LOGIC IN GAMES

Johan van Benthem, Friday, 5 August 2011, ESSLLI Ljubljana

1 Logic and games, an old historical connection

Logic is about individual reasoning, but its origins may lie in multi-agent argumentation (legal debate, …). Semantic, proof-theoretic, but also *pragmatic* intuitions of validity:

A valid conclusion is one for which there is a winning strategy in debate.

Practical uses of games in logic: evaluating formulas, dialogue, comparing or building models, ... They also provide a foundational look at logical constants as game actions:

$\vee$, $\wedge$ are *choices* by different players, $\neg$ is *role switch*, and there are more.

2 Argumentation games, a sketch

Debating games in Middle Ages, modern revival. Illustration: the valid inference $\neg A,\ A \vee B \Rightarrow B$. Proponent (***P***) defends B against Opponent (***O***) who commits to $\neg A$, $A \vee B$. Players speak in turn. 1 ***O*** challenges ***P*** to produce a defense of B, then 2 ***P*** now presses ***O*** on one of his commitments, demanding a choice, 3 ***O*** must respond to this, having nothing else to day. Option 3': ***O*** commits to A. 4' ***P*** points at ***O***'s commitment to $\neg A$, and wins because of ***O***'s self-contradiction. Option 3'': ***O*** commits to B. 4'' ***P*** now uses B as his defense to 1. ***O*** has nothing further to say, and loses.

Proponent has a *winning strategy*: whatever Opponent does, she can counter to win.

An inference is valid *if P has a winning strategy for the conclusion against any O granting the premises*.

3 Model checking games

Leibniz' story of two mathematicians discussing a quantified statement $\forall\varepsilon\ \exists\delta\ \phi(\delta, \varepsilon)$.

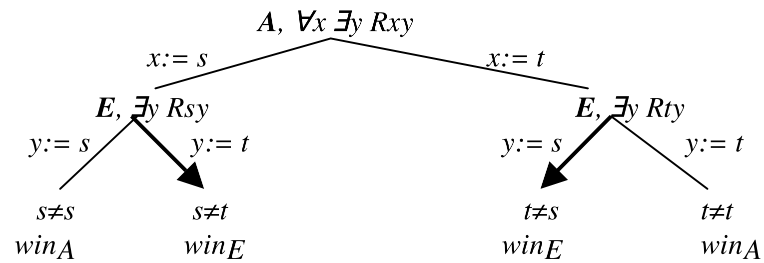
Logical statements ϕ define *evaluation games* **game**$(\phi, \boldsymbol{M}, s)$ in arbitrary models ***M*** with a assignment s from variables to objects. The game is played by Verifier ***V*** who claims that ϕ is true, and Falsifier ***F*** who claims that ϕ is false. Clauses for attack and defense:

atoms	*test* to determine who wins (if true in ***M***, ***V*** wins, otherwise, ***F***)
disjunction $\vee$	is ***V***'s *choice*, and play continues with the chosen disjunct
conjunction $\wedge$	is ***F***'s *choice*, and play continues with the chosen conjunct
negation $\neg$	triggers a *role switch* between the two players
existential quantifiers $\exists x\psi(x)$	let ***V*** *pick an object d* in ***M***, and play continues with $\psi(d)$
universal quantifiers $\forall x\psi(x)$	are the same instruction, but now for player ***F***

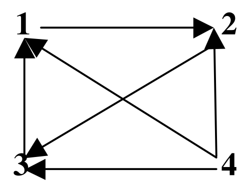
Truth/Strategy Lemma The following two assertions are equivalent:

(a) ***M***, $s \models \phi$ (standard truth), (b) ***V*** has a *winning strategy* in ***game***(ϕ, ***M***, *s*)

Example Two objects *s, t*. The game for the formula $\forall x\ \exists y\ x{\neq}y$ is a simple finite tree, with player ***E***'s winning strategy indicated by the two bold-face arrows:



More challenging cases



Communication Network

Evaluate the first-order formula

$\forall x\ \forall y\ (Rxy \vee \exists z(Rxz\ \&\ Rzy))$:

Who wins the game?

4 Game-theoretic background: determinacy

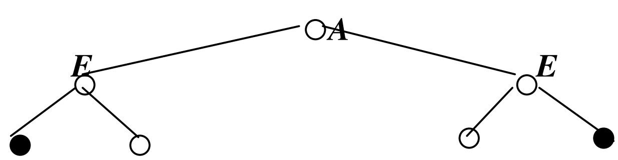
In model checking games, either Verifier or Falsifier has a winning strategy!

Theorem (Zermelo, Euwe) Zero-sum 2-player games of fixed finite depth are determined.

Proof Well-known algorithm, also in AI. Color end nodes *black* that are wins for ***A***, and color the other end nodes *white*, being the wins for ***E***. Then extend this stepwise:

(a) if ***A*** is to move, and at least one child is black, color *n black;* otherwise, *n* is *white*

(b) if ***E*** is to move, and at least one child is white, color *n white;* otherwise, *n* is *black*

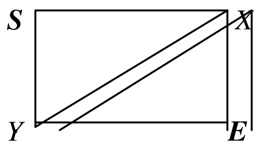


Extends to Chess. A century after Zermelo, still unknown *which player* has a non-losing strategy!

Another take on the proof: The reason is simply the logical law of *Excluded Middle*.

5 One more illustration: 'sabotage', or teaching games

A Student at *S* wants to reach escape *E*, the Teacher wants to prevent him from getting there. Each line is a path that can be traveled. At each round, the Teacher cuts a link anywhere in the diagram, while the Student must travel one link still open to him at his current position:



If Teacher is greedy, and starts cutting a link *S–X* or *S–Y* right in front of the Student, then Student can reach the escape *E*. However, Teacher does have a *winning strategy* for preventing the Student from reaching *E*, by doing something else: what?

6 Digression: game operations

Game-theoretical take on the logical operations:

(a) *conjunction and disjunction are* choices $G \wedge H$, $G \vee H$

(b) *negation is* role switch, often called 'dual' $\neg G$, or G^d

The quantifier $\exists x$ is an atomic game of object picking by Verifier. General operation

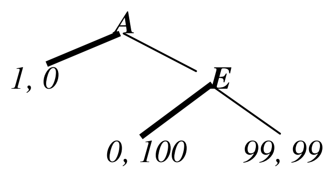
(c) *sequential composition* of games $G ; H$

Other scenarios introduce new logical operations: 'Life' = 'Career' ‘and’ 'Family'.

(d) *parallel composition* of games $G \,||\, H$

7 Game solution with preferences: backward induction

Game theory has much more complex games. Here is a simple social scenario:



Zermelo’s Algorithm generalizes to *BI* procedure for extensive games with preferences:

> “At end nodes, players already have their values marked. At further nodes, once all daughters are marked, the player to move gets her maximal value that occurs on a daughter, while the other, non-active player gets his value on that maximal node.”

Other examples of Backward Induction reasoning: ‘The Election’, ‘The Robbers’.

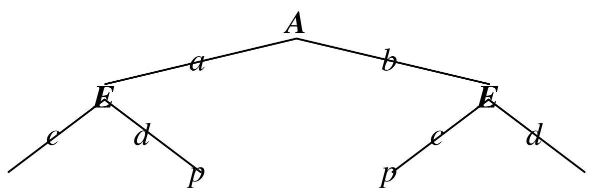
Generalize strategy to any subrelation of the *move* relation (plans may leave things open):

> Mark all moves as ‘active’. Call move *a dominated* if it has a sibling move all of whose reachable endpoints via active nodes are preferred by the current player to all reachable endpoints via *a* itself. Then, at each stage, mark dominated moves as ‘passive’, leaving all others active. In the preference comparison, ‘reachable endpoints’ by an active move are all those that can be reached via further moves that are still active at this stage.

8 Finding the logical form of rationality: action and preference

Many logical notions involved in explaining/predicting this game solution method: action, preference, knowledge, belief, conditionals (philosophical + computational logic). Reasoning is appealing, but not obvious. Balance of information and evaluation is tricky.

Dynamic logic Extensive game trees are process models for Thursday's action logics:



A has two strategies, ***E*** has four. Modal languages can define structure, and strategies:

$[a]\langle d\rangle p \wedge [b]\langle c\rangle p$	player ***E*** has a response ensuring that p always holds
$[move_A]\langle move_E\rangle win_E \vee \langle move_A\rangle[move_E]\neg win_E$	***A*** or ***E*** has a winning strategy
$\{G, \sigma, \mathbf{E}\}\phi$	***E***'s strategy σ makes game G have only runs with ϕ

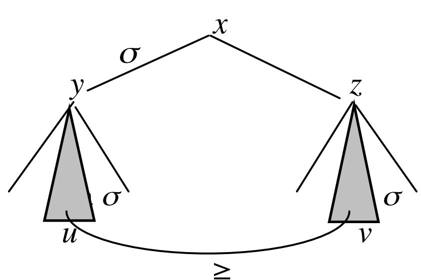
Strategies as relations Example, the above outcome can also be described as $(?turn_A ; left) \cup (?turn_E ; left)$. Many strategies are *PDL* programs, e.g., 'Tit-for-Tat', 'Copy Cat'.

Adding preference $\langle pref_i\rangle\phi$: agent i prefers some node with ϕ to the current one.

Fact The *BI* strategy is the largest relation σ validating for each player i:

$$(turn_i \ \& \ \langle\sigma\rangle[\sigma^*](\mathbf{end} \rightarrow p)) \rightarrow [move\text{-}i]\langle\sigma^*\rangle(\mathbf{end} \ \& \ \langle pref_i\rangle p).$$

Corresponding confluence property *CF* links moves, preference, and 'best actions' σ:



$$\&_i \ \forall x \forall y \ ((Turn_i(x) \wedge x \ \sigma \ y) \rightarrow (x \ move \ y \wedge \forall z \ (x \ move \ z \rightarrow \exists u \exists v \ (end(u) \wedge end(v) \wedge y \ \sigma^* u \wedge z \ \sigma^* v \wedge v \leq_i u)))$$

9 Another perspective: knowledge and information dynamics

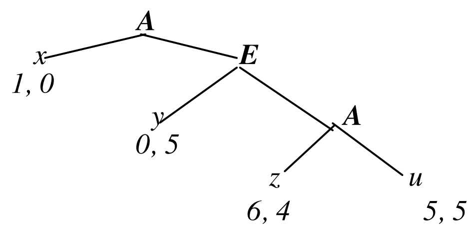
Link with Wednesday's logic of knowledge. Backwards Induction is a procedure creating expectations about optimal behavior, and we can also focus on its information dynamics. Game solution as *prior deliberation*, players remind themselves of everyone's rationality.

At a turn for player i, move a is *dominated* by sibling move b if every history through a ends worse, in terms of i's preference, than every history through b. Now ***rat*** says that

"at the current node, no player ever chose a strictly dominated move coming here".

Iterated announcement of ***rat*** in ***M*** reaches a limit model, where no node is dominated:

Example Solving games by iterated assertions of Rationality.



next stages remove branches…

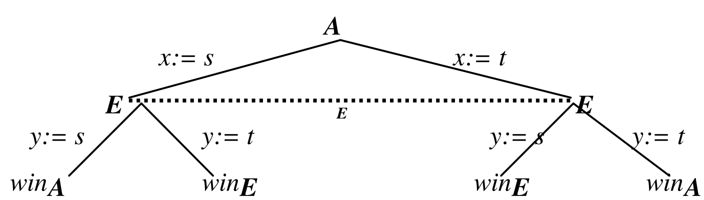
Fact In any game ***M***, the limit sub-model for ***rat*** is the actual sub-tree computed by *BI*.

Announcement limits Limit sub-models $(\varphi, \boldsymbol{M})^{\#}$ emerge in many settings: game solution, Muddy Children, disagreement in epistemology. Two cases: self-fulfilling, self-refuting.

Alternative Analyze BI procedure in terms of dynamic logics for belief revision.

10 Knowledge once more: games with imperfect information

Realistic: observation uncertainty in card games, war. In game for $\forall x\ \exists y\ x{\neq}y$ with $\{s, t\}$, let ***E*** be ignorant of the object chosen by ***A***. The *dotted line* marks ***E***'s uncertainty:



Only 'uniform' strategies can be played in case of uncertainty, ***E*** only has 'left', 'right'. *Determinacy is lost*: neither player has a winning strategy! *Knowledge-action logic*:

De dicto, $K_E(\langle y{:=}t\rangle win_E \vee \langle y{:=}s\rangle\ win_E)$ ***E*** knows that some move will make her win

De re, $\neg K_E \langle y{:=}t\rangle win_E) \wedge \neg K_E \langle y{:=}s\rangle win_E)$ ***E*** does not know which move will make her win

Analyzing players We can also describe special features of players of *ii* games, such as Perfect Memory, a modal interchange axiom between knowledge and action:

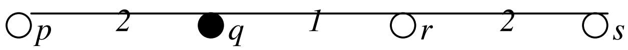
$$K_E[a]\phi \rightarrow [a]K_E\phi$$

Not always valid in logics of action: depends on agents' power of memory/observation.

Commuting diagram for Perfect Memory: $\forall xyz{:}\ ((x\ R_a\ y \wedge\ y \sim_i z) \rightarrow \exists u\ (x \sim_i u \wedge u\ R_a\ z))$.

11 Knowledge games

New imperfect information games in logic. Example: 'be the first to know'. Start in any epistemic model, players have to say something that they know and that is significant. They win if they are the first to know the actual world, or the other player cannot move.



Player *1* has to start. In the actual world *q*, he can say either *{q, r}* or *{q, r, p}* or *{q, r, s}*. While *{q, r, p}* is also true at *p*, he cannot say it there, since it would already be a win. However, he could also say *{q, r, p}* at the world *r* which he cannot distinguish from *q*.

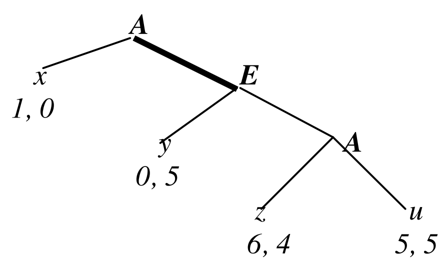
Drawing the game tree of all possible moves shows the players' options, their wins, and it suggests their reasoning. Also instructive to draw: the protocol where agent 2 starts first.

Two kinds of information Assertions produce *factual information* about the real world, but also *procedural information* about what observed moves 'mean' given the protocol. (In terms of dynamic-epistemic logic, $\langle !\varphi \rangle T$ may now be more informative than just φ.)

Fact Neither player has a uniform winning strategy in this game.

But game theory shows that equilibria exist in probabilistic mixed strategies.

12 Problems with the logic so far: 'Paradox of Backwards Induction'

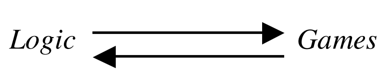


BI says ***A*** will go left at the start, on the basis of reasoning that is available to both players. But now look at actual play. If ***A*** plays *right* (see the black line) what should ***E*** conclude? Does not this mean that ***A*** is not following the *BI* reasoning, and hence that all bets are off as to what he will do later on in the game?

Current debate From game to richer model of players, belief revision policies, *Theory of Play*.

13 Closing the circle: logic *of* games, and logic *as* games

Two directions: use games to analyze logic, use logic to analyze games.



Ideas and results can flow in both directions, but many topics remain to be explored.