GRAPH GAMES AND LOGIC DESIGN, THE CASE OF SABOTAGE AND EDGE CUTTING

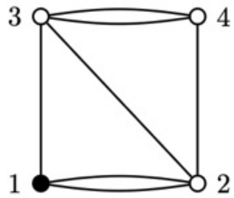
CLAWS 2025, Shantou, 25 October 2025

Johan van Benthem, https://staff.fnwi.uva.nl/j.vanbenthem/, johan@stanford.edu

1. **The sabotage game**

Definition 49 (Sabotage game) Starting from an initial point s, a *Traveler* T moves through a (directed or undirected) graph $\mathbf{G}$ with possibly multiple edges, from one position to the next via an available link, one at a time. The aim is to reach a designated goal region. But in each round of the game, a *Demon* D first cuts one link, anywhere in the graph. Traveler loses when at its turn, no further move can be made at its current position, while that position is not in the goal region.

Example 90 (A simple sabotage game) Traveler starts in point 1 of the graph below, trying to reach the goal point 4. Which player wins?



Game version of Graph Reachability. Who has the winning strategy in this game?

Aim: analyze standard computational tasks (Traveler) under obstruction (Demon).

Benign reinterpretation: a Teacher helping a Student. (Who has the w.s. now?)

The game is determined (one of the players has a w.s.) by Zermelo's Theorem.

2. **Sabotage modal logic**

Standard modal logic describes travel steps. $\mathbf{M}, s \models \Diamond\varphi$ iff $\exists t(Rst \,\&\, \mathbf{M}, t \models \varphi)$.

Model checking in Ptime. SAT decidable, bisimulation-invariant, FO-translatable.

Sabotage modality changes a model. $\mathbf{M}, s \models \blacklozenge\varphi$ iff $\exists tu(Rtu \,\&\, \mathbf{M}\{R := R - \{(t, u)\}, t \models \varphi)$.

Valid, or not valid? $\blacklozenge\Diamond\varphi \to \Diamond\blacklozenge\varphi$, $\Diamond\blacklozenge\varphi \to \blacklozenge\Diamond\varphi$.

Model checking Pspace-complete. Still FO-translatable (translation now needs memory), invariance for extended bisimulation. *SAT is undecidable*. Background: memory logics.

Open problem 1 Find good decidable fragments of SML (cf. FOL).

Sabotage μ-calculus. Winning positions for Traveler:

$$\nu p \cdot (\gamma \vee \blacksquare\Diamond p)$$

In the above game, let γ define {4}. We look at the computation stages 'from above':

The first approximation for the greatest fixed-point is the total set $\{1, 2, 3, 4\}$. Next, taking this set as a denotation for p, all points in the graph satisfy the formula $\blacksquare\Diamond p$, so $\gamma \vee \blacksquare\Diamond p$ denotes the same set, and approximation stops. But our first example showed that the starting point 1 for Traveler should not be in this set. What went wrong?

We need to evaluate in *universe of changing pointed models* ($\mathbf{M}$, s), not a fixed one.

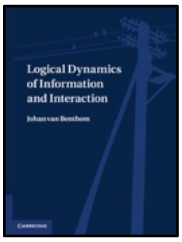
In changed models, complex modal statements may have changed their truth value.

Open problem 2 What happens to the standard theory of the modal μ-calculus: automata, parity games, etc., when we have to work across changing models?

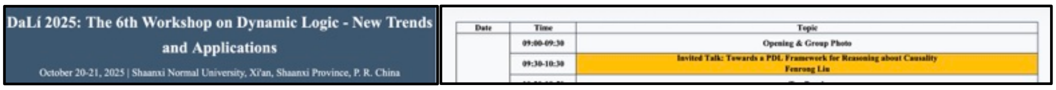
3. **Other edge cutting logics**

Partial observation in *dynamic epistemic logic*.

Example 19 (Drawing a card from the stack) "We both know that the top card on the stack in our game is *Red* or *Black*. I draw that card in your view and see what it is, you do not. Our initial model has two worlds {Red, Black} and epistemic relations $\sim_{\text{me}}, \sim_{\text{you}}$ connect both. After this public event, you still do not know what I drew, but you know that I know. My uncertainty line has disappeared, yours stayed."



Interventions on a variable X in a causal graph cut all incoming links to X.



Open problem 3 Where is the border decidable/undecidable for graph logics?

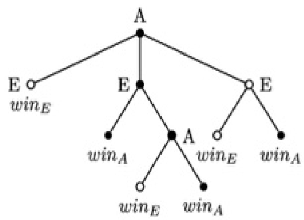
Open problem 4 Is it decidable if a finitely given logic is decidable?

4. **Game logics**

Theorem 24 (Zermelo's Theorem) Each two-player zero-sum game with a fixed finite bound on depth of histories is determined.

Proof. The proof is by a *coloring algorithm*. Color winning end nodes for *A black*, and for *E white*. We now extend this coloring upward in the game tree using the following rule:

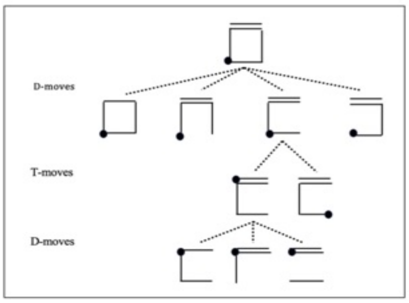
1. If all children of a node have been colored already, and the turn is for *A*, color the node black if at least one of its children is black, otherwise color the node white.
2. The rule for a turn of player *E* is completely analogous.



Definition of the points where player *i* has a winning strategy in modal μ-calculus:

$$\nu p \cdot (end \wedge win_i) \vee (turn_i \wedge \langle move_i \rangle p) \vee (turn_j \wedge [move_j] p)$$

Game tree for the sabotage game, just a part showing the idea:



Richer *temporal SGL* for such game trees, evaluated at pairs consisting of ('game state' (procedural information), 'board state' (Traveler's location in a current graph)).

The language of SGL has proposition letters describing properties of graph positions, including a proposition letter γ for Traveler's goal region, as well as game indicators 'turn_i' marking a turn for player i, 'end' for when the game stops, and 'win_i' for when player i wins. In addition there are Boolean operations as well as two existential modalities $\langle \text{T} \rangle \varphi, \langle \text{D} \rangle \varphi$ intended to describe effects of the players' moves to a next stage of the game in terms of propositions φ true there.

Can describe procedural features of the games, and procedural winning conditions.

Open problem 5 When transition needed from graph logics to game logics?

Open problem 6 Are there translations between the two sorts of logics?

See my paper in

5. Game representation levels

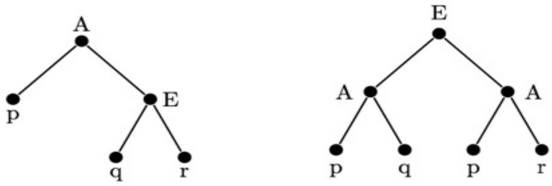


Fig. 10.7: Are these two games the same?

NO, in terms of actions, choice points, *modal logic with bisimulation* as game invariance. YES, if we focus on players' powers for achieving outcomes. A *power* for a player is a set of outcomes a player can force by playing one of her available strategies.

Left-hand game:

powers of A: $\{p\}$ (L), $\{q, r\}$ (R),

powers of E: $\{p, q\}$ (L), $\{p, r\}$ (R)

Right-hand game:

powers of E $\{p, q\}$ (L), $\{p, r\}$ (R),

powers of A: $\{p\}$ (LL), $\{p, r\}$ (LR), $\{p, q\}$ (RL), $\{q, r\}$ (RR)

Notice that the powers of A in the right-hand game contain two supersets of the power $\{p\}$. These weaker powers may be disregarded, as they are less informative than the stronger power 'inside' them, and we end up with the two players having the same powers in both games.

(i) C1 (Monotonicity)
If X is a power for i, and $X \subseteq Y$, then Y is a power for i.

(ii) C2 (Consistency)
If X is a power for i and Y is a power for j, then X, Y overlap.

(iii) C3 (Determinacy)
Either X is a power for i, or $-X$ is a power for j, where $-X$ is the complement of X in the total set of outcomes.

Power equivalence is a coarser game invariance with its own language for invariants.

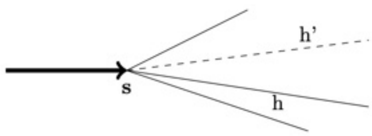
Modal neighborhood logic: $\mathsf{M}, s \models \{G, i\}\varphi$ iff there exists a set power X for player i in game G at stage s such that, for all $t \in X$: $\mathsf{M}, t \models \varphi$

This 'forcing modality' does not distribute over conjunction or disjunction.

Translation into modal game logic by the same idea as with Zermelo's Theorem.

Open problem 7

Extend this to infinite games.



Gale-Stewart Theorem. Weak Determinacy. Now we need a *temporal forcing logic*.

Theorem 25 All infinite zero-sum two-player games with an open set of winning histories for one of the players are determined.

$$\text{DET} \quad \{i\}\varphi \vee \{j\}\neg\varphi, \qquad \text{WD} \quad \{i\}\varphi \vee \{j\}\neg\{i\}\varphi$$

There are many more levels for looking at games, like for computational processes.

Tracking. Connections between levels may be via translation, but a further ubiquitous topic: track dynamic changes at one level faithfully by those at another (if possible).

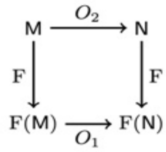


Fig. 4.3: A commuting diagram for tracking.

No game transform at power level tracks public announcement on extensive games.

(c) Tracking can be refuted using our initial example of two different games producing the same powers. If we announce $\neg q$ in the game on the left, the powers for A become $\{p\}, \{r\}$, but these are different from the powers of A in the game to the right after the same announcement, since A has no strategy there for forcing the outcome r.

Open problem 8 Develop a general theory of translation and tracking for game logics living at different levels of representation.

6. **Another perspective: Probability**

For which player is the sabotage game favorable?

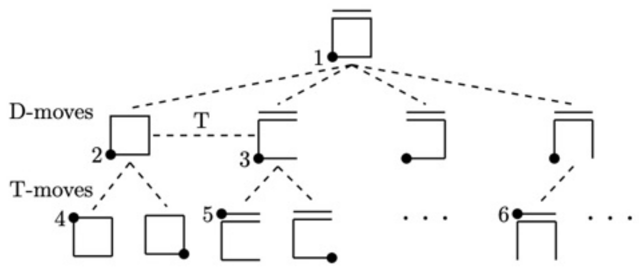
Fact 58 For any first-order sentence, its probability of being true goes to either 0 or 1 as graph size goes to infinity. It can even be decided effectively, for any given formula, which of the two options occurs.

The Zero-One Law even holds for the *fixed-point extension of first-order logic*, which contains the sabotage μ-calculus under translation. By analyzing the sabotage game played on the countable random graph, it can be shown that

Traveler's probability of winning goes to 1 as finite graph size goes to infinity.

7. **Challenge: Imperfect Information**

Let Traveler (and perhaps also Demon) have 'short sight': only limited observation of their environment in the graph. We get an extensive game with *imperfect information*:



In the course of the game, players *may learn more than what they see directly*.

The logic for analyzing such games: *dynamic-epistemic logic* DEL: see literature.

Open problems galore in this area: it's just starting (Fenrong Liu, Dazhu Li)