novel poison modality $\blacklozenge$ tracks moves of $\mathbb{O}$ selecting successors of a current state *and* poisoning them.

[1] In what follows we will refer to (W, R) also as attack graphs even though, technically speaking, they are inversions of attack graphs.

[2] A detailed comparison of the two games is not in the scope of this paper. Albeit very similar, the two games are from a technical point of view slightly different, and are adequate with respect to slightly different notions: the poison game is adequate w.r.t. the existence of non-empty admissible sets; the game from [41] is adequate w.r.t. the membership of one given argument to at least one admissible set.

[3] The result holds also with a weaker condition requiring every weakly connected component of the directed graph to be finite.

3.1 Syntax & Semantics

The poison modal language $\mathcal{L}^{\mathfrak{p}}$ is defined by the following BNF:

$$\mathcal{L}^{\mathfrak{p}} : \varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Diamond\varphi \mid \blacklozenge\varphi,$$

where $p \in \mathbf{P} \cup \{\mathfrak{p}\}$ with $\mathbf{P}$ a countable set of propositional atoms and $\mathfrak{p}$ a distinguished atom called *poison atom*. We will also discuss multi-modal variants of the above language, denoted $\mathcal{L}^{\mathfrak{p}}_n$, where $n \geq 1$ denotes the number of distinct pairs $(\Diamond_i, \blacklozenge_i)$ of modalities, with $1 \leq i \leq n$ and where each $\blacklozenge_i$ comes equipped with a distinct poison atom $\mathfrak{p}_i$.

This language is interpreted on Kripke models $\mathcal{M} = (W, R, V)$, as defined above. When confusion may arise, we will write $W^{\mathcal{M}}$, $R^{\mathcal{M}}$ and $V^{\mathcal{M}}$ to refer to the elements of the model $\mathcal{M}$. We will note $w \in \mathcal{M}$ to say $w \in W^{\mathcal{M}}$. A pointed model is a pair $(\mathcal{M}, w)$ with $w \in \mathcal{M}$. We call $\mathfrak{M}$ the set of all pointed models and $\mathfrak{M}^{\emptyset}$ the set of pointed models $(\mathcal{M}, w)$ such that $V^{\mathcal{M}}(\mathfrak{p}) = \emptyset$, that is the class of pointed models where no state satisfies $\mathfrak{p}$.

We define now an operation $\bullet$ on models which, given an input model and a state, modifies its function V by adding that state to $V(\mathfrak{p})$. Formally, for $\mathcal{M} = (W, R, V)$ and $w \in W$:

$$\mathcal{M}^{\bullet}_w = (W, R, V)^{\bullet}_w = (W, R, V'),$$

where $\forall p \in \mathbf{P}, V'(p) = V(p)$ and $V'(\mathfrak{p}) = V(\mathfrak{p}) \cup \{w\}$. We are now equipped to formally define the semantics of $\mathcal{L}^{\mathfrak{p}}$.

Definition 3.1 (Satisfaction relation). Let $\mathcal{M} \in \mathfrak{M}$. The satisfaction relation of PML is defined recursively as follows:

$$(\mathcal{M}, w) \models p \iff w \in V(p), \forall p \in \mathbf{P} \cup \{\mathfrak{p}\}$$
$$(\mathcal{M}, w) \models \neg\varphi \iff (\mathcal{M}, w) \not\models \varphi$$
$$(\mathcal{M}, w) \models \varphi \wedge \psi \iff (\mathcal{M}, w) \models \varphi \text{ and } (\mathcal{M}, w) \models \psi$$
$$(\mathcal{M}, w) \models \Diamond\varphi \iff \exists v \in W, wRv, (\mathcal{M}, v) \models \varphi$$
$$(\mathcal{M}, w) \models \blacklozenge\varphi \iff \exists v \in W, wRv, (\mathcal{M}^{\bullet}_v, v) \models \varphi.$$

The poison formula $\blacklozenge\varphi$ is then true at w in $\mathcal{M}$ if and only if φ is true at a successor w' of w in the model obtained from $\mathcal{M}$ by adding w' to the valuation of the poison atom $\mathfrak{p}$. Validity in a model and in a frame are defined in the usual way, but the relevant class of models to specify PML is $\mathfrak{M}^{\emptyset}$, that is, those models where $\mathfrak{p}$ starts with an empty valuation. PML is therefore the set of formulas which are valid in $\mathfrak{M}^{\emptyset}$. Similarly, PML_n is the set of formulas of $\mathcal{L}^{\mathfrak{p}}_n$ which are valid in $\mathfrak{M}^{\emptyset}$. We introduce some auxiliary definitions.

Definition 3.2 (Poison modal theory). The poison modal theory of a pointed model $(\mathcal{M}, w)$ with $\mathcal{M} \in \mathfrak{M}$ is the set $\mathbb{T}^{\mathfrak{p}}(\mathcal{M}, w) \subseteq \mathcal{L}^{\mathfrak{p}}$ of formulas defined as follows:

$$\mathbb{T}^{\mathfrak{p}}(\mathcal{M}, w) = \{\varphi \in \mathcal{L}^{\mathfrak{p}} \mid (\mathcal{M}, w) \models \varphi\}.$$

Definition 3.3 (Poison relation). The poisoning relation $\overset{\bullet}{\rightarrow}$ between two pointed models is defined as:

$$(\mathcal{M}, w) \overset{\bullet}{\rightarrow} (\mathcal{M}', w') \iff wR^{\mathcal{M}}w' \text{ and } \mathcal{M}' = \mathcal{M}^{\bullet}_{w'}.$$

Furthermore, we denote $(\mathcal{M}, w)^{\bullet}$ the set of all pointed models accessible from $\mathcal{M}$ via a poisoning relation, formally:

$$(\mathcal{M}, w)^{\bullet} = \{(\mathcal{M}', w') \mid (\mathcal{M}, w) \overset{\bullet}{\rightarrow} (\mathcal{M}', w')\}.$$

Definition 3.4 (Poison modal equivalence). Two pointed models $(\mathcal{M}, w)$ and $(\mathcal{M}', w')$ are poison modally equivalent—in symbols, $(\mathcal{M}, w) \overset{\mathfrak{p}}{\leftrightsquigarrow} (\mathcal{M}', w')$—if and only if, $\forall\varphi \in \mathcal{L}^{\mathfrak{p}}$:

$$(\mathcal{M}, w) \models \varphi \iff (\mathcal{M}', w') \models \varphi.$$

3.2 Validity and Expressivity: Examples

Fact 1. *Let $p \in \mathbf{P}$ and $\varphi, \psi \in \mathcal{L}^{\mathfrak{p}}$. The following formulas are validities of* PML *(w.r.t. class $\mathfrak{M}^{\emptyset}$):*

$$\neg\mathfrak{p} \wedge \blacksquare\mathfrak{p} \tag{3}$$
$$\Box\bot \rightarrow \blacksquare\varphi \tag{4}$$
$$\blacksquare p \leftrightarrow \Box p \tag{5}$$
$$\Box\mathfrak{p} \rightarrow (\blacksquare\varphi \leftrightarrow \Box\varphi) \tag{6}$$
$$\blacksquare(\varphi \wedge \psi) \leftrightarrow (\blacksquare\varphi \wedge \blacksquare\psi) \tag{7}$$
$$\blacksquare\neg\varphi \rightarrow (\Box\bot \vee \neg\blacksquare\varphi) \tag{8}$$

Proofs are omitted. It can also be immediately noticed that PML is not closed under uniform substitution. For instance, the schematic version $\blacksquare\varphi \leftrightarrow \Box\varphi$ of Formula (5) is clearly invalid.

To illustrate the expressive power of PML, we show that it is possible to express the existence of cycles in the modal frame, a property not expressible in the standard modal language. Consider the class of formulas δ_n, with $n \in \mathbb{N}_{>0}$, defined inductively as follows, with $i < n$: [Base] $\delta_1 = \Diamond\mathfrak{p}$; [Step] $\delta_{i+1} = \Diamond(\neg\mathfrak{p} \wedge \delta_i)$.

Fact 2. *Let $\mathcal{M} = (W, R, V) \in \mathfrak{M}^{\emptyset}$, then for $i, n \in \mathbb{N}_{>0}$ there exists $w \in W$ such that $(\mathcal{M}, w) \models \blacklozenge\delta_n$ if and only if there exists a cycle of length $i \leq n$ in the frame (W, R).*

Proof. Observe that the formula $\blacklozenge\delta_n$ has only one occurrence of the poison modality $\blacklozenge$. As $V(\mathfrak{p}) = \emptyset$ by assumption, the only poisoned state when we go through the formula is the one poisoned by $\blacklozenge$. The formula then states that one can reach that unique poisoned state without passing through other poisoned states in n steps. It follows that a cycle exists whose length i is n or smaller. □

A direct consequence of Fact 2 is that PML is not bisimulation invariant. In particular, its formulas are not preserved by tree-unravelings and it does not enjoy the tree model property.

3.3 Winning Strategies of the Poison Game

PML can express winning positions (that is, states in a graph in which a player has a winning strategy) in a natural way. Given a frame (W, R), nodes satisfying formulas $\blacklozenge\Box\mathfrak{p}$ are winning for $\mathbb{O}$ as she can move to a dead end for $\mathbb{P}$. So are also nodes satisfying formula $\blacklozenge\Box\blacklozenge\Box\mathfrak{p}$: she can move to a node in which, no matter which successor $\mathbb{P}$ chooses, she can then push her to a dead end. In general, winning positions for $\mathbb{O}$ are defined by the following infinitary $\mathcal{L}^{\mathfrak{p}}$-formula:

$$\blacklozenge\Box\mathfrak{p} \vee \blacklozenge\Box\blacklozenge\Box\mathfrak{p} \vee \ldots \tag{9}$$

Dually, winning positions for $\mathbb{P}$ are defined by the following infinitary $\mathcal{L}^{\mathfrak{p}}$-formula:

$$\blacksquare\Diamond\neg\mathfrak{p} \wedge \blacksquare\Diamond\blacksquare\Diamond\neg\mathfrak{p} \wedge \ldots \tag{10}$$

Remark 1 (Credulous admissibility and PML)**.** *By Theorem 2.2, formula* (10)*, interpreted on the inversion of an argumentation framework, expresses the property "there exist credulously admissible arguments in the framework". To the best of our knowledge, this is the first modal characterization of the notion, albeit an infinitary one.*[4]

4 EXPRESSIVITY OF PML

4.1 Translation into First-Order Logic

Let $\mathcal{L}$ be the language of the binary fragment of first-order logic (FOL) with equality. We present here a translation of the language of PML into $\mathcal{L}$.

*Definition 4.1 (*FOL*translation).* Let $p, q, \ldots \in \mathbf{P}$ be propositional atoms, we call $P, Q, \ldots$ their corresponding first-order predicate. The first-order predicate for the poison atom $\mathfrak{p}$ is $\mathfrak{P}$. Let N be a finite set of variables, and x a designated variable, the translation $ST_x^N : \mathcal{L}^{\mathfrak{p}} \to \mathcal{L}$ is defined inductively as follows:

$$\begin{aligned}
ST_x^N(p) &= P(x), \forall p \in \mathbf{P}\\
ST_x^N(\neg\varphi) &= \neg ST_x^N(\varphi)\\
ST_x^N(\varphi \wedge \psi) &= ST_x^N(\varphi) \wedge ST_x^N(\psi)\\
ST_x^N(\Diamond\varphi) &= \exists y \left(xRy \wedge ST_y^N(\varphi)\right)\\
ST_x^N(\blacklozenge\varphi) &= \exists y \left(xRy \wedge ST_y^{N\cup\{y\}}(\varphi)\right)\\
ST_x^N(\mathfrak{p}) &= \mathfrak{P}(x) \vee \bigvee_{y\in N}(y = x).
\end{aligned}$$

The definition is naturally extended to inputs consisting of sets of formulas. Let us briefly comment on the translation. A state is poisoned either if it is in the valuation of $\mathfrak{p}$, or if it has been poisoned by traversing a link instantiating the semantics of the poison modality, in which case the world is added to N which 'book-keeps' the set of poisoned states. It is worth noticing that the translation does not, in general, returns a formula with only one free variable. It does so, however, when N is set to $\emptyset$. We move now to proving that the translation is correct.

Lemma 4.2. *For a model $\mathcal{M}$ and an assignment g:*

$$\mathcal{M}^\bullet_w \models ST_x^N(\varphi)[g] \Longleftrightarrow \mathcal{M} \models ST_x^{N\cup\{y\}}(\varphi)[g_{y:=w}].$$

Sketch of proof. We prove the lemma by induction on the structure of φ (standard cases are omitted).

$\boxed{\varphi = \mathfrak{p}}$ This case (part of the induction base) is established by the following series of equivalences, using the definitions of the $^\bullet$ operation on models and of the standard translation.

$$\begin{aligned}
\mathcal{M}^\bullet_w \models ST_x^N(\mathfrak{p})[g] &\Leftrightarrow \mathcal{M}^\bullet_w \models \left(\mathfrak{P}(x) \vee \bigvee_{y\in N}(y = x)\right)[g]\\
&\Leftrightarrow \mathcal{M} \models \left(\mathfrak{P}(x) \vee \bigvee_{y\in N\cup\{w\}}(y = x)\right)[g]\\
&\Leftrightarrow \mathcal{M} \models ST_x^{N\cup\{y\}}(\mathfrak{p})[g_{y:=w}].
\end{aligned}$$

[4] Formulas (9) and (10) call naturally for a fixpoint extension of PML. Such an extension poses interesting technical challenges very similar to those charted in [7] for a μ-calculus extension of sabotage modal logic.

$\boxed{\varphi = \blacklozenge\psi}$ with $\psi \in \mathcal{L}^{\mathfrak{p}}$. The case is established by the following series of equivalences, using the definitions of the $^\bullet$ operation on models, of the standard translation, the semantics of $\blacklozenge$ and $\wedge$, and the induction hypothesis.

$$\begin{aligned}
\mathcal{M}^\bullet_w \models ST_x^N(\blacklozenge\psi)[g] &\Leftrightarrow \mathcal{M}^\bullet_w \models \exists y \left(xRy \wedge ST_y^{N\cup\{y\}}(\psi)\right)[g]\\
&\Leftrightarrow \exists v, g(x)Rv, \mathcal{M}^\bullet_w \models ST_y^{N\cup\{y\}}(\psi)[g_{y:=v}]\\
&\Leftrightarrow \exists v, g(x)Rv, \mathcal{M} \models ST_y^{N\cup\{y,z\}}(\psi)[g_{y:=v,z:=w}]\\
&\Leftrightarrow \mathcal{M} \models \exists y \left(xRy \wedge ST_y^{N\cup\{y,z\}(\psi)}\right)[g_{z:=w}]\\
&\Leftrightarrow \mathcal{M} \models ST_x^{N\cup\{z\}}(\blacklozenge\psi)[g_{z:=w}].
\end{aligned}$$

This completes the proof. □

Theorem 4.3. *Let $(\mathcal{M}, w)$ be a pointed model and $\varphi \in \mathcal{L}^{\mathfrak{p}}$ a formula, we have then:*

$$(\mathcal{M}, w) \models \varphi \Longleftrightarrow \mathcal{M} \models ST_x^{\emptyset}(\varphi)[x := w].$$

Sketch of proof. The proof is by induction on the structure of φ (standard cases are omitted).

$\boxed{\varphi = \mathfrak{p}}$ This case (part of the base case) is established through the following series of simple equivalences:

$$\begin{aligned}
(\mathcal{M}, w) \models \mathfrak{p} &\Leftrightarrow \mathcal{M} \models \mathfrak{P}(x)[x := w]\\
&\Leftrightarrow \mathcal{M} \models \mathfrak{P}(x) \vee \bigvee_{y\in\emptyset}(y = x)[x := w]\\
&\Leftrightarrow \mathcal{M} \models ST_x^{\emptyset}(\mathfrak{p})[x := w].
\end{aligned}$$

$\boxed{\varphi = \blacklozenge\psi}$ with $\psi \in \mathcal{L}^{\mathfrak{p}}$. This case is established through the following series of equivalences, using the semantics of $\blacklozenge$, the definition of the standard translation and Lemma 4.2:

$$\begin{aligned}
(\mathcal{M}, w) \models \blacklozenge\psi &\Leftrightarrow \exists v, wRv, (\mathcal{M}^\bullet_v, w) \models \psi\\
&\Leftrightarrow \exists v, wRv, \mathcal{M}^\bullet_v \models ST_x^{\emptyset}(\psi)[x := w]\\
&\Leftrightarrow \exists v, wRv, \mathcal{M} \models ST_x^{\{y\}}(\psi)[x := w, y := v]\\
&\Leftrightarrow \mathcal{M} \models \exists y \left(xRy \wedge ST_x^{\{y\}}(\psi)\right)[x := w]\\
&\Leftrightarrow \mathcal{M} \models ST_x^{\emptyset}(\blacklozenge\psi)[x := w].
\end{aligned}$$

This completes the proof. □

4.2 Poison Bisimulation

As observed earlier PML is not bisimulation invariant. In what follows we define a notion of bisimulation tailored to PML.

Definition 4.4 (p-bisimulation). Two pointed models $(\mathcal{M}_1, w_1)$ and $(\mathcal{M}_2, w_2)$ are said to be p-bisimilar, written $(\mathcal{M}_1, w_1) \overset{\mathfrak{p}}{\rightleftharpoons} (\mathcal{M}_2, w_2)$, if there exists a relation $Z \subseteq W^{\mathcal{M}_1} \times W^{\mathcal{M}_2}$ (the p-bisimulation relation) such that $w_1 Z w_2$ and, for any states $w \in W^{\mathcal{M}_1}$ and $v \in W^{\mathcal{M}_2}$, whenever wZv the following clauses are satisfied:

Atom: For any atom $p \in \mathbf{P} \cup \{\mathfrak{p}\}$, $w \in V^{\mathcal{M}_1}(p)$ iff $v \in V^{\mathcal{M}_2}(p)$.

Zig$_\Diamond$: If there exists $w' \in \mathcal{M}_1$ such that $wR^{\mathcal{M}_1}w'$ then there exists $v' \in \mathcal{M}_2$ such that $vR^{\mathcal{M}_2}v'$ and $(\mathcal{M}_1, w')Z(\mathcal{M}_2, v')$.

Zag$_\Diamond$: If there exists $v' \in \mathcal{M}_2$ such that $vR^{\mathcal{M}_2}v'$ then there exists $w' \in \mathcal{M}_1$ such that $wR^{\mathcal{M}_1}w'$ and $(\mathcal{M}_1, w')Z(\mathcal{M}_2, v')$.

Zig$_\blacklozenge$: If there exists $(\mathcal{M}_1', w_1')$ such that $(\mathcal{M}_1, w_1) \overset{\bullet}{\rightarrow} (\mathcal{M}_1', w_1')$, then there exists $(\mathcal{M}_2', w_2')$ such that $(\mathcal{M}_2, w_2) \overset{\bullet}{\rightarrow} (\mathcal{M}_2', w_2')$ and $(\mathcal{M}_1', w_1')Z(\mathcal{M}_2', w_2')$.

Zag$_\blacklozenge$: If there exists $(\mathcal{M}_2', w_2')$ such that $(\mathcal{M}_2, w_2) \overset{\bullet}{\rightarrow} (\mathcal{M}_2', w_2')$, then there exists $(\mathcal{M}_1', w_1')$ such that $(\mathcal{M}_1, w_1) \overset{\bullet}{\rightarrow} (\mathcal{M}_1', w_1')$ and $(\mathcal{M}_1', w_1')Z(\mathcal{M}_2', w_2')$.

An example of a p-bisimulation relation is depicted in Figure 2. Observe that, unlike for bisimulation in the standard modal language, p-bisimulation involves transitions between pointed models in the clauses Zig$_\blacklozenge$ and Zag$_\blacklozenge$. For a simple example of two models which are not p-bisimilar consider a model consisting of just one reflexive point, and its unraveling in an infinite chain.

Remark 2 (Argumentation and p-bisimulation)**.** *As argued in [19], modal bisimulation formalizes a natural notion of similarity of argumentation frameworks that preserves important argumentation-theoretic notions. It has for instance been shown [21, Th. 6] that, given two totally bisimilar models $\mathcal{M}_1$ and $\mathcal{M}_2$, a set of arguments denoted by p in $\mathcal{M}_1$ is admissible (respectively, complete, stable or grounded) in the frame of $\mathcal{M}_1$, if and only if the set of arguments denoted by p in $\mathcal{M}_2$ is admissible (respectively, complete, stable or grounded) in the frame of $\mathcal{M}_2$. Which strengthening of the notion of bisimulation is needed to guarantee the preservation of credulous admissibility across frameworks was mentioned as an open question in [19]. P-bisimulation provides an elegant answer.*

4.3 Characterization

The aim of this section is to establish a characterization theorem (Theorem 4.7) in the tradition of [35]. The standard proof methods can be adapted easily to fit PML. We start by precisely relating p-bisimulation with poison modal equivalence.

Theorem 4.5. *For two pointed models $(\mathcal{M}_1, w_1)$ and $(\mathcal{M}_2, w_2)$, if $(\mathcal{M}_1, w_1) \overset{\text{p}}{\rightleftharpoons} (\mathcal{M}_2, w_2)$ then $(\mathcal{M}_1, w_1) \overset{\text{p}}{\rightsquigarrow} (\mathcal{M}_2, w_2)$.*

Sketch of proof. The proof is by induction on the structure of formulas. The base case is covered by the atomic condition of the definition of p-bisimulation. For the inductive case, we provide details only for the $\blacklozenge$ modality. Let Z be the p-bisimulation relation. Suppose that $(\mathcal{M}_1, w_1) \models \blacklozenge\varphi$, then given the semantics of $\blacklozenge$, there exists $(\mathcal{M}_1', w_1')$ such that $(\mathcal{M}_1, w_1) \overset{\bullet}{\rightarrow} (\mathcal{M}_1', w_1')$ and $(\mathcal{M}_1', w_1') \models \varphi$. From the clause Zig$_\blacklozenge$ of a p-bisimulation we know that there exists $(\mathcal{M}_2', w_2')$ such that $(\mathcal{M}_2, w_2) \overset{\bullet}{\rightarrow} (\mathcal{M}_2', w_2')$ and $(\mathcal{M}_1, w_1')Z(\mathcal{M}_2, w_2')$. By the induction hypothesis, we have $(\mathcal{M}_1', w_1') \overset{\text{p}}{\rightsquigarrow} (\mathcal{M}_2', w_2')$ which brings that $(\mathcal{M}_2', w_2') \models \varphi$, from which we conclude $(\mathcal{M}_2, w_2) \models \blacklozenge\varphi$. The direction from $(\mathcal{M}_2, w_2) \models \blacklozenge\varphi$ to $(\mathcal{M}_1, w_1) \models \blacklozenge\varphi$ is similar and uses the Zag$_\blacklozenge$ condition. □

Remark 3 (Credulous admissibility and p-bisimulation)**.** *Formula* (10) *expresses the existence of credulous admissible arguments (Remark 1), and is invariant for p-bisimulation (Theorem 4.5). It directly follows that, given two p-bisimilar pointed models $(\mathcal{M}_1, w_1)$ and $(\mathcal{M}_2, w_2)$, the frame of $\mathcal{M}_1$ contains credulously admissible arguments if and only if the frame of $\mathcal{M}_2$ does.*

For the converse result, some auxiliary definitions are needed.[5] Let $\mathcal{M} = (W, R, V)$ be a model. A set of FOL formulas $\Gamma(x)$ from $\mathcal{L}$ with one free variable x is realized by $\mathcal{M}$ if there exists $w \in W$ s.t. $\mathcal{M} \models \Gamma(x)[x := w]$. We say that $\mathcal{M}$ (viewed as a FOL structure) is ω-saturated if for any finite set $X \subseteq W$, the expansion $\mathcal{M}_X$ realizes every set $\Gamma(x) \in \mathcal{L}_X$ (i.e., the expansion of $\mathcal{L}$ with constants for the elements in X) whenever every finite subset $\Gamma'(x) \subseteq \Gamma(x)$ is realized in $\mathcal{M}_X$.

Theorem 4.6. *For any two ω-saturated models $(\mathcal{M}_1, w_1)$ and $(\mathcal{M}_2, w_2)$, if $(\mathcal{M}_1, w_1) \overset{\text{p}}{\rightsquigarrow} (\mathcal{M}_2, w_2)$ then $(\mathcal{M}_1, w_1) \overset{\text{p}}{\rightleftharpoons} (\mathcal{M}_2, w_2)$.*

Sketch of proof. We show that $\overset{\text{p}}{\rightsquigarrow}$ is itself a p-bisimulation. The base case holds trivially. The proof for the Zig$_\Diamond$ and Zag$_\Diamond$ proceed in the usual manner. We need to prove that the conditions Zig$_\blacklozenge$ and Zag$_\blacklozenge$ are verified. Zig$_\blacklozenge$ Let us assume that $(\mathcal{M}_1, w_1) \overset{\text{p}}{\rightsquigarrow} (\mathcal{M}_2, w_2)$ and that there exists a pointed model $(\mathcal{M}_1', w_1')$ such that $(\mathcal{M}_1, w_1) \overset{\bullet}{\rightarrow} (\mathcal{M}_1', w_1')$. We show that there exists $(\mathcal{M}_2', w_2')$ such that $(\mathcal{M}_2, w_2) \overset{\bullet}{\rightarrow} (\mathcal{M}_2', w_2')$ and $(\mathcal{M}_1', w_1') \overset{\text{p}}{\rightsquigarrow} (\mathcal{M}_2', w_2')$. First, observe that for any finite $\Gamma \subseteq \mathbb{T}^{\text{p}}(\mathcal{M}_1', w_1')$, by Theorem 4.3, the following equivalences hold:

$$\begin{aligned}(\mathcal{M}_1, w_1) \models \blacklozenge \bigwedge \Gamma &\Leftrightarrow (\mathcal{M}_2, w_2) \models \blacklozenge \bigwedge \Gamma \\ &\Leftrightarrow \mathcal{M}_2 \models ST_x^{\emptyset}\left(\blacklozenge \bigwedge \Gamma\right)[x := w_2] \\ &\Leftrightarrow \mathcal{M}_2 \models \exists y \left(xR^{\mathcal{M}_2}y \wedge ST_y^{\{y\}}\left(\bigwedge \Gamma\right)\right)[x := w_2].\end{aligned}$$

Since $\mathcal{M}_2$ is ω-saturated by assumption, we have:

$$\exists y \in \mathcal{M}_2, \mathcal{M}_2 \models ST_y^{\{y\}}\left(\mathbb{T}^p(\mathcal{M}_1', w_1')\right).$$

By Theorem 4.3, there exists a pointed model $(\mathcal{M}_2', w_2')$ such that $(\mathcal{M}_2, w_2) \overset{\bullet}{\rightarrow} (\mathcal{M}_2', w_2')$ and $\mathcal{M}_2' \models ST_x^{\emptyset}\left(\mathbb{T}^{\text{p}}(\mathcal{M}_1', w_1')\right)[x := w_2']$. Again by Theorem 4.3, it follows that $(\mathcal{M}_1', w_1') \overset{\text{p}}{\rightsquigarrow} (\mathcal{M}_2', w_2')$. Zag$_\blacklozenge$ The proof is by a similar argument. □

Theorem 4.7. *An $\mathcal{L}$ formula is equivalent to the translation of an $\mathcal{L}^{\text{p}}$ formula if and only if it is invariant for p-bisimulation.*

Proof. Left-to-right By Theorem 4.5. Right-to-left Let $\varphi \in \mathcal{L}$ be a formula with only one free variable x. Let us assume that φ is invariant under p-bisimulation and let C be the following set: $C(\varphi) = \{ST_x^{\emptyset}(\psi) \mid \psi \in \mathcal{L}^{\text{p}} \text{ and } \varphi \models ST_x^{\emptyset}(\psi)\}$. Let us now show that $C(\varphi) \models \varphi$, that is to say: for any pointed model $(\mathcal{M}, w)$ if $\mathcal{M} \models C(\varphi)[x := w]$ then $\mathcal{M} \models \varphi[x := w]$. To do so, let us first show that $\Sigma = ST_x^{\emptyset}\left(\mathbb{T}^{\text{p}}(\mathcal{M}, w)\right) \cup \{\varphi\}$ is consistent.

Let us assume for the sake of the contradiction that Σ is inconsistent. By the compactness of FOL, $\models \varphi \rightarrow \neg \bigwedge \Gamma$ for some finite $\Gamma \subseteq ST_x^{\emptyset}\left(\mathbb{T}^{\text{p}}(\mathcal{M}, w)\right)$. By the definition of $C(\varphi)$ this means that $\neg \bigwedge \Gamma \in C(\varphi)$ and so $\neg \bigwedge \Gamma \in ST_x^{\emptyset}\left(\mathbb{T}^{\text{p}}(\mathcal{M}, w)\right)$ which is in contradiction with $\Gamma \subseteq ST_x^{\emptyset}\left(\mathbb{T}^{\text{p}}(\mathcal{M}, w)\right)$.

Let us now show that $\mathcal{M} \models \varphi[x := w]$. As Σ is consistent, there exists a pointed model $(\mathcal{M}', w')$ such that $(\mathcal{M}', w') \models \Sigma$. From this it is immediate to see that $(\mathcal{M}, w) \overset{\text{p}}{\rightsquigarrow} (\mathcal{M}', w')$. Let us now consider two ω-saturated elementary extensions $(\mathcal{M}_\omega, w)$

[5] Cf. [12, Ch. 2].

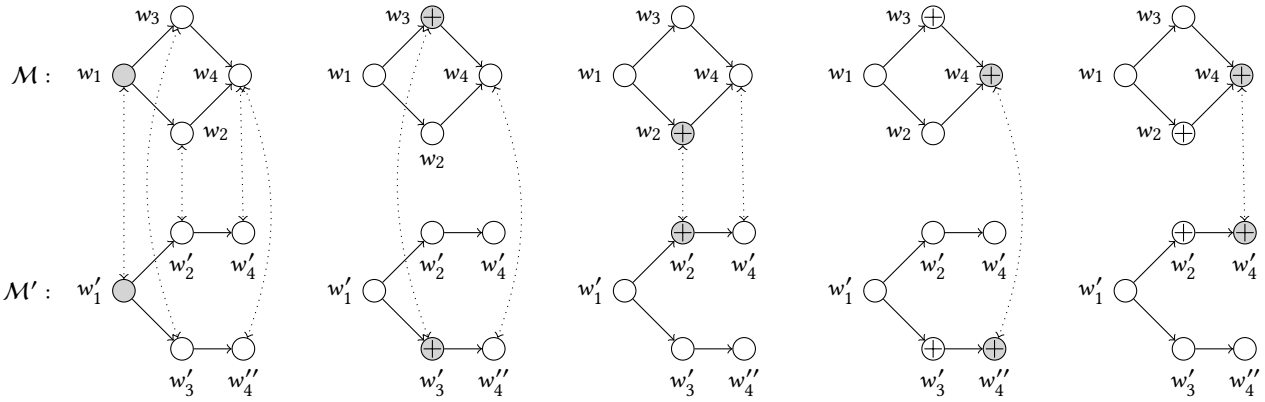

Figure 2: Two p-bisimilar models $\mathcal{M}$ and $\mathcal{M}'$ (leftmost models), and the models reachable via poisoning from them (crosses denote poisoned states). Dotted lines represent links of the p-bisimulation instantiating conditions of Definition 4.4.

and $(\mathcal{M}'_\omega, w')$ of $(\mathcal{M}, w)$ and $(\mathcal{M}', w')$. Such extensions exist by standard argument (cf. [14, Proposition 3.2.6]). As FOL is invariant under elementary extensions, from $\mathcal{M}' \models \varphi[x := w]$ we can conclude that $\mathcal{M}'_\omega \models \varphi[x := w]$. As we have assumed that φ is invariant under p-bisimulation and thanks to Theorem 4.6, we get $\mathcal{M}_\omega \models \varphi[x := w]$ which brings $\mathcal{M} \models \varphi[x := w]$. We have thus shown that $C(\varphi) \models \varphi$.

To conclude the proof, we need to show that $C(\varphi) \models \varphi$ implies that φ is equivalent to the translation of an $\mathcal{L}^{\mathfrak{p}}$ formula. As $C(\varphi) \models \varphi$, from the deduction and the compactness theorems of FOL, $\models \bigwedge \Gamma \to \varphi$ for some finite $\Gamma \subset C(\varphi)$. By definition of $C(\varphi)$ we also have $\models \varphi \to \bigwedge \Gamma$ and so $\models \varphi \leftrightarrow \bigwedge \Gamma$. □

Finally, it is worth mentioning a simple example of a property which is not expressible in PML. Properties of this type are for instance those involving counting quantifiers, e.g.: the current state has at least n successors. It is easy to devise two p-bisimilar models for which the above property holds for one, but not the other.

5 UNDECIDABILITY

In this section we tackle the question of the decidability of PML.

5.1 Undecidability of PML_3

In this section we establish the undecidability of PML_3, that is the variant of PML with three standard modalities and three poison modalities, each with a distinct poison atom. We call R, R_1 and R_2 the three accessibility relations of a model of PML_3. The satisfaction problem for PML_3 can be defined as follows:

Data: A PML_3 formula $\varphi \in \mathcal{L}^{\mathfrak{p}}_3$.

Problem: Is there $(\mathcal{M}, w)$, with $\mathcal{M} \in \mathfrak{M}^{\emptyset}$, such that $(\mathcal{M}, w) \models \varphi$?

Theorem 5.1. *The satisfaction problem for* PML_3 *is undecidable.*

Sketch of proof. We reduce the problem of the $\mathbb{N} \times \mathbb{N}$ tilling in a similar way as for the proof of undecidability of hybrid logic $\mathcal{H}(\downarrow)$ presented in [34]. Let us recall $\mathbb{N} \times \mathbb{N}$ tilling problem. Given a finite set of colors C, a tile is a 4-tuple of colors (its 4 sides). The $\mathbb{N} \times \mathbb{N}$ tilling problem is then defined as follows:

Data: A finite set T of tiles.

Problem: Can the infinite grid $\mathbb{N} \times \mathbb{N}$ be tiled using only tiles in T and such that two adjacent tiles share the same color on their common edge ?

This problem is known to be undecidable [24]. So let T be a finite set of tiles. We claim that the following formula is satisfiable if and only if T tiles the grid $\mathbb{N} \times \mathbb{N}$:

$$\varphi_T = \alpha \wedge \beta \wedge \gamma \wedge \Box \left(\delta^1_T \wedge \delta^2_T \wedge \delta^3_T \right) \tag{11}$$

with $\alpha, \beta, \gamma, \delta^1_T, \delta^2_T$ and δ^3_T such as:

$$\alpha = q \wedge \Box(\neg q \wedge \Diamond q) \wedge \Box\blacksquare_1 \Diamond(q \wedge \Diamond \mathfrak{p}) \wedge \Box\blacksquare_2 \Diamond(q \wedge \Diamond \mathfrak{p})$$

$$\beta = \bigwedge_{i=1,2} (\Box\Diamond_i \top \wedge \blacksquare\Box(q \to \Box(\Diamond_i \mathfrak{p} \to \Box_i \mathfrak{p})))$$

$$\gamma = \blacksquare\Box\,(q \to \Box(\Box_1\Box_2\neg\mathfrak{p} \vee \Box_2\Box_1\mathfrak{p}))$$

$$\delta^1_T = \bigvee_{t \in T} \left(p_t \wedge \bigwedge_{t' \in T, t' \neq t} \neg p_{t'} \right)$$

$$\delta^2_T = \bigwedge_{t \in T} \left(p_t \to \Box_2 \bigvee_{t' \in T, \mathit{left}(t') = \mathit{right}(t)} p_t \right)$$

$$\delta^3_T = \bigwedge_{t \in T} \left(p_t \to \Box_1 \bigvee_{t' \in T, \mathit{bottom}(t') = \mathit{top}(t)} p_t \right).$$

In these formulas, the modality $\Diamond_1$ represent vertical moves on the grid, the modality $\Diamond_2$ horizontal moves and the modality $\Diamond$ moves from any point to any point on the grid (i.e., a universal modality). For a tile $t \in T$, the predicate p_t models the fact that t is placed on the point and the four predicates $\mathit{top}(t)$, $\mathit{right}(t)$, $\mathit{bottom}(t)$ and $\mathit{left}(t)$ represent the four colors of t. For $(\mathcal{M}, w) \models \varphi_T$, the subformulas of φ_T are interpreted as follow:

- α: w is a q-world, its R-successors are not q and link back to it, and the set of its R-successors is closed under R_1 and R_2.
- β: for all R-successor of w, accessibility relations R_1 and R_2 are total functions.
- γ: accessibility relations R_1 and R_2 commute.
- $\Box(\delta^1_T \wedge \delta^2_T \wedge \delta^3_T)$: only one tile is present at each node and horizontal and vertical tiling are correct.

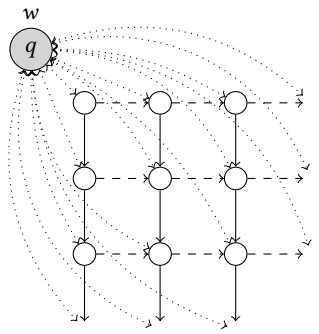

Figure 3: A model of formula φ_T. R is represented by dotted links, R_1 by dashed link and R_2 by plain links.

By the construction of φ_T, it is easy to see that φ_T is satisfiable if and only if there is a tiling of the grid $\mathbb{N} \times \mathbb{N}$ with T (Figure 3). □

5.2 Failure of FMP for PML

It is unclear however whether PML *with only* one standard and one poison modality is also undecidable. We suspect it is, and we can show it fails to have the finite model property.

Proposition 5.2. PML *does not have the finite model property.*

Proof. We provide a formula whose models are all infinite. Let us consider $\varphi = \alpha \wedge \beta \wedge \gamma \wedge \delta \wedge \epsilon$ with the sub-formulas defined below.

- $\alpha = \neg q \wedge \Diamond \top \wedge \Box q \wedge \Box(\Diamond \top \wedge \Box \neg q)$: the current state falsifies q and all its successors (there exists at least one) are q and have in turn successors (at least one) which all falsify q.
- $\beta = \blacksquare\Box\Diamond \mathfrak{p}$: after any poisoning a state is reached whose successors can reach the poisoned state in one step. In other words, all successors of the current state have successors linked via symmetric edges.
- $\gamma = \blacksquare\Box\Diamond(\neg q \wedge \Diamond \mathfrak{p}) \wedge \Box\Box\neg\blacklozenge\Diamond \mathfrak{p}$: after any poisoning a state is reached whose successors are not reflexive loops (right conjunct), and can reach a $\neg q$ state which can in turn reach the poison state. In other words, all successors of the current state lay on cycles of length 3.
- $\delta = \Box\Box\blacksquare\Box(q \rightarrow \Diamond \mathfrak{p})$: all successors of the current state's successors are such that after any poisoning, and further q-successor can reach back to the poisoned state.
- $\epsilon = \Box\blacklozenge\neg\Diamond(q \wedge \Diamond(\neg q \wedge \Diamond \mathfrak{p}))$: all successors of the current state are such that there is one successor that can be poisoned and such that none of its successors satisfies q and can reach the poisoned state in two steps via a $\neg q$ state.

Now let $(\mathcal{M}, w) \models \varphi$. Then, w is followed by distinct successors w' (α) that have successors w'' which are linked back to their predecessors w' by symmetric edges (β). These w'' states also have successors, different from w' which also have w' as successor (γ) and which are also successors of w' (δ). Hence, w'' is followed by an infinite path of distinct states. Finally, there exists one such w'' which has no other predecessor than w' (ϵ), that is, w'' is the root of an infinite sequence of distinct states which are all successors of w'. One such model is depicted in Figure 4. □

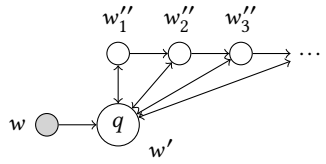

Figure 4: An infinite model of φ from Proposition 5.2.

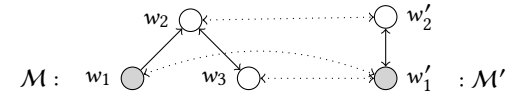

Figure 5: Two p-bisimilar models from Proposition 6.1.

6 DISCUSSION & CONCLUSIONS

In this last section we relate PML with memory and hybrid logics establishing results about their relative expressivity (Propositions 6.1 and 6.3). Such results are useful to position PML precisely within the landscape of existing extensions of the standard modal language. We then conclude by charting a few lines of future research.

6.1 PML and Memory Logics

As hinted at in the introduction, PML is tightly related to memory logics [3, 4]. The simplest memory logic, $\mathcal{M}(\textcircled{r}, \textcircled{k})$, extends modal semantics by considering frames (W, R, M) where $M \subseteq W$ is a set of states that have been 'memorized'. The standard modal language is then extended with two operators $\textcircled{r}$ and $\textcircled{k}$ defined as:

$$((W, R, M, V), w) \models \textcircled{r}\varphi \iff ((W, R, M \cup \{w\}, V), w) \models \varphi$$
$$((W, R, M, V), w) \models \textcircled{k} \iff w \in M,$$

where V is a valuation function. Intuitively, the $\textcircled{r}$ stores the current state in the memory M, serving a similar purpose to our poisoning operation, and the nullary operator $\textcircled{k}$ works precisely as our atom $\mathfrak{p}$. Intuitively, PML can be seen as a memory logic in which storing states via $\textcircled{r}$ occurs only after traversing an edge in the underlying frame. Technically, PML is a proper fragment of $\mathcal{M}(\textcircled{r}, \textcircled{k})$:

Proposition 6.1. $\mathcal{M}(\textcircled{r}, \textcircled{k})$ *is strictly more expressive than* PML.

Proof. First of all observe that PML models and $\mathcal{M}(\textcircled{r}, \textcircled{k})$ are exactly the same type of structures, where M of the latter type of models corresponds to the truth-set of $\mathfrak{p}$ in the former type of models. We show that $\mathcal{M}(\textcircled{r}, \textcircled{k})$ is at least as expressive as PML, by providing a truth-preserving embedding of the latter into the former. Such an embedding is provided by the following translation (clauses for Boolean connectives and $\Diamond$ are omitted as straightfoward):

$$MT(\mathfrak{p}) = \textcircled{k}$$
$$MT(\blacklozenge\varphi) = \Diamond\textcircled{r}MT(\varphi)$$

It is easy to see that such translation is truth-preserving.

To show that PML is strictly less expressive than $\mathcal{M}(\textcircled{r}, \textcircled{k})$ it suffices to provide two p-bisimilar pointed models which can be distinguished by a formula of $\mathcal{M}(\textcircled{r}, \textcircled{k})$. Such models are depicted in Figure 5. The model on the right satisfies formula $\textcircled{r}\Diamond\Diamond\textcircled{k}$ while the model on the left falsifies it. □

6.2 PML and Hybrid Logics

PML has, perhaps unsurprisingly, also tight links with hybrid logics. We show how PML can be embedded into $\mathcal{H}(\downarrow)$ defined by [34]:

$$\mathcal{L}^{\mathcal{H}(\downarrow)} : \varphi := p \mid i \mid \neg\varphi \mid \varphi \wedge \varphi \mid \ \Diamond\varphi \mid \downarrow x.\varphi,$$

with $p \in \mathbf{P} \cup \{\mathfrak{p}\}$ a propositional atom, and $i \in \mathbf{N}$ a nominal. We write $\models_{\mathbf{H}}$ the satisfaction relation for $\mathcal{H}(\downarrow)$. Given an assignment $g : \mathbf{N} \to W$, g^x_m is called a x-variant of g if $\forall i \in \mathbf{N}, g(i) = g^x_m(i)$ and $g^x_m(x) = m$. The semantics is then $(M, g, m) \models_{\mathbf{H}} i \Leftrightarrow m = g(i)$ and $(M, g, m) \models_{\mathbf{H}} \downarrow x.\varphi \Leftrightarrow (M, g^x_m, m) \models_{\mathbf{H}} \varphi$. We can then set up the translation $HT^S : \mathcal{L}^{\mathfrak{p}} \to \mathcal{L}^{\mathcal{H}(\downarrow)}$ as follows, with $S \subseteq \mathbf{N}$:

$$\begin{aligned}
HT^S(p) &= p \\
HT^S(\mathfrak{p}) &= \mathfrak{p} \vee \bigvee_{i \in N} i \\
HT^S(\neg\varphi) &= \neg HT^S(\varphi) \\
HT^S(\varphi \wedge \psi) &= HT^S(\varphi) \wedge HT^S(\psi) \\
HT^S(\Diamond\varphi) &= \Diamond HT^S(\varphi) \\
HT^S(\blacklozenge\varphi) &= \Diamond\left(\downarrow x.HT^{S\cup\{x\}}(\varphi)\right),
\end{aligned}$$

with x a "fresh variable" never used before. We also need a way to transform PML models into hybrid models. Let $\mathcal{M} = (W, R, V)$ be a PML-model, we define $M = (W, R, V')$, the hybrid extension of $\mathcal{M}$, by extending the valuation V so that $\forall p \in \mathbf{P} \cup \{\mathfrak{p}\}, V'(p) = V(p)$, $\forall w \in W, \exists i \in \mathbf{N}, V'(i) = \{w\}$ and $\forall i \in \mathbf{N}, |V'(i)| = 1$. We can now proceed to show the translation HT^S is correct.

Lemma 6.2. *Let $\mathcal{M} = (W, R, V)$ be a* PML*-model and $M = (W, R, V')$ its hybrid extension. Let us consider $v, w \in W$ and g an assignment. Then for $\varphi \in \mathcal{L}^{\mathfrak{p}}$ a* PML*-formula and any set S, we have:*

$$(M^\bullet_v, g, w) \models_{\mathbf{H}} HT^S(\varphi) \Leftrightarrow (M, g^x_v, w) \models_{\mathbf{H}} HT^{S\cup\{x\}}(\varphi).$$

Sketch of proof. We show this result by induction on the structure of φ. The proof is trivial for the propositional case and the Boolean connectives, hence we only present cases for the poison atom and the poison modality.

$\boxed{\varphi = \mathfrak{p}}$ The claim is proven by the following series of equivalences using the definition of the poison operation $\bullet$:

$$\begin{aligned}
(M^\bullet_v, g, w) \models_{\mathbf{H}} HT^S(\mathfrak{p}) &\Leftrightarrow (M^\bullet_v, g, w) \models_{\mathbf{H}} \mathfrak{p} \vee \bigvee_{i \in S} i \\
&\Leftrightarrow (M, g, w) \models_{\mathbf{H}} \mathfrak{p} \vee \bigvee_{i \in S\cup\{x\}} i \\
&\Leftrightarrow (M, g^x_v, w) \models_{\mathbf{H}} HT^{S\cup\{x\}}(\mathfrak{p}).
\end{aligned}$$

$\boxed{\varphi = \blacklozenge\psi}$ with $\psi \in \mathcal{L}^{\mathfrak{p}}$ The claim is proven by the following series of equivalences using the definition of $\bullet$, the semantics of $\downarrow$ and the induction hypothesis :

$$\begin{aligned}
&(M^\bullet_v, g, w) \models_{\mathbf{H}} HT^S(\blacklozenge\psi) \\
\Leftrightarrow\ &(M^\bullet_v, g, w) \models_{\mathbf{H}} \Diamond\left(\downarrow y.HT^{S\cup\{y\}}(\psi)\right) \\
\Leftrightarrow\ &\exists u \in W, wRu, (M^\bullet_v, g, u) \models_{\mathbf{H}} \downarrow y.HT^{S\cup\{y\}}(\psi) \\
\Leftrightarrow\ &\exists u \in W, wRu, (M^\bullet_v, g^y_u, u) \models_{\mathbf{H}} HT^{S\cup\{y\}}(\psi) \\
\Leftrightarrow\ &\exists u \in W, wRu, (M, (g^y_u)^x_v, u) \models_{\mathbf{H}} HT^{S\cup\{x,y\}}(\psi) \\
\Leftrightarrow\ &\exists u \in W, wRu, (M, g^x_v, u) \models_{\mathbf{H}} \downarrow y.HT^{S\cup\{x,y\}}(\psi) \\
\Leftrightarrow\ &(M, g^x_v, w) \models_{\mathbf{H}} \Diamond\left(\downarrow y.HT^{S\cup\{x,y\}}(\psi)\right) \\
\Leftrightarrow\ &(M, g^x_v, w) \models_{\mathbf{H}} HT^{S\cup\{x,y\}}(\blacklozenge\psi).
\end{aligned}$$

This completes the proof. □

Proposition 6.3. *Let $\mathcal{M} = (W, R, V)$ be a* PML*-model, $M = (W, R, V')$ its hybrid extension, g an assignment and $\varphi \in \mathcal{L}^{\mathfrak{p}}$ a* PML*-formula, we have:*

$$(\mathcal{M}, w) \models \varphi \Longleftrightarrow (M, g, w) \models_{\mathbf{H}} HT^{\emptyset}(\varphi).$$

Sketch of proof. The proof is done by induction on the structure of φ. We only present the non classical cases.

$\boxed{\varphi = \mathfrak{p}}$ The claim is proven by the following series of equivalences using the defintion of a hybrid extension:

$$\begin{aligned}
(\mathcal{M}, w) \models \mathfrak{p} &\Leftrightarrow w \in V(\mathfrak{p}) \\
&\Leftrightarrow w \in V'(\mathfrak{p}) \\
&\Leftrightarrow (M, g, w) \models_{\mathbf{H}} \mathfrak{p} \\
&\Leftrightarrow (M, g, w) \models_{\mathbf{H}} HT^{\emptyset}(\mathfrak{p}).
\end{aligned}$$

$\boxed{\varphi = \blacklozenge\psi}$ with $\psi \in \mathcal{L}^{\mathfrak{p}}$ The claim is proven by the following series of equivalences the definition of $\bullet$, the semantics of $\downarrow$, the induction hypothesis and Lemma 6.2:

$$\begin{aligned}
(\mathcal{M}, w) \models \blacklozenge\psi &\Leftrightarrow \exists v \in W, wRv, (\mathcal{M}^\bullet_v, v) \models \psi \\
&\Leftrightarrow \exists v \in W, wRv, (M^\bullet_v, g, v) \models_{\mathbf{H}} HT^{\emptyset}(\psi) \\
&\Leftrightarrow \exists v \in W, wRv, (M, g^x_v, v) \models_{\mathbf{H}} HT^{\{x\}}(\psi) \\
&\Leftrightarrow \exists v \in W, wRv, (M, g, v) \models_{\mathbf{H}} \downarrow x.HT^{\{x\}}(\psi) \\
&\Leftrightarrow (M, g, w) \models_{\mathbf{H}} \Diamond(\downarrow x.HT^{\{x\}}(\psi)) \\
&\Leftrightarrow (M, g, w) \models_{\mathbf{H}} HT^{\emptyset}(\blacklozenge\psi).
\end{aligned}$$

This completes the proof. □

6.3 Conclusions

The paper has introduced and studied a modal logic PML that arises naturally from a game-theoretic approach to a central decision problem in argumentation theory: the existence of credulously admissible sets. Our results provide new links between abstract argumentation theory [18], games on graphs [11] and modal logic [12]. Many directions for future research present themselves, at several levels. From the logic point of view, several technical problems remain open concerning PML: Can the logic be axiomatised via a Hilbert calculus, possibly using insights from the memory logic literature (e.g., [5])? Can PML be embedded in a fixed-variable fragment of FOL (thereby shedding light on the precise level of saturation that would suffice for Theorem 4.6)? Is PML (with one modal operator and one sabotage operator) decidable? Can the logic be extended in a natural way with a least-fixpoint operator, for instance to express formulas (9) and (10)? From the argumentation theory point of view, a natural question is whether the poison game can be adapted to capture the property of existence of skeptically admissible arguments, that is, arguments that belong to *all* admissible sets in a framework.

REFERENCES

[1] C. Areces, R. Fervari, and G. Hoffmann. 2013. Tableaux for Relation-Changing Modal Logics. In *Proceedings of the 9th International Symposium on Frontiers of Combining Systems (FroCoS'13) (LNAI)*, P. Fontaine, C. Ringeissen, and R. Schmidt (Eds.), Vol. 8152. 263–278.

[2] C. Areces, R. Fervari, and G. Hoffmann. 2015. Relation-Changing Modal Operators. *Logic Journal of the IGPL* (2015).

[3] C. Areces, D. Figueira, S. Figueira, and S. Mera. 2011. The Expressive Power of Memory Logics. *Review of Symbolic Logic* 4, 2 (2011), 290–218.

[4] C. Areces, D. Figueira, and S. Mera. 2008. Expressive power and decidability for memory logics. In *Proceedings of WoLLIC 2008 (LNCS)*, Vol. 5110. 56–68.

[5] C. Areces, D. Figueira, and S. Mera. 2009. Completeness results for memory logics. In *Proceedings of LFCS'09 (LNCS)*, Vol. 5407. 16–30.

[6] G. Aucher, P. Balbiani, L. F. Del Cerro, and A. Herzig. 2009. Global and local graph modifiers. *Electronic Notes in Theoretical Computer Science* 231 (2009), 293–307.

[7] G. Aucher, J. van Benthem, and D. Grossi. 2017. Modal logics of sabotage revisited. *Journal of Logic and Computation* 28, 2 (2017), 269–303.

[8] P. Baroni, M. Caminada, and M. Giacomin. 2011. An Introduction to Argumentation Semantics. *The Knowledge Engineering Review* 26, 4 (2011), 365–410.

[9] P. Baroni and M. Giacomin. 2009. Semantics of Abstract Argument Systems. In *Argumentation in Artifical Intelligence*, I. Rahwan and G. R. Simari (Eds.). Springer.

[10] T. Bench-Capon and P. Dunne. 2007. Argumentation in Artificial Intelligence. *Artificial Intelligence* 171, 10 (2007), 619–641.

[11] C. Berge. 1996. Combinatorial Games on Graph. *Discrete Mathematics* 151 (1996), 59–65.

[12] P. Blackburn, M. de Rijke, and Y. Venema. 2001. *Modal Logic.* Cambridge University Press.

[13] P. Blackburn, J. van Benthem, and F. Wolter. 2006. *Handbook of modal logic.* Vol. 3. Elsevier.

[14] C. C. Chang and H. J. Keisler. 1973. *Model Theory.* North-Holland.

[15] Y. Dimopulos and V. Magirou. 1994. A Graph-Theoretic Approach to Default Logic. *Information and Computation* 112 (1994), 239–256.

[16] Y. Dimopulos and A. Torres. 1996. Graph theoretical structures in logic programs and default theories. *Theoretical Computer Science* 170 (1996), 209–244.

[17] P. Duchet and H. Meyniel. 1993. Kernels in directed graphs: a poison game. *Discrete mathematics* 115, 1-3 (1993), 273–276.

[18] P. M. Dung. 1995. On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-Person Games. *Artificial Intelligence* 77, 2 (1995), 321–358.

[19] D. Gabbay and D. Grossi. 2014. When are two arguments the same? Equivalence in abstract argumentation. In *Johan van Benthem on Logic and Information Dynamics*, A. Baltag and S. Smets (Eds.). Springer.

[20] H. Galeana-Sánchez and V. Neumann-Lara. 1984. On Kernels and Semikernels of Digraphs. *Discrete Mathematics* 48 (1984), 67–76.

[21] D. Grossi. 2010. On the Logic of Argumentation Theory. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, W. van der Hoek, G. Kaminka, Y. Lespérance, and S. Sen (Eds.). IFAAMAS, 409–416.

[22] D. Grossi. 2011. Argumentation Theory in the View of Modal Logic. In *Post-proceedings of the 7th International Workshop on Argumentation in Multi-Agent Systems (LNAI)*, P. McBurney and I. Rahwan (Eds.). 190–208.

[23] D. Grossi and W. Van der Hoek. 2014. Justified Beliefs by Justified Arguments. In *Proceedings of KR'14.*

[24] D. Harel. 1983. Recurring dominoes: Making the highly undecidable highly understandable (preliminary report). In *International Conference on Fundamentals of Computation Theory.* Springer, 177–194.

[25] B. Kooi and B. Renne. 2011. Arrow update logic. *The Review of Symbolic Logic* 4, 4 (2011), 536–559.

[26] C. Löding and P. Rohde. 2003. Model Checking and Satisfiability for Sabotage Modal Logic. In *FSTTCS 2003 (LNCS)*, P. K. Pandya and J. Radhakrishnan (Eds.), Vol. 2914. Springer, 302–313.

[27] C. Mierzewski and F. Zaffora Blando. [n. d.]. The Modal Logic(s) of Poison Games.

[28] S. Modgil and M. Caminada. 2009. Proof Theories and Algorithms for Abstract Argumentation Frameworks. In *Argumentation in AI*, I. Rahwan and G. Simari (Eds.). Springer, 105–132.

[29] M. J. Osborne and A. Rubinstein. 1994. *A Course in Game Theory.* MIT Press.

[30] T. Przymusinski. 1990. Extended Stable Semantics for Normal and Disjunctive Programs. In *Proceedings of the 7th International Conference on Logic Programming.* MIT Press.

[31] P. Rohde. 2006. On the mu-Calculus Augmented with Sabotage. In *Foundations of Software Science and Computation Structures, 7th International Conference, FOSSACS 2004, Held as Part of the Joint European Conferences on Theory and Practice of Software, ETAPS 2004, Barcelona, Spain, March 29 - April 2, 2004, Proceedings (LNCS)*, Vol. 3921. 142–156.

[32] C. Shi, S. Smets, and F. Velázquez-Quesada. 2017. Argument-Based Belief in Topological Structures. In *Proceedings of TARK'17 (EPTCS)*, Vol. 251. 489–503.

[33] C. Shi, S. Smets, and F. Velázquez-Quesada. 2018. Beliefs Supported by Binary Arguments. *Journal of Applied Non-Classical Logic* (2018), 1–24.

[34] B. Ten Cate and M. Franceschet. 2005. On the complexity of hybrid logics with binders. In *International Workshop on Computer Science Logic.* Springer, 339–354.

[35] J. van Benthem. 1983. *Modal Logic and Classical Logic.* Bibliopolis.

[36] J. Van Benthem. 2005. An essay on sabotage and obstruction. In *Mechanizing Mathematical Reasoning.* Springer, 268–276.

[37] J. van Benthem. 2011. *Logical Dynamics of Information and Interaction.* Cambridge University Press.

[38] J. van Benthem. 2014. *Logic in games.* MIT press.

[39] J. van Benthem and A. Gheerbrant. 2010. Game Solution, Epistemic Dynamics and Fixed-Point Logics. *Fundamenta Informaticae* 1, 4 (2010), 19–41.

[40] H. van Ditmarsch, W. van Der Hoek, and B. Kooi. 2007. *Dynamic epistemic logic.* Vol. 337. Springer Science & Business Media.

[41] G. Vreeswijk and H. Prakken. 2000. Credulous and Sceptical Argument Games for Preferred Semantics. In *Proceedings of the 7th European Workshop on Logic for Artificial Intelligence (JELIA'00) (LNAI).* Springer, 239–253.