Credulous Acceptability, Poison Games and Modal Logic

Extended Abstract

Davide Grossi
University of Groningen
Groningen, the Netherlands
d.grossi@rug.nl

Simon Rey
Sorbonne Université, ENS Paris Saclay
Paris, France
srey@ens-paris-saclay.fr

ABSTRACT

The Poison Game is a two-player game in which players alternatively move a token on a graph's nodes and such that one player can influence which edges the other player is able to traverse. It operationalizes the notion of existence of credulously acceptable arguments in an argumentation framework or, equivalently, 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 acceptability 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.

ACM Reference Format:
Davide Grossi and Simon Rey. 2019. Credulous Acceptability, Poison Games and Modal Logic . In *Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), Montreal, Canada, May 13–17, 2019,* IFAAMAS, 3 pages.

1 INTRODUCTION

In abstract argumentation theory [5], an *argumentation framework* (or *attack graph*) is a directed graph $(A, \rightarrow)$ [11]. 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 [11]: (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. Arguments contained in some admissible set are said to be *credulously acceptable* arguments. In the terminology of graph theory, admissible sets are semi-kernels of the inverted attack graph $(A, \rightarrow^{-1})$. They form the basis of all main argumentation semantics first developed in [11], and they are central to the influential graph-theoretic systematization of logic programming and default reasoning pursued in [8, 9, 17].

One key reasoning task in abstract argumentation is then to decide whether a given argumentation framework contains at least one non-empty admissible set [12]. Interestingly, the notion has an elegant operationalization in the form of a two-player game, called *Poison Game* [10], or *game for credulous acceptance* [16, 22]. 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. The paper also defines a notion of poison bisimulation, which answers another open question [13], namely a notion of structural equivalence tailored to

credulous acceptability. 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 initiated in [14]. PML sits at the intersection of two lines of research in modal logic: dynamic logic, concerned with the study of operators which transform semantic structures [3, 19, 21]; and game logics, concerned with the analysis of game structures [4, 20]. To date, only [15] has presented work on a modal logic inspired by the Poison Game, 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.

2 POISON MODAL LOGIC (PML)

2.1 The Poison Game

The Poison Game [10] is a two-player ($\mathbb{P}$ and $\mathbb{O}$) 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 of w_1 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 moves available. This game has the remarkable property that, when (W, R) is finite, $\mathbb{P}$ has a winning-strategy if and only if there exists a non-empty semikernel in the graph [10], and therefore if and only if the attack graph (W, R^{-1}) contains credulously acceptable arguments.

2.2 Syntax and semantics

The language $\mathcal{L}^p$ is defined by the following grammar in BNF:

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

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 use multi-modal variants of the above language, denoted $\mathcal{L}^p_n$, 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 $\mathcal{M} = (W, R, V)$ [6]. 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 set $\mathfrak{M}^\emptyset = \{(\mathcal{M}, w) \in \mathfrak{M} \mid V^{\mathcal{M}}(\mathfrak{p}) = \emptyset\}$. We define now an operation $\bullet$ on models which adds a specific state to $V(\mathfrak{p})$. Formally, for $\mathcal{M} = (W, R, V)$ and $w \in W$, we have:

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

where $\forall p \in \mathbf{P}, V'(p) = V(p)$ and $V'(\mathfrak{p}) = V(\mathfrak{p}) \cup \{w\}$.

We are now equipped to describe the semantics for the ◆ modality (the other clauses are standard):

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

We introduce some auxiliary definitions. The poison relation between pointed models $\rightarrow^\bullet \in \mathfrak{M}^2$ is defined as: $(\mathcal{M}, w) \rightarrow^\bullet (\mathcal{M}', w')$ iff $wR^{\mathcal{M}}w'$ and $\mathcal{M}' = \mathcal{M}_{w'}^\bullet$. We say that $(\mathcal{M}, w)$ and $(\mathcal{M}', w')$ are poison modally equivalent, written $(\mathcal{M}, w) \leftrightsquigarrow (\mathcal{M}', w')$, if and only if, $\forall\varphi \in \mathcal{L}^\mathrm{p}: (\mathcal{M}, w) \models \varphi \Leftrightarrow (\mathcal{M}', w') \models \varphi$.

2.3 Validities and Expressible Properties

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

$$\blacksquare p \leftrightarrow \Box p$$
$$\Box \mathfrak{p} \rightarrow (\blacksquare\varphi \leftrightarrow \Box\varphi)$$
$$\blacksquare(\varphi \rightarrow \psi) \rightarrow (\blacksquare\varphi \rightarrow \blacksquare\psi).$$

To illustrate the expressive power of PML, we show that it is possible to express the existence of cycles [7, p. 4] in the modal frame, a property not expressible in the standard modal language.

Fact 2. *Consider the formula δ_n, with $n \in \mathbb{N}_{>0}$, defined as:*

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

Let $(\mathcal{M}, w) \in \mathfrak{M}^\emptyset$ with $\mathcal{M} = (W, R, V)$, 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 a cycle of length $i \leq n$ in the frame (W, R).[1]

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 $\mathcal{L}^\mathrm{p}$-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 $\mathcal{L}^\mathrm{p}$-formula: $\neg\mathfrak{p} \wedge \Diamond\blacksquare\neg\mathfrak{p} \wedge \Diamond\blacksquare\Diamond\blacksquare\neg\mathfrak{p} \wedge \cdots$.

Remark 1. *By Duchet and Meyniel's theorem [10], these formulas interpreted on the inversion of a framework, express the existence of credulously acceptable arguments.*

3 EXPRESSIVITY OF PML

Definition 3.1 (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 $ST_x^N : \mathcal{L}^\mathrm{p} \rightarrow \mathcal{L}$ is defined inductively as follows (where $\mathcal{L}$ is the first-order correspondence language):

$$ST_x^N(p) = P(x), \forall p \in \mathbf{P}$$
$$ST_x^N(\neg\varphi) = \neg ST_x^N(\varphi)$$

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

Theorem 3.2. *For $(\mathcal{M}, w) \in \mathfrak{M}$ and $\varphi \in \mathcal{L}^\mathrm{p}$ a formula, we have:*

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

Definition 3.3 (p-bisimulation). A relation $Z \subseteq \mathfrak{M}^2$ is a p-bisimulation if, together with the standard clauses for bisimulation [6, p. 64]:

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

Zag$_\blacklozenge$: as expected.

Remark 2. *As argued in [13], bisimulation captures a natural notion of 'similarity' of argumentation frameworks that preserves argumentation-theoretic notions. 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 (resp., 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 (resp., complete, stable or grounded) in the frame of $\mathcal{M}_2$ [14, Th. 6] . How to strengthen bisimulation to preserve the existence of credulously acceptable arguments across frameworks was mentioned as an open question in [13]. The p-bisimulation relation provides an elegant answer.*

4 FURTHER RESULTS

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

Theorem 4.1. *For any two pointed models $(\mathcal{M}_1, w_1)$ and $(\mathcal{M}_2, w_2)$, if $(\mathcal{M}_1, w_1) \rightleftharpoons (\mathcal{M}_2, w_2)$ then $(\mathcal{M}_1, w_1) \leftrightsquigarrow (\mathcal{M}_2, w_2)$.*

Theorem 4.2. *For any two ω-saturated models $(\mathcal{M}_1, w_1)$ and $(\mathcal{M}_2, w_2)$, if $(\mathcal{M}_1, w_1) \leftrightsquigarrow (\mathcal{M}_2, w_2)$ then $(\mathcal{M}_1, w_1) \rightleftharpoons (\mathcal{M}_2, w_2)$.*

Theorem 4.3. *A $\mathcal{L}$ formula is equivalent to the translation of a $\mathcal{L}^\mathrm{p}$ formula if and only if it is p-bisimulation invariant.*

We establish then the undecidability of PML_3, that is, PML with three pairs of standard and poison modalities, and three poison atoms (language $\mathcal{L}_3^\mathrm{p}$). We only consider satisfiability with respect to models with an empty valuation for the poison atom (in $\mathfrak{M}^\emptyset$).

Theorem 4.4. *The satisfiability problem for PML_3 is undecidable.*

The proof is done by reduction of the $\mathbb{N} \times \mathbb{N}$ tiling problem in a similar way as the undecidability proof for hybrid logic in [18]. Whether PML is also undecidable remains an open question. We suspect however that it is indeed the case given the following:

Theorem 4.5. PML *does not have the Finite Model Property.*

Finally, in order to position PML more precisely within the landscape of related logics, we show that PML is a proper fragment of the memory logic known as $\mathcal{M}(\text{ⓡ}, \text{ⓚ})$ [1, 2]: $\blacklozenge$ acts as ⓡ, and $\mathfrak{p}$ as ⓚ. We also show that PML can be embedded into $\mathcal{H}(\downarrow)$ [18], the hybrid logic containing only one binder thus not containing nominals nor $@_i$ operators.

---

[1]The cycle is not always of length n as formula $\blacklozenge\delta_n$ allows for the repetition of nodes.

REFERENCES

[1] 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.

[2] 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.

[3] 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.

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

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

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

[7] Bela Bollobas. 2012. *Graph theory: an introductory course.* Vol. 63. Springer Science & Business Media.

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

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

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

[11] 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.

[12] P. Dunne and M. Wooldridge. 2009. Complexity of Abstract Argumentation. In *Argumentation in Artifical Intelligence*, I. Rahwan and G. Simari (Eds.). Springer,

85–104.

[13] 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.

[14] 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.

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

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

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

[18] 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.

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

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

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

[22] 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.