# Credulous Acceptability, Poison Games and Modal Logic (*extended abstract*)

Davide Grossi[1] and Simon Rey[2]*

[1] University of Groningen, Groningen, Netherland
d.grossi@rug.nl
[2] ENS Paris-Saclay, Cachan, France
srey@ens-paris-saclay.fr

## 1 Introduction

In abstract argumentation theory [3], an argumentation framework is a directed graph $(A, \rightarrow)$ [6]. For $x, y \in A$ such that $x \rightarrow y$ we say that $x$ attacks $y$. An *admissible* set, of a given attack graph, is a set $X \subseteq A$ such that [6]: (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 $y$. Such sets are also called *credulously admissible* argument. They form the basis of all main argumentation semantics first developed in [6], and they are central to the influential graph-theoretic systematization of logic programming and default reasoning pursued in [4].

One key reasoning tasks is then to decide whether a given argumentation framework contains at least one non-empty admissible set [7]. Interestingly, the notion has an elegant operationalization in the form of a two-player game, called *Poison Game* [5], or *game for credulous acceptance* [11, 16]. 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 express the modal reasoning involved in the notion of credulous admissibility. This explores research lines presented in [9]. The paper also defines a suitable notion of p-bisimulation, which answers another open question [8], namely a notion of structural equivalence tailored for it. More broadly we see the present paper as a contribution to bridging concepts from abstract argumentation theory, games on graphs and modal logic.

This paper is a natural continuation of the line of work interfacing abstract argumentation and modal logic. PML sits at the intersection of two lines of research in modal logic: dynamic logic concerned with the study of operators which transform semantics structures [1, 13, 15]; and game logics analyzing games through logic [2, 14]. To the best of our knowledge, only [10] (private communication) presented a preliminary work on a modal logic inspired by the Poison Game.

## 2 Poison Modal Logic (PML)

### 2.1 The Poison Game

The Poison Game [5] is a two-player ($\mathbb{P}$, the proponent, and $\mathbb{O}$, the opponent), win-lose, perfect-information game played on a directed graph $(W, R)$. The game starts by $\mathbb{P}$ selecting a node $w_0 \in W$. After this initial choice, $\mathbb{O}$ selects $w_1$ a successor of $w_0$, $\mathbb{P}$ then selects a successor $w_1$ and so on. However, while $\mathbb{O}$ can choose any successor of the current node, $\mathbb{P}$ can select only 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. What makes this game interesting for us is that the existence of a winning strategy for $\mathbb{P}$, if $(W, R)$ is finite, can be shown to be equivalent to the existence of a (non-empty) *credulously admissible* argument in the graph [5].

## 2.2 Syntax and semantics

The poison modal language $\mathcal{L}^{\mathfrak{p}}$ is defined by the following grammar in 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 touch on 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 $\mathcal{M} = (W, R, V)$. A pointed model is a pair $(\mathcal{M}, w)$ with $w \in \mathcal{M}$. We will call $\mathfrak{M}$ the set of all pointed models and $\mathfrak{M}^{\emptyset}$ the set of pointed models $(\mathcal{M}, w)$ such that $V^{\mathcal{M}}(\mathfrak{p}) = \emptyset$. We define now an operation $\bullet$ on models which modifies valuation $V$ by adding a specific 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 equiped to describe the semantics for the $\blacklozenge$ modality (the other clauses are standard):

$$(\mathcal{M}, w) \models \blacklozenge\varphi \iff \exists v \in W, wRv, (\mathcal{M}^{\bullet}_v, v) \models \varphi.$$

We introduce some auxiliary definitions. The poisoning relation between two pointed models $\overset{\bullet}{\to} \in \mathfrak{M}^2$ is defined as: $(\mathcal{M}, w) \overset{\bullet}{\to} (\mathcal{M}', w') \iff R^{\mathcal{M}}(w, w')$ and $\mathcal{M}' = \mathcal{M}^{\bullet}_{w'}$. Furthermore, we denote $(\mathcal{M}, w)^{\bullet} \subset \mathfrak{M}$ the set of all pointed models accessible from $\mathcal{M}$ via a poisoning relation. Two pointed models $(\mathcal{M}, w)$ and $(\mathcal{M}', w')$ are poison modally equivalent, written $(\mathcal{M}, w) \overset{p}{\leftrightsquigarrow} (\mathcal{M}', w')$, if and only if, $\forall\varphi \in \mathcal{L}^{\mathfrak{p}}$: $(\mathcal{M}, w) \models \varphi \iff (\mathcal{M}', w') \models \varphi$.

## 2.3 Validities and Expressible Properties

**Fact 1.** *Let $\varphi, \psi \in \mathcal{L}^{\mathfrak{p}}$ be two formulas, then the following formulas are valid in* PML*:*

$$\begin{aligned} &\blacksquare p \leftrightarrow \Box p \\ &\Box\mathfrak{p} \to (\blacksquare\varphi \leftrightarrow \Box\varphi) \\ &\blacksquare(\varphi \to \psi) \to (\blacksquare\varphi \to \blacksquare\psi). \end{aligned}$$

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 $\delta_n$, with $n \in \mathbb{N}_{>0}$, defined inductively as follows, with $i < n$:

$$\begin{aligned} \delta_1 &= \Diamond\mathfrak{p} \\ \delta_{i+1} &= \Diamond(\neg\mathfrak{p} \wedge \delta_i). \end{aligned}$$

**Fact 2.** *Let $\mathcal{M} = (W, R, V) \in \mathfrak{M}^{\emptyset}$, then for $n \in \mathbb{N}_{>0}$ 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 $(W, R)$.*

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.

PML (or, more precisely, its infinitary version) can express winning positions 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

$\mathbb{P}$. It is also the case for 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: $\mathfrak{p} \vee \blacklozenge\Box\mathfrak{p} \vee \blacklozenge\Box\blacklozenge\Box\mathfrak{p} \vee \cdots$. Dually, winning positions for $\mathbb{P}$ are defined by the following infinitary $\mathcal{L}^{\mathfrak{p}}$-formula: $\neg\mathfrak{p} \wedge \Diamond\blacksquare\neg\mathfrak{p} \wedge \Diamond\blacksquare\Diamond\blacksquare\neg\mathfrak{p} \wedge \cdots$.

## 3 Expressivity of PML

**Definition 1** (FOL translation)**.** *Let $p, q, \ldots$ in $\mathbf{P}$ be propositional atoms, then their corresponding first-order predicates are called $P, Q, \ldots$. The predicate for the poison atom $\mathfrak{p}$ is $\mathfrak{P}$. Let $N$ be a (possibly empty) set of variables, and $x$ a designated variable, then the translation $ST_x^N : \mathcal{L}^{\mathfrak{p}} \to \mathcal{L}$ is defined inductively as follows (where $\mathcal{L}$ is the first-order correspondence language):*

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

**Theorem 1.** *Let $(\mathcal{M}, w)$ be a pointed model and $\varphi \in \mathcal{L}^{\mathfrak{p}}$ a formula, we have then:*

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

A relation $Z \subseteq \mathfrak{M} \times \mathfrak{M}$ is a p-bisimulation if, together with the standard clauses for bisimulation:

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

Zag$_\blacklozenge$: as expected.

Invariance under the existence of a p-bisimulation (in symbols, $\overset{p}{\rightleftharpoons}$) can be proven to characterize the fragment of FOL which is equivalent to PML.

**Theorem 2.** *For any two pointed models $(\mathcal{M}_1, w_1)$ and $(\mathcal{M}_2, w_2)$, if $(\mathcal{M}_1, w_1) \overset{p}{\rightleftharpoons} (\mathcal{M}_2, w_2)$ then $(\mathcal{M}_1, w_1) \overset{p}{\leftrightsquigarrow} (\mathcal{M}_2, w_2)$.*

**Theorem 3.** *For any two $\omega$-saturated models $(\mathcal{M}_1, w_1)$ and $(\mathcal{M}_2, w_2)$, if $(\mathcal{M}_1, w_1) \overset{p}{\leftrightsquigarrow} (\mathcal{M}_2, w_2)$ then $(\mathcal{M}_1, w_1) \overset{p}{\rightleftharpoons} (\mathcal{M}_2, w_2)$.*

**Theorem 4.** *A $\mathcal{L}$ formula is equivalent to the translation of an $\mathcal{L}^{\mathfrak{p}}$ formula if and only if it is p-bisimulation invariant.*

## 4 Undecidability

In this section we establish the undecidability of $\mathsf{PML}_3$ that corresponds to $\mathcal{L}_3^{\mathfrak{p}}$. We call $R, R_1$ and $R_2$ the three accessibility relations of a model of $\mathsf{PML}_3$. In this variant we only consider models whose poison valuation is empty.

We show that the satisfaction problem for $\mathsf{PML}_3$ is undecidable. To do so we reduce the problem of the $\mathbb{N} \times \mathbb{N}$ tilling in a similar way as the undecidability proof for hybrid logic presented in [12].

**Theorem 5.** *The satisfaction problem for* $\mathsf{PML}_3$ *is undecidable.*

Based on this result we postulate that $\mathsf{PML}$ is also undecidable, especially since we can show that:

**Theorem 6.** $\mathsf{PML}$ *does not have the Finite Model Property.*

# 5 Conclusion

In this article we presented a modal logic to describe the Poison Game which is thus able to detect credulously admissible arguments. This paper is a first exploration of this logic: we gave a first-order translation, a suitable notion of bisimulation and we proved the undecidability of a variant of $\mathsf{PML}$.

# References


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

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

[3] P. Baroni and M. Giacomin. Semantics of abstract argument systems. In I. Rahwan and G. R. Simari, editors, *Argumentation in Artifical Intelligence*. Springer, 2009.

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

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

[6] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–358, 1995.

[7] P. Dunne and M. Wooldridge. Complexity of abstract argumentation. In I. Rahwan and G. Simari, editors, *Argumentation in Artifical Intelligence*, pages 85–104. Springer, 2009.

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

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

[10] C. Mierzewski and F. Zaffora Blando. The modal logic(s) of poison games.

[11] S. Modgil and M. Caminada. Proof theories and algorithms for abstract argumentation frameworks. In I. Rahwan and G. Simari, editors, *Argumentation in AI*, pages 105–132. Springer, 2009.

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

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

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

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

[16] G. Vreeswijk and H. Prakken. Credulous and sceptical argument games for preferred semantics. In *Proceedings of the 7th European Workshop on Logic for Artificial Intelligence (JELIA'00)*, LNAI, pages 239–253. Springer, 2000.