Out-of-equilibrium belief revision and dynamic perspective on strategic reasoning

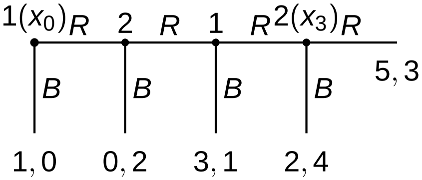
Cédric Dégremont

ILLC, Amsterdam

20/07/07

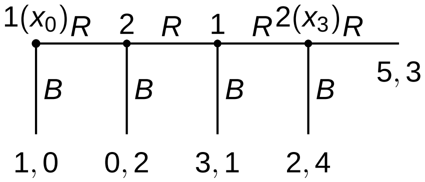
joint work with Johan van Benthem and Jonathan Zvesper

4-legs Centipede Game



- Backward induction path is B at x_0
- 2 is rational (at x_3), 1 is rational and believes that 2 is rational (at x_2)...

4-legs Centipede Game



- Backward induction path is *B* at x_0
- 2 is rational (at x_3), 1 is rational and believes that 2 is rational (at x_2)...
- What if 1 plays Right?
- The description of rational behavior must include rules of conduct for all conceivable situations - including those where 'the others' behaved irrationally, in the sense that the theory will set for them. [von Neumann and Morgenstern]
- Any description of rational behavior which claims to be complete must be immune to defections from it. [Reny]

Outline

- Dynamic logic modeling of players' strategic reasoning
- Assumptions under which backward induction is immune

Johan van Benthem. Rational dynamics and epistemic logic in games. to appear in *International Journal of Game Theory*.

Johan van Benthem. Dynamic logic for belief revision. *Journal of Applied Non-classical Logics*, 17(2), 2007.

Philip J. Reny. Common belief and the theory of games with perfect information. *Journal of Economic Theory*, 59(2):257–274, April 1993.

Rational belief revision

- Is backward induction immune (in the sense that no player can be better off than in the equilibrium outcome by playing out of equilibrium)? No, in general it is not.
- It will depends on players belief revision policy (Reny, Stalnaker, Board). What is a "rational" belief revision policy?
- Sleeping Beauty wakes up hungry and not knowing whether we are $\{Monday, Tuesday\} \times \{8AM, 9AM\}$ nor whether she will find the shop $\{< Open >, < Closed >\}$. What is she expecting?
- She strongly believes (takes it as hard information) that the shop will be open if we are Tuesday at 9 AM and closed otherwise ($\rightarrow$ PAL or Lexicographic upgrade) and weakly believes (takes it as soft information) that we are Monday at 9 AM (Elite change).
- She is expecting to find the shop closed. But she goes there anyway and finds the shop open.

Rational belief revision (2)

$(B * 1)$	$B * \phi \vdash \phi$
$(B * 2)$	If $\phi \nvdash \bot$ then $B * \phi \nvdash \bot$
$(B * 3)$	If $B \cup \phi \nvdash \psi$ then $B * \phi \nvdash \psi$
$(B * 4)$	If $B \cup \phi \vdash \psi$ and $B * \phi \nvdash \psi$ then $B * \phi \subset E \subseteq B \cup \phi$ implies $E \vdash \bot$

Reny $c_i(x)$ reaches x

- ϕ for a move: **consistency maintenance** or strong revision.
- **contraction**-oriented or abduction-oriented

Dynamic turn in modal modeling

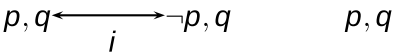
Static modeling of games

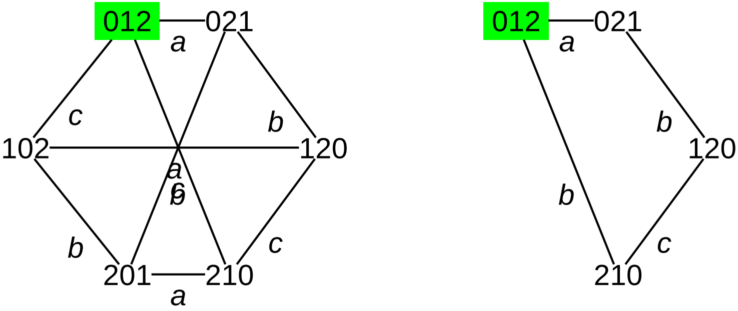
- Modal modeling of non-cooperative games
- Relational structures (LTS / Epistemic)
- Static aspects: Nash Equilibrium
- Goal? Reasoning?

Dynamic turn in epistemic logic

- Classical approach (modal structures): modeling knowledge and belief
- Dynamic aspects: learning new information
- Modification of epistemic structures $\langle W, (\sim_i)_{i \in N} \subseteq W \times W, V : Prop \rightarrow \wp(W) \rangle$

PAL: Reducing epistemic structures





$$\mathcal{M}, 012 \models [\neg(201 \vee 102)!]K_c 012$$

Games structures and PDL

- Structures

$$\langle W, N, A, (\overset{i\ a}{\rightarrow})_{a\in A, i\in N}, (\succ_i)_{i\in N}, (\sim_i)_{i\in N}\rangle$$

  with a valuation $V : Prop \rightarrow \wp(W)$

- Language

  $\alpha ::= ai \mid a \mid i \mid \succ_i \mid \rightarrow \mid \alpha \cup \alpha \mid \alpha ; \alpha \mid \alpha \cap \alpha \mid \alpha^{-1}$

  $\phi ::= \mid p \mid \bot \mid \phi \rightarrow \phi \mid \langle\alpha\rangle\phi \mid K_i\phi$

- Semantics

$$\begin{aligned} R_{\alpha\cap\beta} &= R_\alpha \cap R_\beta \\ R_{\beta^{-1}} &= \{(v, w) | wR_\beta v\} \\ R_{\alpha;\beta} &= \{(v, w) | \text{Il y a un } x\ vR_\alpha x R_\beta w\} \end{aligned}$$

Which rationality?

- Concept of (ir)rationality of a *decision*
- Modal logic: local and Internal perspective
- $$\langle (i^{-1}; i) \cap {\succ_i}^{-1} \rangle \top$$
- a state is irrational if the (last) decision leading to it was irrational
- Decision irrationality = suboptimal choice

Rationality "Announcements"

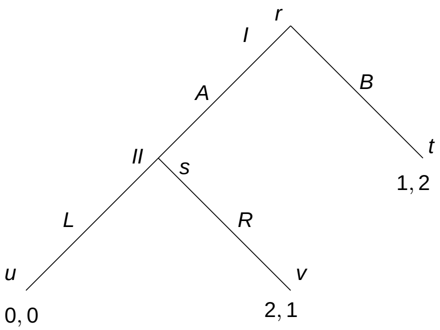


Figure: Entry game

Rationality "Announcements" (2)

Same game after a reduction step

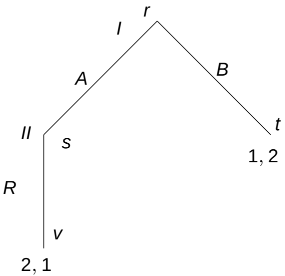


Figure: Game of figure 1 after another step of reduction.

Rationality "Announcements" (3)

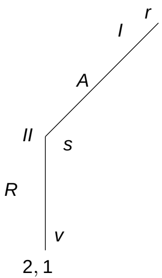


Figure: Game after two reductions.

Backward induction / EF Rationalizability

- Generic games - no indifference
- Generic games $\rightarrow$ unique solution under backward induction
- Coincide on generic games (van Benthem)
- Characterization of BI solution as a sub-model

Proposition

A maximal path survives iterated announcement of common rationality ($\bigwedge_{i \in N} rat_i$) iff its edges are contained in the union of subgame perfect equilibrium (EF Rationalizability?)

Example

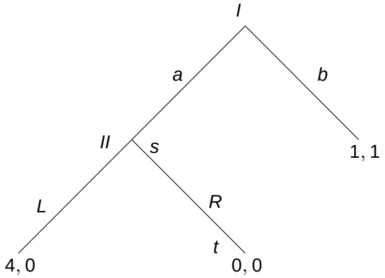


Figure:

DR+PR	BI
{a,L}	{a,L}
{a,R}	
{b,R}	{b,R}

Qualitative Decision-Making

- The effect of a decision rule will be cashed out as the elimination of states that would be reached only if the player did not respect it, in our case if she acted in a irrational way, i.e. as relativization of a game to his rational core.
- An interactive situation $\mathcal{G}$ being given, the set of moves accepted by the decision rule ”do not take a strictly dominated move" at a state w where i is to play is $\{a \in A \mid \mathcal{G}, w \models [!Rat_i]\langle a \rangle \top\}$ where Rat_i is the concept of rationality.
- Use information you trust (the most) to simplify the decision-making situation you are facing until your decision rule returns a singleton.

Types and belief hierarchy

- Hard facts: decision rule (e.g. rationality) and preferences
- Beliefs hierarchy (1st and higher order beliefs)
- Decision rule (e.g. rationality) and belief hierarchy
- Expectations about other players' strategies is not a primitive

Linking static beliefs to cognitive actions

From types...

1's type

			Rat_1
		$B_1 Rat_2$	
	$B_1 B_2 Rat_1$		
$B_1 B_2 B_1 Rat_2$			
...			
	... to assumptions		
			$Rat_1 !$
		$Rat_2 !$;	$\top !$
	$Rat_1 !;$	$\top !;$	$\top !$
$Rat_2 !;$	$\top !;$	$\top !;$	$\top !$

- $Rat_2 !; Rat_1 !; Rat_2 !; Rat_1 !;$
- Sufficient and necessary assumptions (and thus beliefs) to make *Bottom* the only decision selected by the choice rule.

2's expectations and out-of-equilibrium beliefs

2's type

B_2				Rat_1
B_2			$B_1 Rat_2$	
B_2		$B_1 B_2 Rat_1$		
B_2	$B_1 B_2 B_1 Rat_2$			
...				
	... to the following thought experiment			
				Rat_1
			$Rat_2!$;	$\top!$
		$Rat_1!$;	$\top!$;	$\top!$
	$Rat_2!$;	$\top!$;	$\top!$;	$\top!$

- If I were 1 and had the following type, I would choose *Bottom*
- Now, what if 1 plays 2?
- 2 withdraw (at least) one of his assumptions about 1's type to maintain consistency!

Sufficient belief revision policy for immunity of backward induction (conjecture)

- We don't revise our beliefs about other players' payoff function.
- Dynamic framework sketched
- Less confident in higher-order beliefs

Sufficient belief revision policy for immunity of backward induction (conjecture)

- We don't revise our beliefs about other players' payoff function.
- Dynamic framework sketched
- Less confident in higher-order beliefs

In generic (no payoff ties) two-players extensive games of perfect information:

(Conjectured Proposition)

If there is common initial belief in rationality, rationality, the belief revision policy of the players satisfies

$$
\begin{array}{ll}
(B*1) & B*\phi \vdash \phi \\
(B*2) & \text{If } \phi \nvdash \bot \text{ then } B*\phi \nvdash \bot \\
(B*3) & \text{If } B\cup\phi \nvdash \psi \text{ then } B*\phi \nvdash \psi \\
(B*4) & \text{If } B\cup\phi \vdash \psi \text{ and } B*\phi \nvdash \psi \text{ then } B*\phi \subset E \subseteq B\cup\phi \text{ implies } E \vdash \bot
\end{array}
$$

and if players higher order beliefs are strictly less entrenched then backward induction is immune (to out of equilibrium play).

Invariance lemma

Conditions of invariance under adding higher order beliefs.

Conjectured Lemma

Iterated beliefs of rationality to the degree n overrides any higher order belief in a subgame of length at most $n + 1$

Example:

- Adding a 2nd order belief to 1's type ($B_1 B_2 Right_1$) is not safe!
- Adding a 4th order belief to 1's type ($B_1 B_2 B_1 B_2 Right_1$) is.

Completeness

- Syntactic proof
- $\langle (i^{-1}; i) \cap \succ_i{}^{-1} \rangle \top$
- $Irrat_i := \downarrow y.[\,]^*([\,]\bot \to \downarrow x.@_y \langle i^{-1}; i \rangle [\,]^*([\,]\bot \to \langle \succ_i \rangle x)$.
- In words i is irrational whenever every currently accessible terminal states is strictly worse for i than any of world that would be accessible if he had made a different choice at his previous move.
- $\mathcal{M}' = (\mathcal{M} \cup ||\langle i^{-1} \rangle \top||^{\mathcal{M}_0}) \upharpoonright Rat_i$ (in fact generated submodel...)
- $\mathcal{M}' = \mathcal{M} \upharpoonright Rat_i$
- $\langle (!Rat_i)^* \rangle \langle \leftarrow^* \rangle [\leftarrow] \bot$

Summary

- Dynamic (belief revision) framework for strategic reasoning
- (Arguably) reasonable belief revision policy under which backward induction is immune

Major open Problems:

- Complete compositional analysis
- Decidability of the fragment and implementation problem $(\mathcal{M}, root \models [Rat! \cup Pref]^*[]^*\phi)$:(
- Token semantics (Bonnay/Egré) and bounded rationality (or syntactic counterpart)
- Multi-agent case
- Simulatenous moves and framework for forward induction
- Precise connection to quantitative frameworks (Battigalli)