Decisions, Games & Logic '08

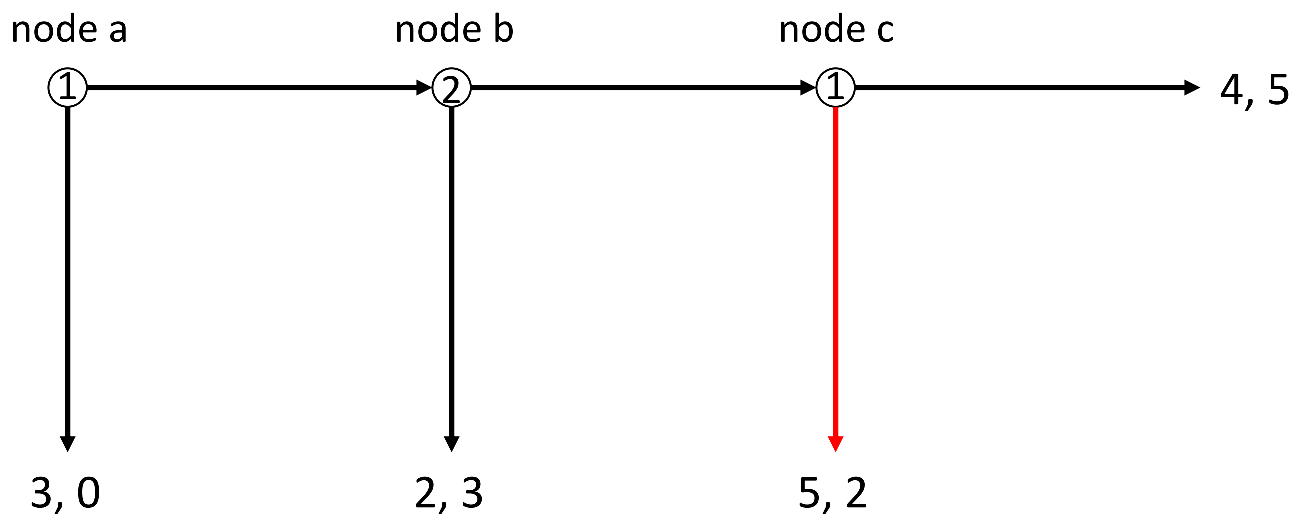
Amsterdam, July 1st 2008

Logical Foundations of Game Theory

Lecture 2: Dynamic Games

Oliver Board, University of Pittsburgh

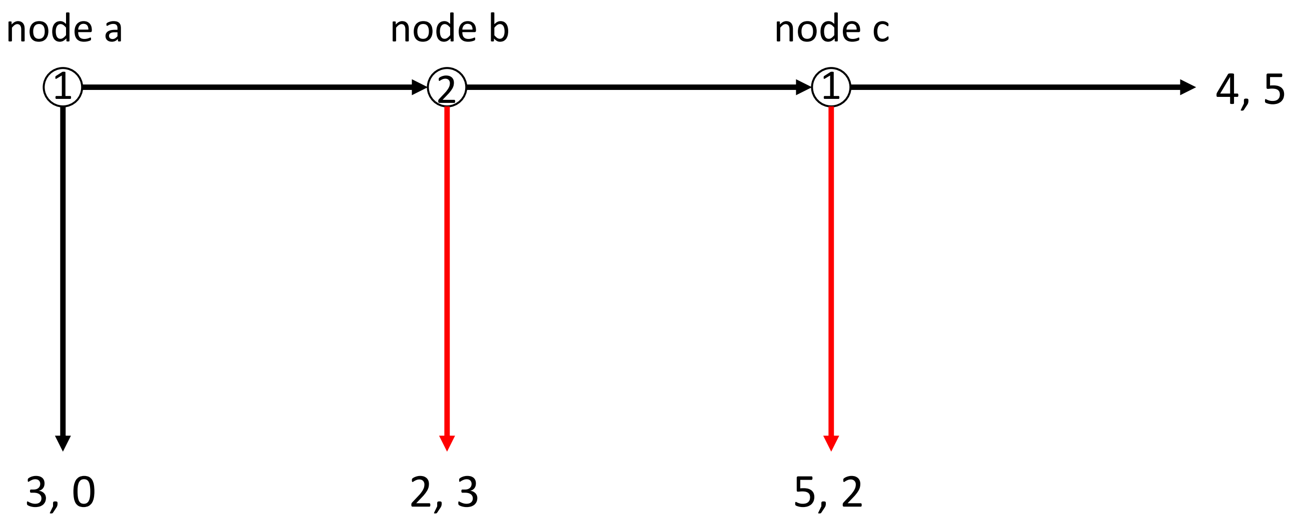
A Motivating Example



A backward induction argument: assume common knowledge of rationality

$Rat_1 \rightarrow down_c$

A Motivating Example

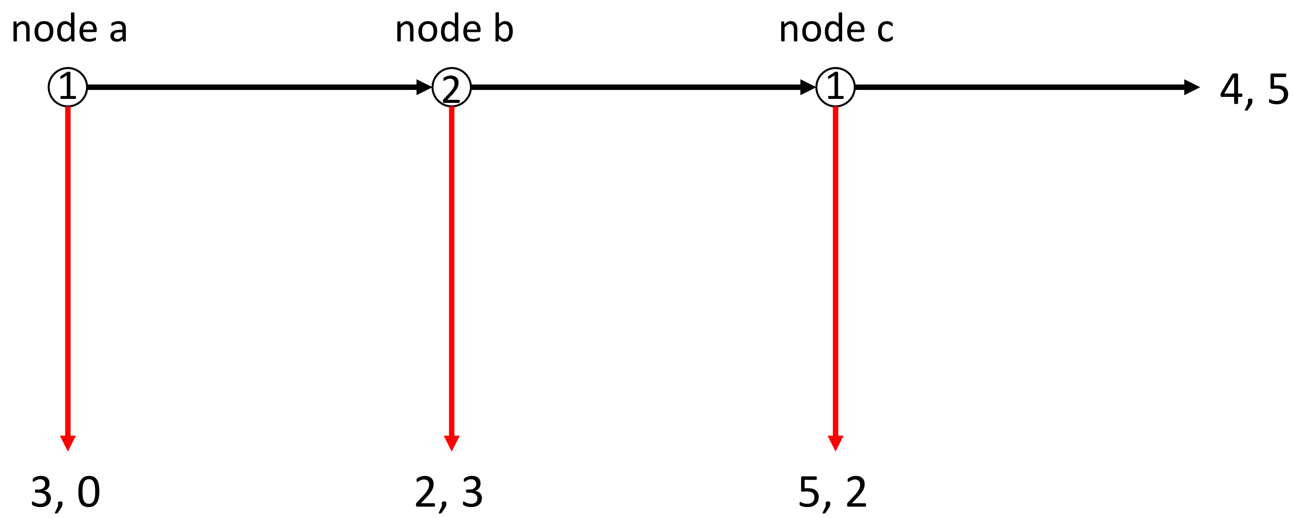


A backward induction argument: assume common knowledge of rationality

$Rat_1 \rightarrow down_c$

$Rat_2 \,\&\, K_2 Rat_1 \rightarrow K_2 down_c \rightarrow down_b$

A Motivating Example



A backward induction argument: assume common knowledge of rationality

$Rat_1 \rightarrow down_c$

$Rat_2 \ \& \ K_2 Rat_1 \rightarrow K_2 down_c \rightarrow down_b$

$Rat_1 \ \& \ K_1 Rat_2 \ \& \ K_1 K_2 Rat_1 \rightarrow K_1 down_b \rightarrow down_a$

But what if common knowledge of rationality breaks down as the game progresses?

Extensive Form Games

An extensive form game with complete information and perfect recall consists of

$$\langle n, H, \iota, \mathcal{P}, U_i \rangle$$

- n is a set of players
- H is a finite set of sequences, which satisfies

  (i) $\emptyset \in H$

  (ii) If $(a^k)_{k=1,\dots,K} \in H$ and $L < K$, then $(a^k)_{k=1,\dots,L} \in H$.

  Each element $h \in H$ is a *history*, and each component of h is an *action*.

  A history $(a^k)_{k=1,\dots,K} \in H$ is *terminal* if there is no a^{K+1} such that $(a^k)_{k=1,\dots,K+1} \in H$.
  We use $A(h) = \{a \mid (h, a) \in H\}$ to denote the set of actions available after history h, and Z to denote the set of terminal histories.

Extensive form games contd.

- $\iota : H \setminus Z \to n$ tells us who moves at each non-terminal history
- $\mathcal{P}$ is a partition of $H \setminus Z$ into *information sets*. If $h' \in \mathcal{P}(h)$, we assume

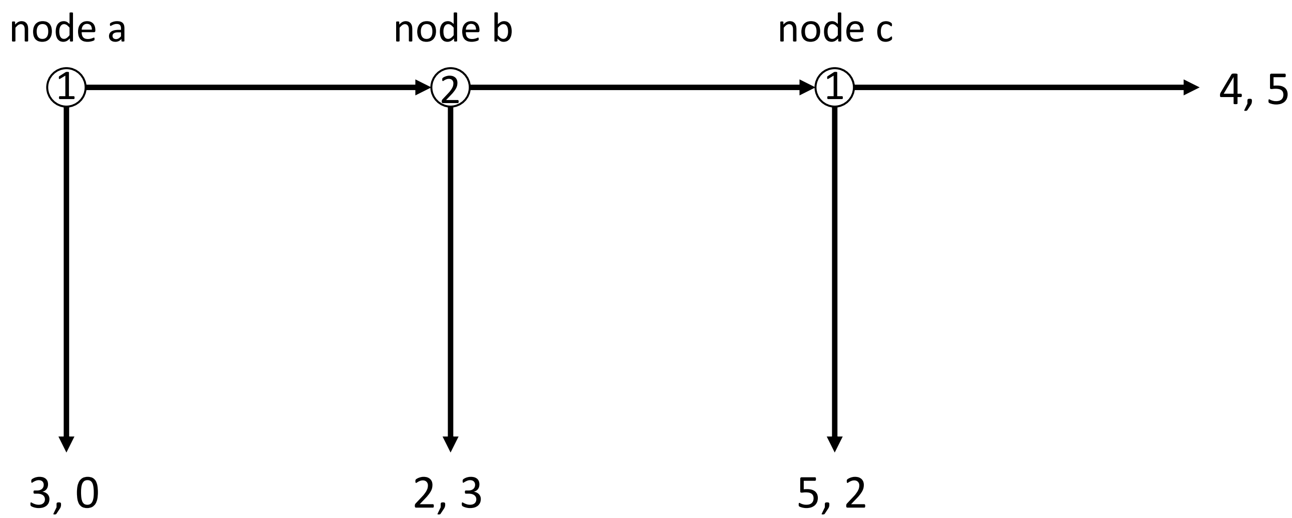
  (i) $\iota(h) = \iota(h')$ (players know when it is their turn to move)

  (ii) $A(h) = A(h')$ (players know which actions are available to them)

  So for any information set $P \in \mathcal{P}$, we can write $\iota(P)$ for the player on move, and write A_P for the actions available to her. Further, we can partition $\mathcal{P}$ into sets $\mathcal{P}_i$. *Perfect recall*: let $X_i(h)$ denote *i*'s experience at history h, i.e. the sequence of information sets she encounters and the actions she takes. Then if $h, h' \in P$ for some $P \in \mathcal{P}_i$, $X_i(h) = X_i(h')$.

- $U_i : Z \to \mathbb{R}$ gives *i*'s utility at each terminal history

Example



- $n = \{1,2\}$
- $H = \{\emptyset, d, a, ad, aa, aad, aaa\}$
- $\iota(\emptyset) = \iota(aa) = 1;\ \iota(a) = 2$
- $\mathcal{P} = \big\{\{\emptyset\}, \{a\}, \{aa\}\big\}$
- $u(d) = (3,0);\ u(ad) = (2,3);\ u(aad) = (5,2);\ u(aaa) = (4,5)$

Strategies

A strategy for player i specifies an action for each of her information sets:

$$s_i \in S_i = \times_{P \in \mathcal{P}_i} A_P$$

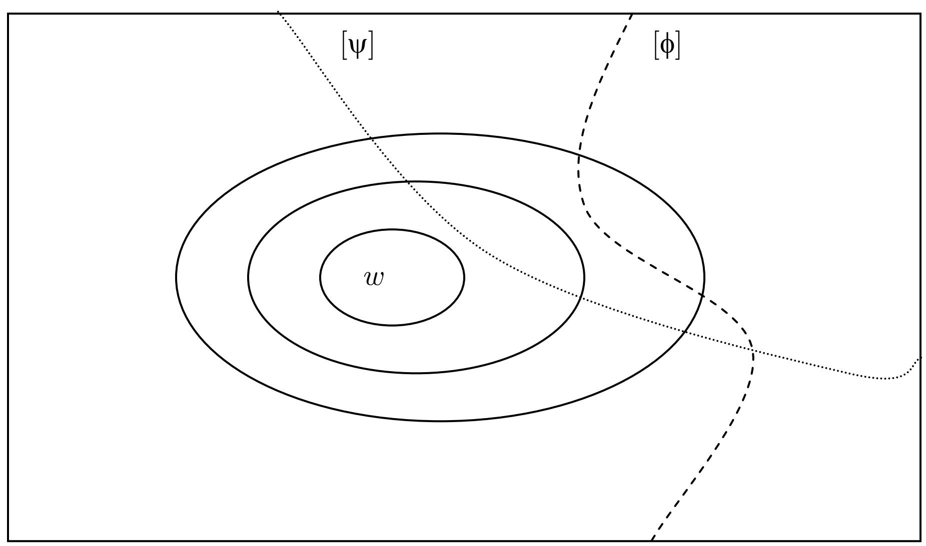
We use $S_i(P)$ to denote the set of strategies for player i that are consistent with information set $P \in \mathcal{P}_j$ being reached (e.g. $S_1(aa) = \{ad, aa\}$)

Let $h(s)$ denote the (terminal) history resulting when strategy profile s is played. Then:

$u_i : S \to \mathbb{R}$ denotes player i's strategic form utility function, i.e. $u_i(s) = U_i(h(s))$

Modeling Belief Revision

In an extensive form game, a player's knowledge/beliefs may change as the game progresses. To model rational play, we need to know what her beliefs will (or would) be at every node where she is on move, including those she thought wouldn't be reached at the beginning of the game.



Models of Games

A model of an extensive form game consists of:

$$\langle W, \preccurlyeq_1, \dots, \preccurlyeq_n, p_1, \dots, p_n, f_1, \dots, f_n \rangle$$

$\preccurlyeq_i$ is vector of *plausibility orderings* for player i, one for each w.

Each $\preccurlyeq_i^w$ is a complete and transitive binary relation on W

Beliefs at an information set: suppose $P \in \mathcal{P}_i$ is one of player i's information sets

Her beliefs at P are given by

$$p_{i,P}^w(E) = \frac{p_i(min_i^w([P]) \cap E)}{p_i\left(min_i^w([P])\right)}$$

where $min_i^w(X)$ is the set of worlds which are $\preccurlyeq_i^w$-minimal in X, and $[P]$ is the set of worlds where strategies consistent with P are played.

Expected Utility

Expected utility is defined just as in the static case, except that the EU at a given information set is computed given beliefs at that information set.

The expected utility for player i at information set P if she plays s_i at state w is given by

$$EU_{i,P}^{w}(s_i) = \sum_{x \in W} p_{i,P}^{w}(x) \cdot u_i\big(s_i, f_{-i}(x)\big)$$

for all $P \in \mathcal{P}_i$ such that $s_i \in S_i(P)$.

Note: this expression is valid only for strategies of player i that are consistent with information set P being reached. To evaluate a given strategy, we don't need to know what would happen at information sets that can't be reached it that strategy is played.

Rationality

At any given state w, we say that player i is *rational* at information set $P \in \mathcal{P}_i$ if her chosen strategy reaches that information set, and her expected utility conditional on P being reached is at least as high as it would be if she did something else instead.

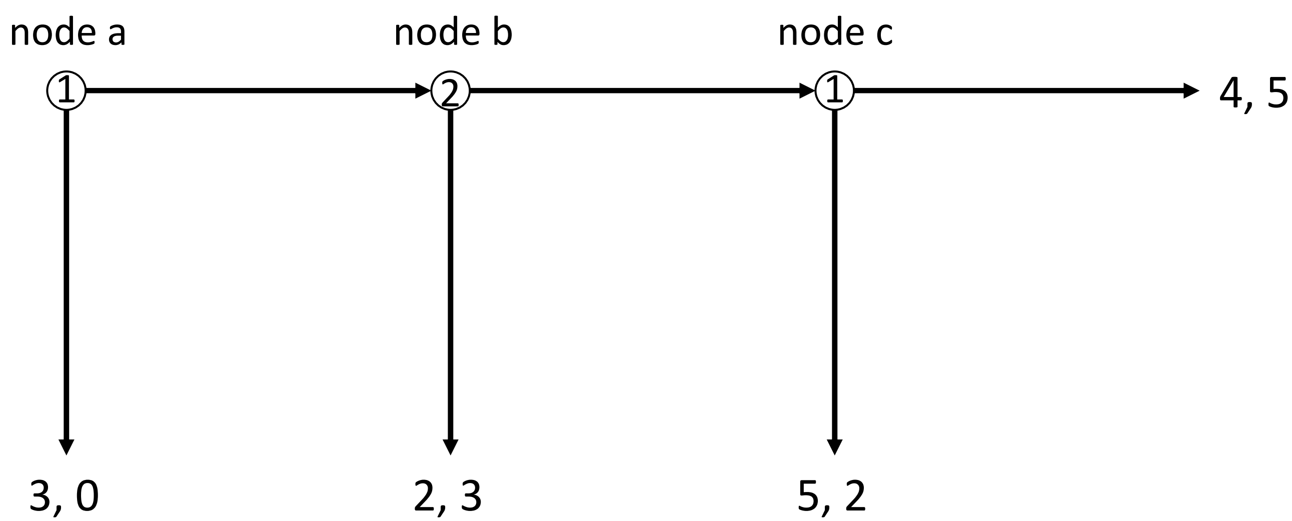
$$Rat_{i,P} = \{w \mid f_i(w) \in S_i(P) \text{ and } EU_{i,P}^{w}\big(f_i(w)\big) \geq EU_{i,P}^{w}(s_i), \text{for all } s_i \in S_i(P)\}$$

And player i is rational if she's rational at every information set reached by her chosen strategy.

$$Rat_i = \{w \mid w \in Rat_{i,P} \text{ for all } P \text{ with } f_i(w) \in S_i(P)\}$$

Define common belief in rationality (CBR) in the usual way

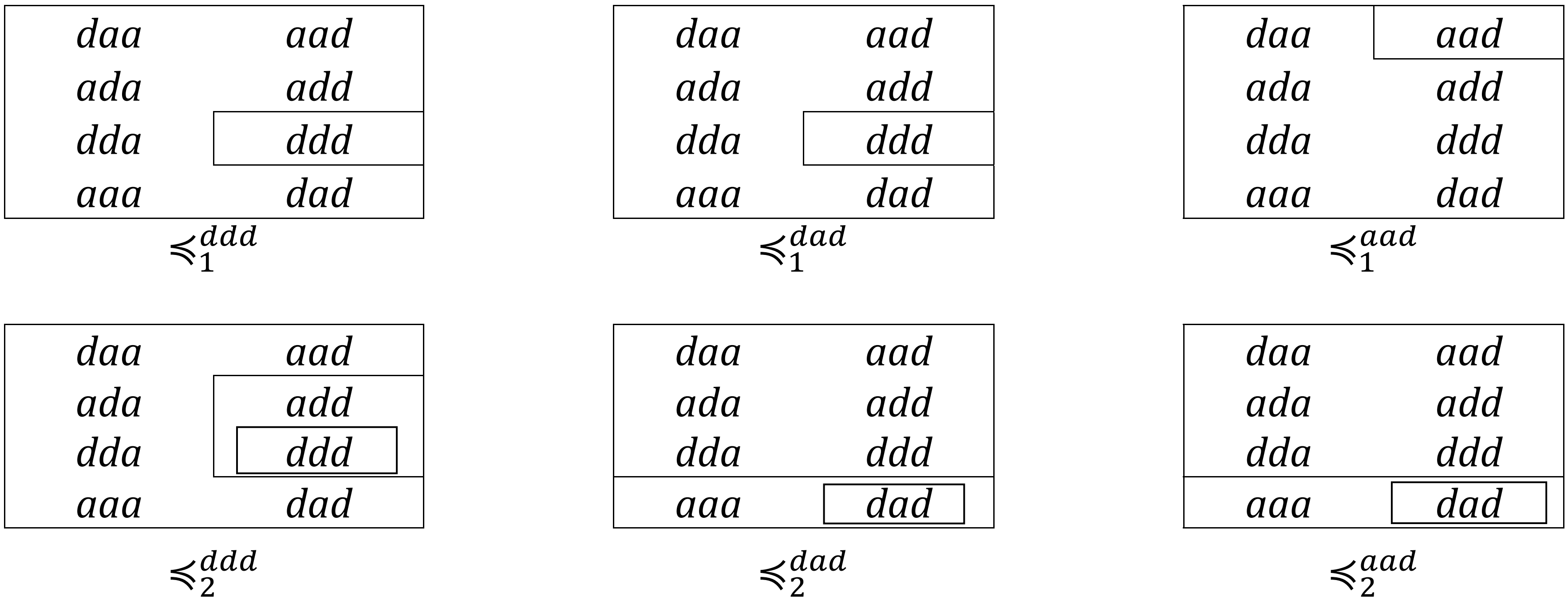
Example



$W = \{aaa, aad, ada, add, daa, dad, dda, ddd\}$

$p_1(w) = p_2(w) = \left(\frac{1}{8}\right)$ for all w

$f(aaa) = aaa$, etc.



First, notice $ddd \in Rat_1 \cap Rat_2$ (why?). Similarly, $dad \in Rat_1 \cap Rat_2$

What about aad? By the time node b is reached, player 2 revises his beliefs and thinks player 1 will move across at node c; hence it is rational to move across at node b. Player 1 believes that player 2 will do this, so it's rational for her to move across at node a (and of course down at node c). So, $aad \in Rat_1 \cap Rat_2$.

In fact, $\{aad, dad, ddd\} \subseteq CBR$

Implications of CBR

If common belief in rationality doesn't imply backward induction, what does it imply?

For every information set $P \in \mathcal{P}$, define the P-subgame to be the strategic form game G_P with strategy sets $S_i(P)$ for each player i, but the same (strategic form) utility functions as before.

Then, CBR characterizes the following iterated deletion process:

Round 1: in each P-subgame, delete any strictly dominated strategy for player $\iota(P)$

Round 2, 3,…: continue with IDSDS in whatever is left of the original game

	a	d
~~aa~~	~~4, 5~~	~~2, 3~~
ad	5, 2	2, 3
da	3, 0	3, 0
dd	3, 0	3, 0

$G_{\{\emptyset\}}$

	a	d
aa	4, 5	2, 3
ad	5, 2	2, 3

$G_{\{a\}}$

	a
~~aa~~	~~4, 5~~
ad	5, 2

$G_{\{aa\}}$

Rather disappointing!

Saving Backward Induction

What assumptions can we make about the players to guarantee that they will play the backward induction outcome?

The best rationalization principle (Battigalli GEB 1996)

> *A player should always believe that her opponents are implementing one of the "most rational"* (*or "least irrational"*) *strategy profiles which are consistent with her information*.

Basic idea: if possible, assume that your opponent has beliefs such that her previous move(s) were rational.

And if possible, assume that she is rational and believes that you're rational…

In the game we just analyzed, "across" by player 1 is consistent with 1's rationality, with 1's belief in 2's rationality, but not with 1's belief in 2's belief in 1's rationality.

Forward Induction and Backward Induction

The best rationalization principle captures forward induction reasoning.

How can forward induction reasoning give us the backward induction outcome?

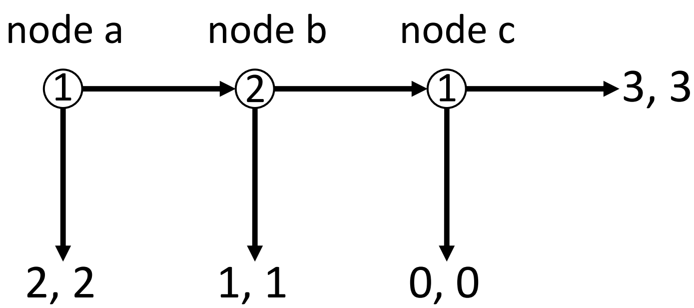
Forward induction reasoning makes inferences from players' past moves about what kind of people they are (e.g. you did this, so you can't be rational; you did that, so you could be rational)

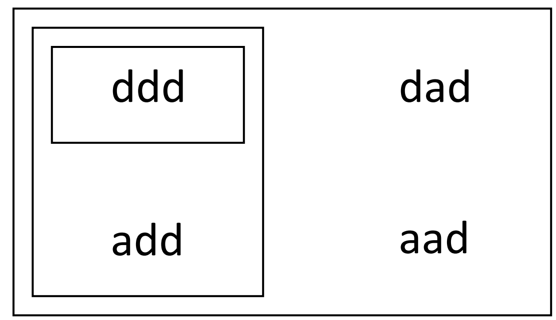
Backward induction makes inferences from what kind of people your opponents are to what they will do in the future (e.g. you're rational, so you'll play "down" next time)

Past moves $\xrightarrow[\text{FI}]{}$ Beliefs about your rationality $\xrightarrow[\text{BI}]{}$ Beliefs about your future moves

Forward Induction and Rich Models

To formalize the best rationalization principle, we need to work with models that have "enough" states





$\preccurlyeq_1^{ddd}$ *and* $\preccurlyeq_2^{ddd}$

In this model, there is no way to interpret "across" by player 1 as a rational move

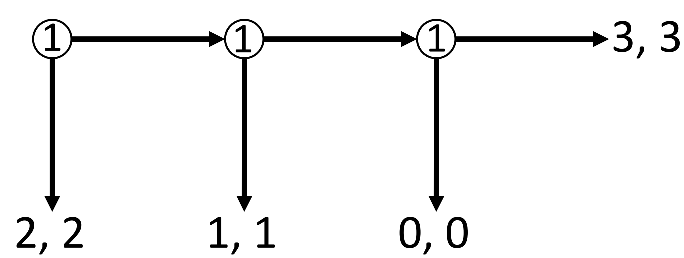
Board (GEB 2004) proves the existence of a model that is "belief complete"

Food for Thought

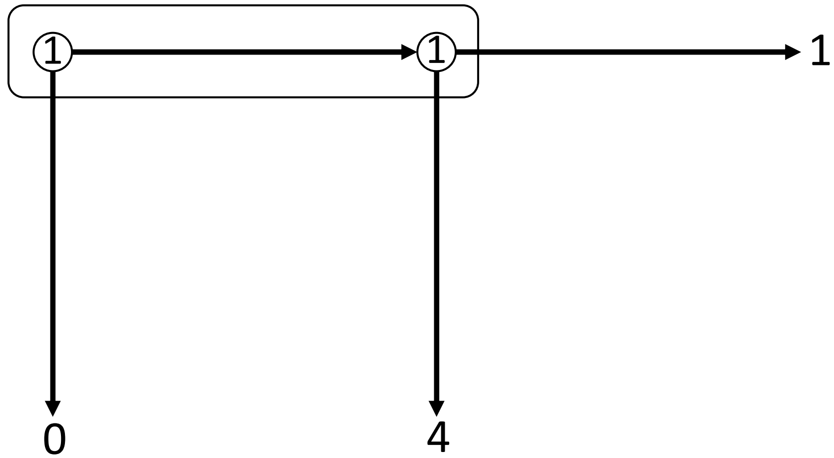
1. Actions vs. strategies

In our definition of rationality, we assumed that at a given information set, the player on move chose optimally among all strategies that reached the information set. Implicitly, this definition assumes she can control her future as well as her current actions.

If only the current action is chosen, a rational player (with common belief in rationality) may move "down" at the first node in this game:



2. Absent-mindedness



(a) How should we analyze games with absentmindedness?

(b) Mixed strategies vs. behavior strategies.

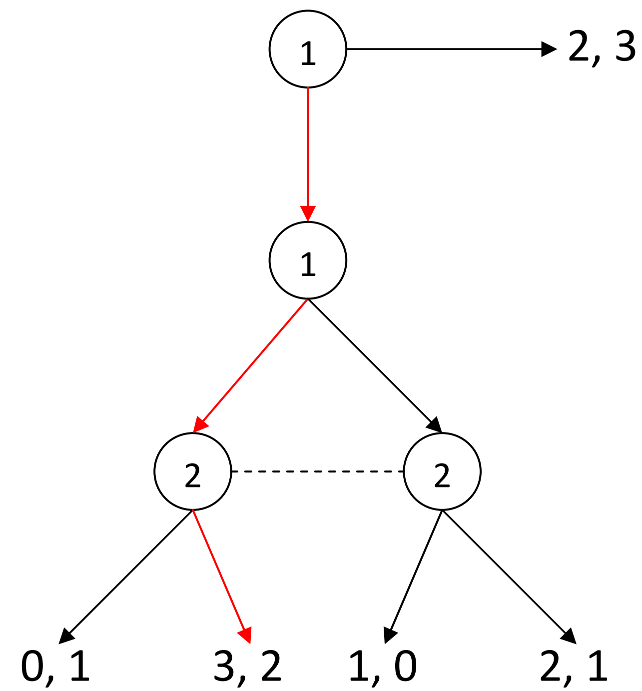
(c) Does the player make one or two choices?

(d) Should the two roles of information sets (defining what the players know and defining strategies) be separated?

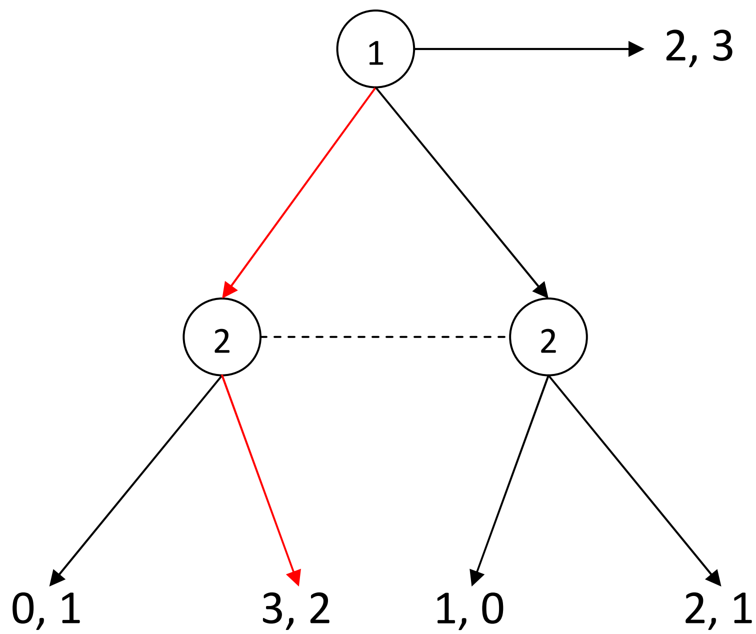
3. Extensive form solution concepts

Can we provide an epistemic characterization of the following solution concepts?

(a) Subgame perfect equilibrium



(b) Perfect Bayesian equilibrium



(c) The intuitive criterion