MS&E 233
Game Theory, Data Science and AI
Lecture 5

Vasilis Syrgkanis

Assistant Professor

Management Science and Engineering

(by courtesy) Computer Science and Electrical Engineering

Institute for Computational and Mathematical Engineering

Computational Game Theory for Complex Games

(1)
- Basics of game theory and zero-sum games (T)
- Basics of online learning theory (T)
- Solving zero-sum games via online learning (T)
- *HW1: implement simple algorithms to solve zero-sum games*
- Applications to ML and AI (T+A)
- *HW2: implement boosting as solving a zero-sum game*

(2)
- **Basics of extensive-form games**
- Solving extensive-form games via online learning (T)
- *HW3: implement agents to solve very simple variants of poker*

(3)
- General games and equilibria (T)
- Online learning in general games, multi-agent RL (T+A)
- *HW4: implement no-regret algorithms that converge to correlated equilibria in general games*

Data Science for Auctions and Mechanisms

(4)
- Basics and applications of auction theory (T+A)
- Learning to bid in auctions via online learning (T)
- *HW5: implement bandit algorithms to bid in ad auctions*

(5)
- Optimal auctions and mechanisms (T)
- Simple vs optimal mechanisms (T)
- *HW6: calculate equilibria in simple auctions, implement simple and optimal auctions, analyze revenue empirically*

(6)
- Optimizing mechanisms from samples (T)
- Online optimization of auctions and mechanisms (T)
- *HW7: implement procedures to learn approximately optimal auctions from historical samples and in an online manner*

Further Topics

(7)
- Econometrics in games and auctions (T+A)
- A/B testing in markets (T+A)
- *HW8: implement procedure to estimate values from bids in an auction, empirically analyze inaccuracy of A/B tests in markets*

Guest Lectures

- Mechanism Design for LLMs, Renato Paes Leme, Google Research
- Auto-bidding in Sponsored Search Auctions, Kshipra Bhawalkar, Google Research

Extensive Form Games

History and Progress

Historical Challenge in Game Theory and AI

Nash1950

A Three-Man Poker Game

As an example of the application of our theory to a more or less realistic case we include the simplified poker game given below. The rules are as follows:

(a) The deck is large, with equally many *high* and *low* cards, and a hand consists of one card.

(b) Two chips are used to ante, open, or call.

(c) The players play in rotation and the game ends after all have passed or after one player has opened and the others have had a chance to call.

(d) If no one bets the antes are retrieved.

(e) Otherwise the pot is divided equally among the highest hands which have bet.

Kuhn1950

A SIMPLIFIED TWO-PERSON POKER [*]

H. W. Kuhn[1]

A fascinating problem for the game theoretician is posed by the common card game, Poker. While generally regarded as partaking of psychological aspects (such as bluffing) which supposedly render it inaccessible to mathematical treatment, it is evident that Poker falls within the general theory of games as elaborated by von Neumann and Morgenstern [1]. Relevant probability problems have been considered by Borel and Ville [2] and several variants are examined by von Neumann [1] and by Bellman and Blackwell [3].
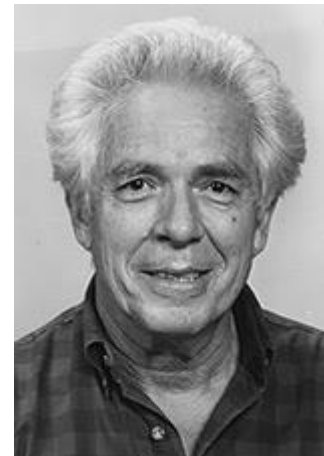
Waterman1970

Generalization Learning Techniques for Automating the Learning of Heuristics[1]

D. A. Waterman

Carnegie-Mellon University, Pittsburgh, Pennsylvania

Many Recent Success Stories

Science

RESEARCH ARTICLE

Superhuman AI for heads-up no-limit poker: Libratus beats top professionals

NOAM BROWN AND TUOMAS SANDHOLM   Authors Info & Affiliations

Science

RESEARCH ARTICLE | MACHINE LEARNING

Mastering the game of Stratego with model-free multiagent reinforcement learning

JULIEN PEROLAT , BART DE VYLDER , DANIEL HENNES , EUGENE TARASSOV , [...], AND KARL TUYLS   +29 authors   Authors Info & Affiliations

Science

REPORT

A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play

DAVID SILVER , THOMAS HUBERT , JULIAN SCHRITTWIESER , IOANNIS ANTONOGLOU , [...], AND DEMIS HASSABIS   +8 authors   Authors Info & Affiliations
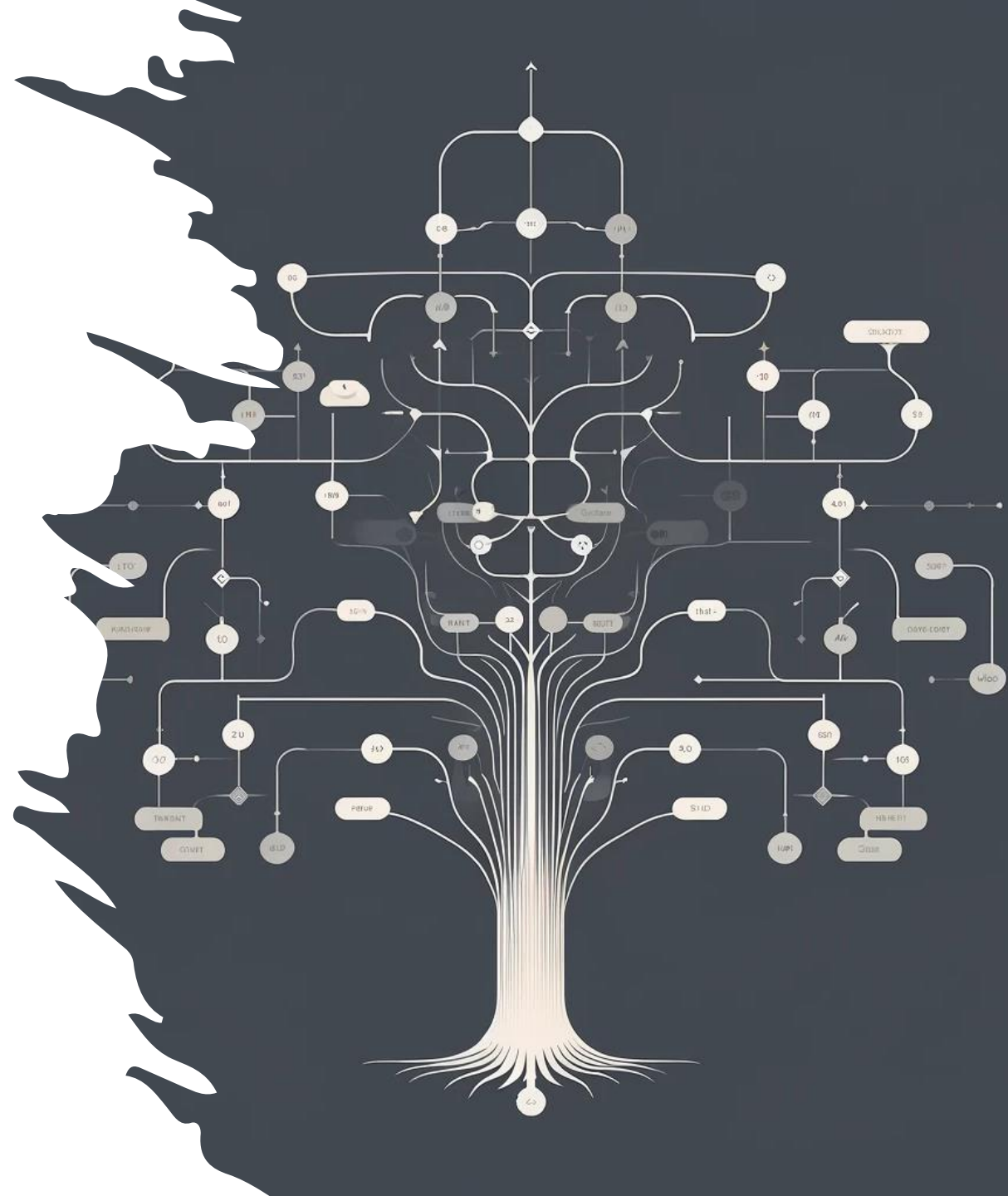
Key Elements to Success

New approaches to approximate the "continuation value of the game" via deep learning and other domain specific techniques

Scalable algorithmic methods to compute approximate Nash equilibria of zero-sum games via learning dynamics

# Extensive Form Games

The Basics
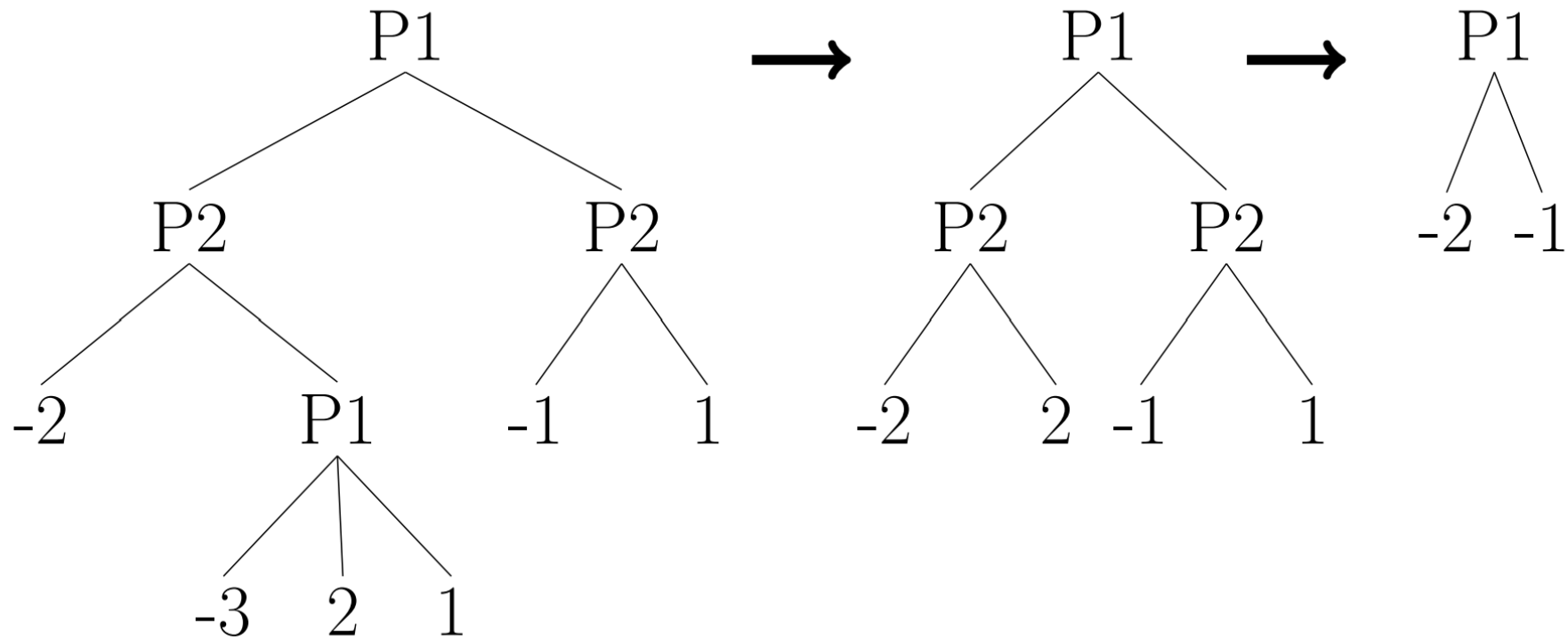
Perfect Information Games

- Players take turns in choosing actions
- All actions are publicly observable
- The "state" of the game is publicly observable
- Some sequence of actions lead to terminal states
- Each player receives some utility/loss at a terminal state
- In zero-sum games: utility of player 1 equals loss of player 2

Tree Representation
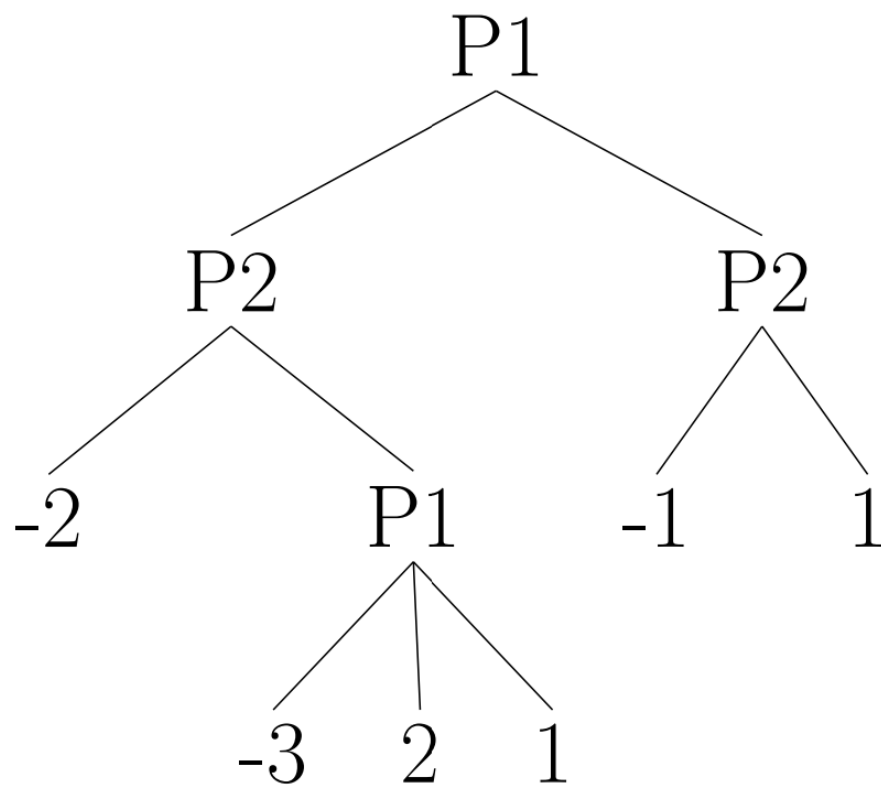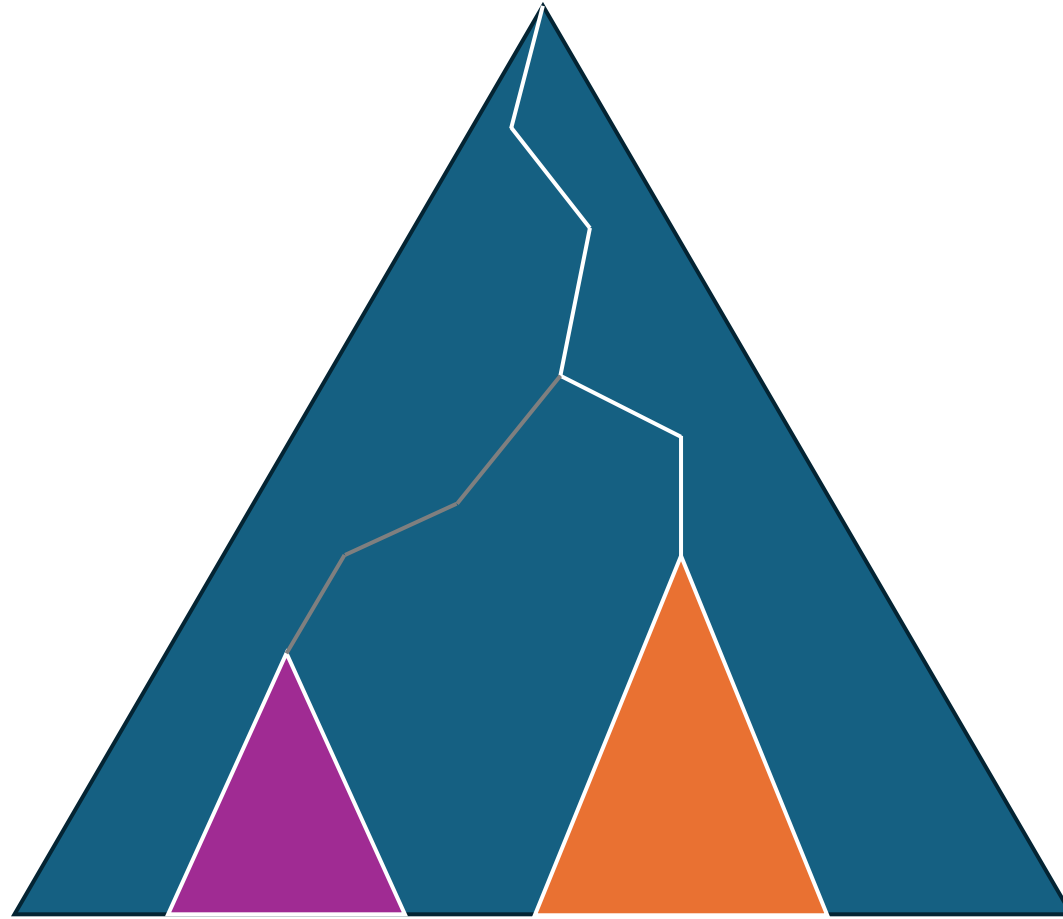
Solving Games via Backwards Induction

Imperfect Information Games

Players don't have perfect knowledge about the "state" of the game
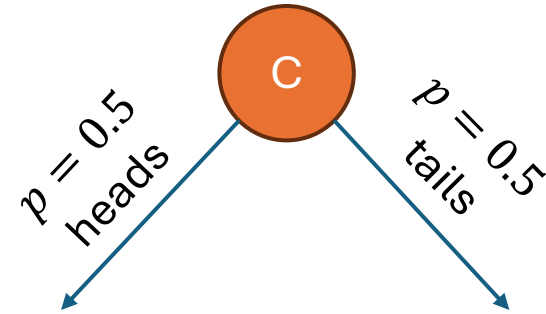
Why are Imperfect Information Games Hard

The optimal strategy in the orange sub-tree can depend on how we play and what happens in the purple sub-tree

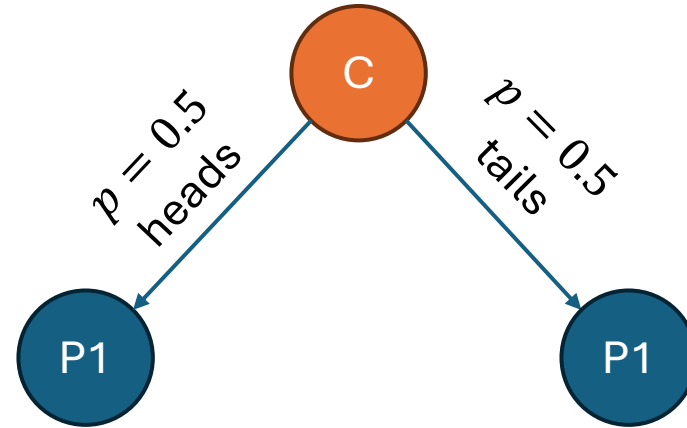# A Simple Game

Rules of the game

- Nature (chance) flips a coin
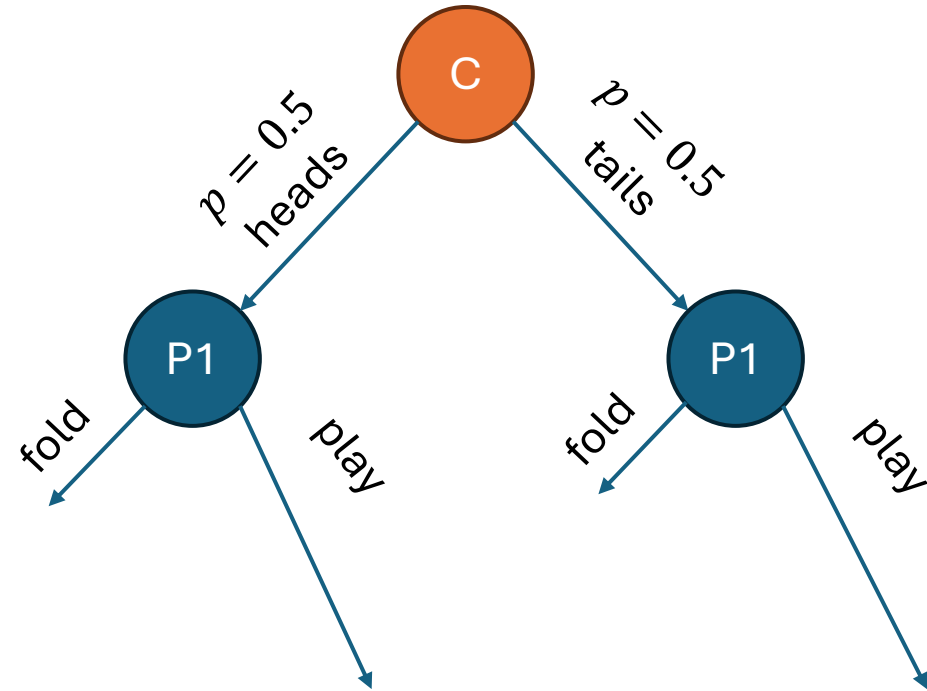
A Simple Game

Rules of the game

- Nature (chance) flips a coin

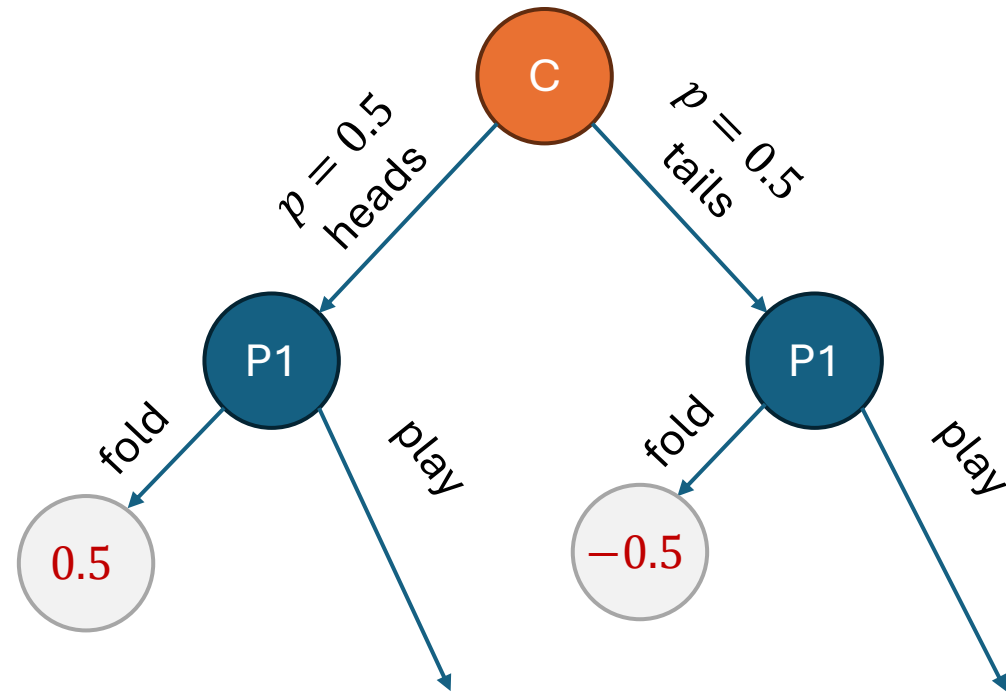- Player one sees the outcome of the coin

A Simple Game

Rules of the game

- Nature (chance) flips a coin
- Player one sees the outcome of the coin
- Player one chooses whether to fold or play

A Simple Game

- Nature (chance) flips a coin
- Player one sees the outcome of the coin
- Player one chooses whether to fold or play
- If they fold with heads they win 0.5 if they fold with tails they lose 0.5

A Simple Game

- Nature (chance) flips a coin

- Player one sees the outcome of the coin

- Player one chooses whether to fold or play

- If they fold with heads they win 0.5 if they fold with tails they lose 0.5

- If they play then it is player two turn

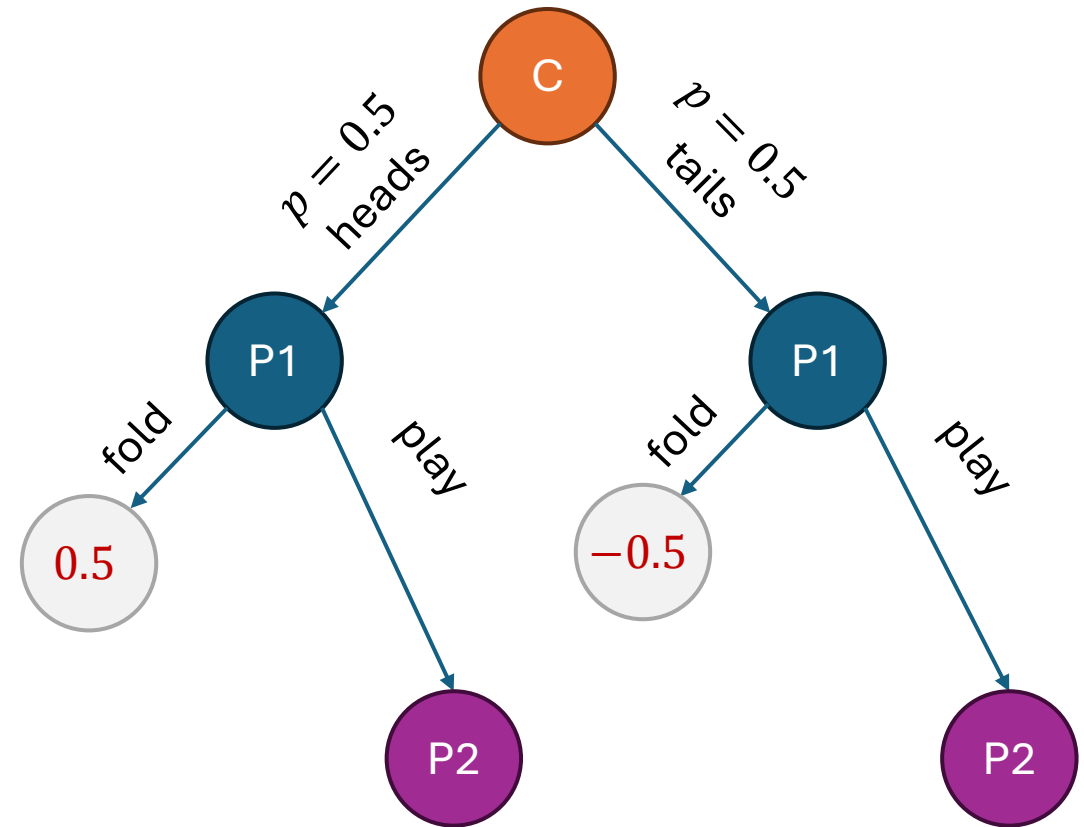# A Simple Game

Rules of the game

- Nature (chance) flips a coin

- Player one sees the outcome of the coin

- Player one chooses whether to fold or play

- If they fold with heads they win 0.5 if they fold with tails they lose 0.5

- If they play then it is player two turn

- Player two doesn't see the outcome of the coin
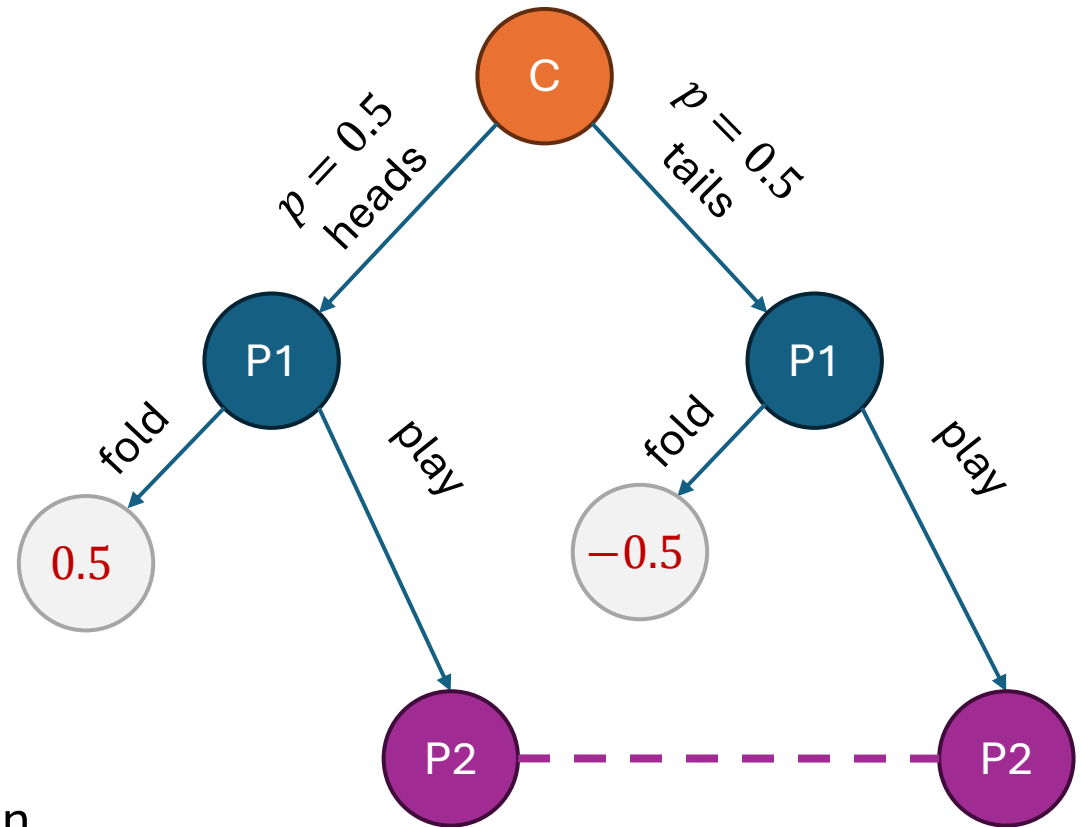
A Simple Game

Rules of the game

- Nature (chance) flips a coin

- Player one sees the outcome of the coin

- Player one chooses whether to fold or play

- If they fold with heads they win 0.5 if they fold with tails they lose 0.5

- If they play then it is player two turn

- Player two doesn't see the outcome of the coin

- They choose either heads or tails

A Simple Game
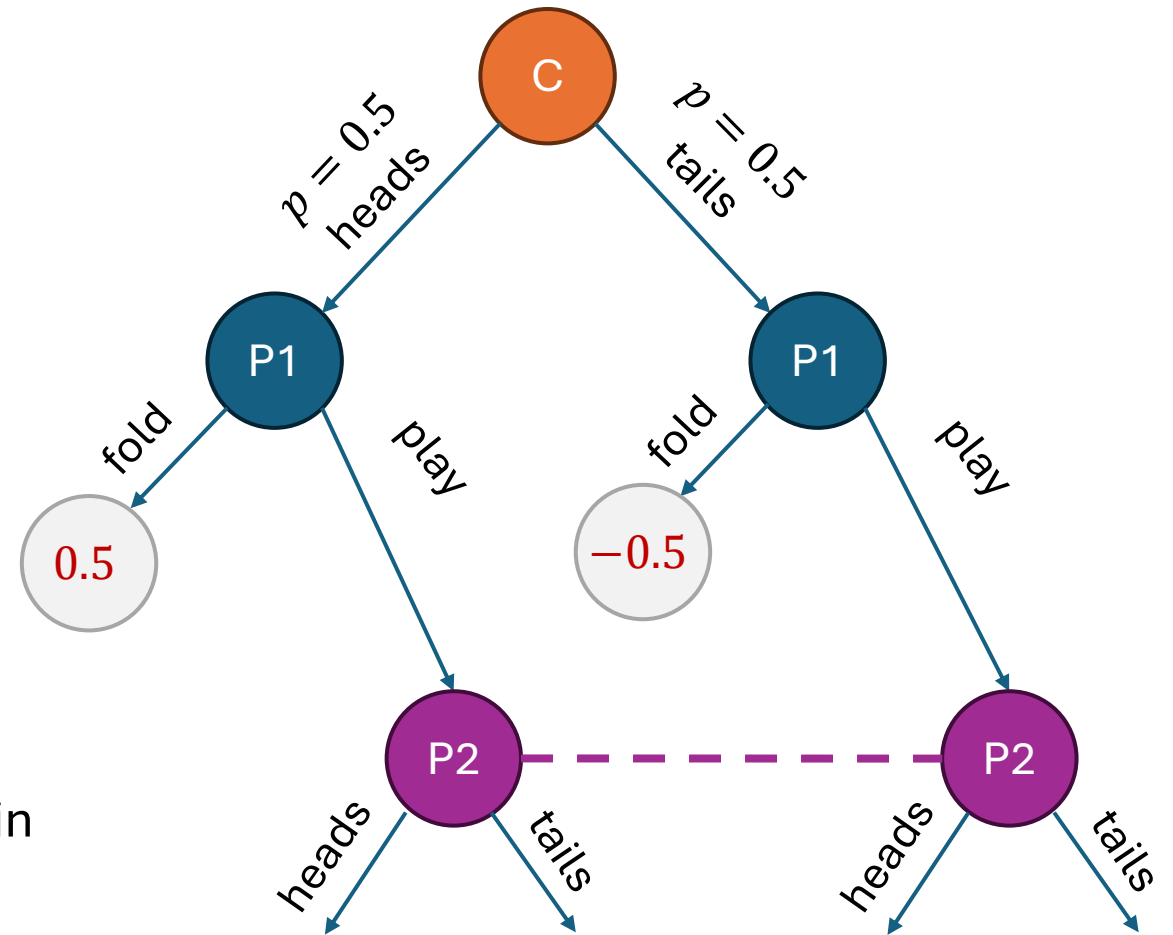
Rules of the game

- Nature (chance) flips a coin

- Player one sees the outcome of the coin

- Player one chooses whether to fold or play

- If they fold with heads they win 0.5 if they fold with tails they lose 0.5

- If they play then it is player two turn

- Player two doesn't see the outcome of the coin

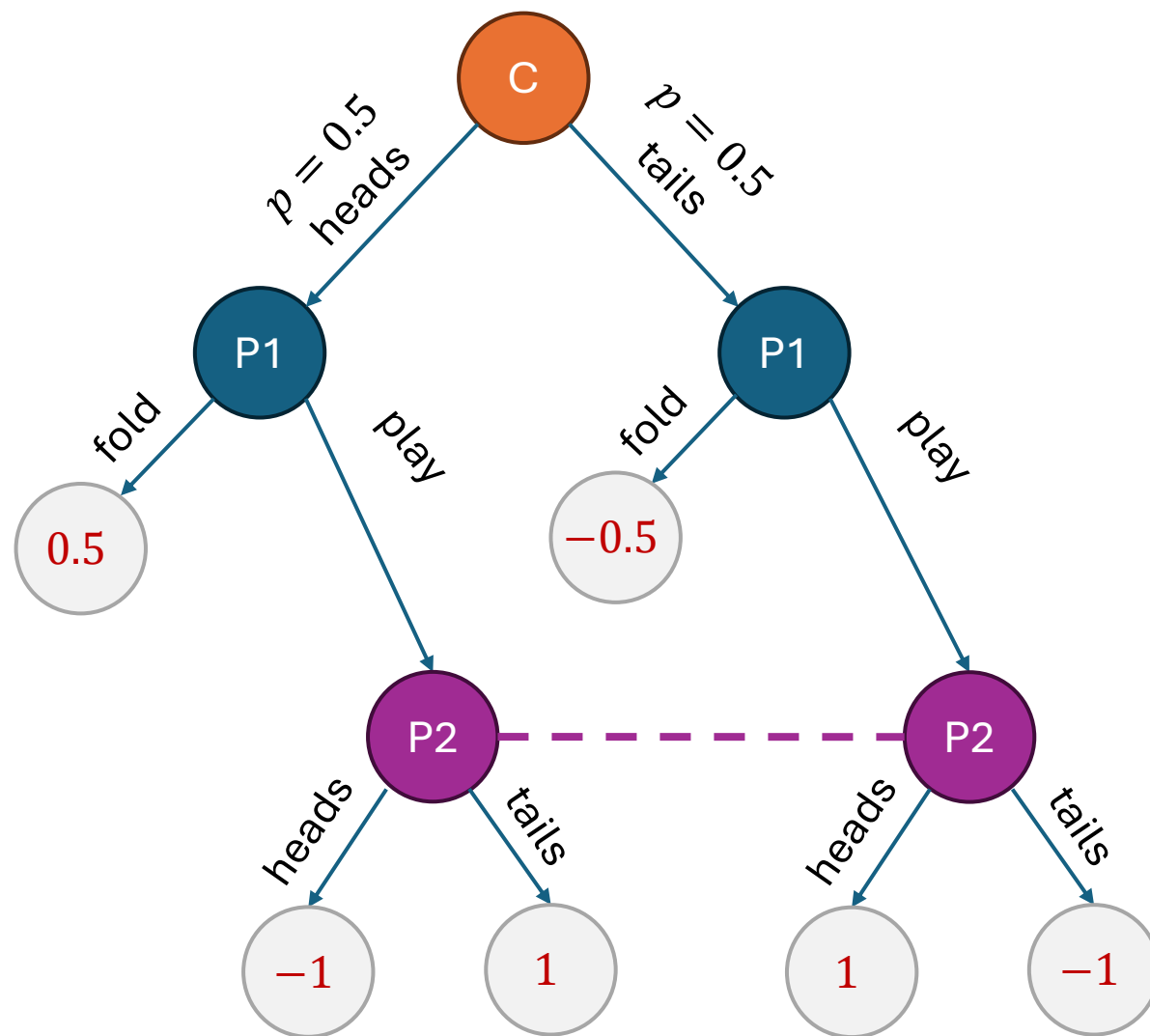- They choose either heads or tails

- If they match the coin they win 1 (P1 loses 1) if they don't match they lose 1 (P1 wins 1)
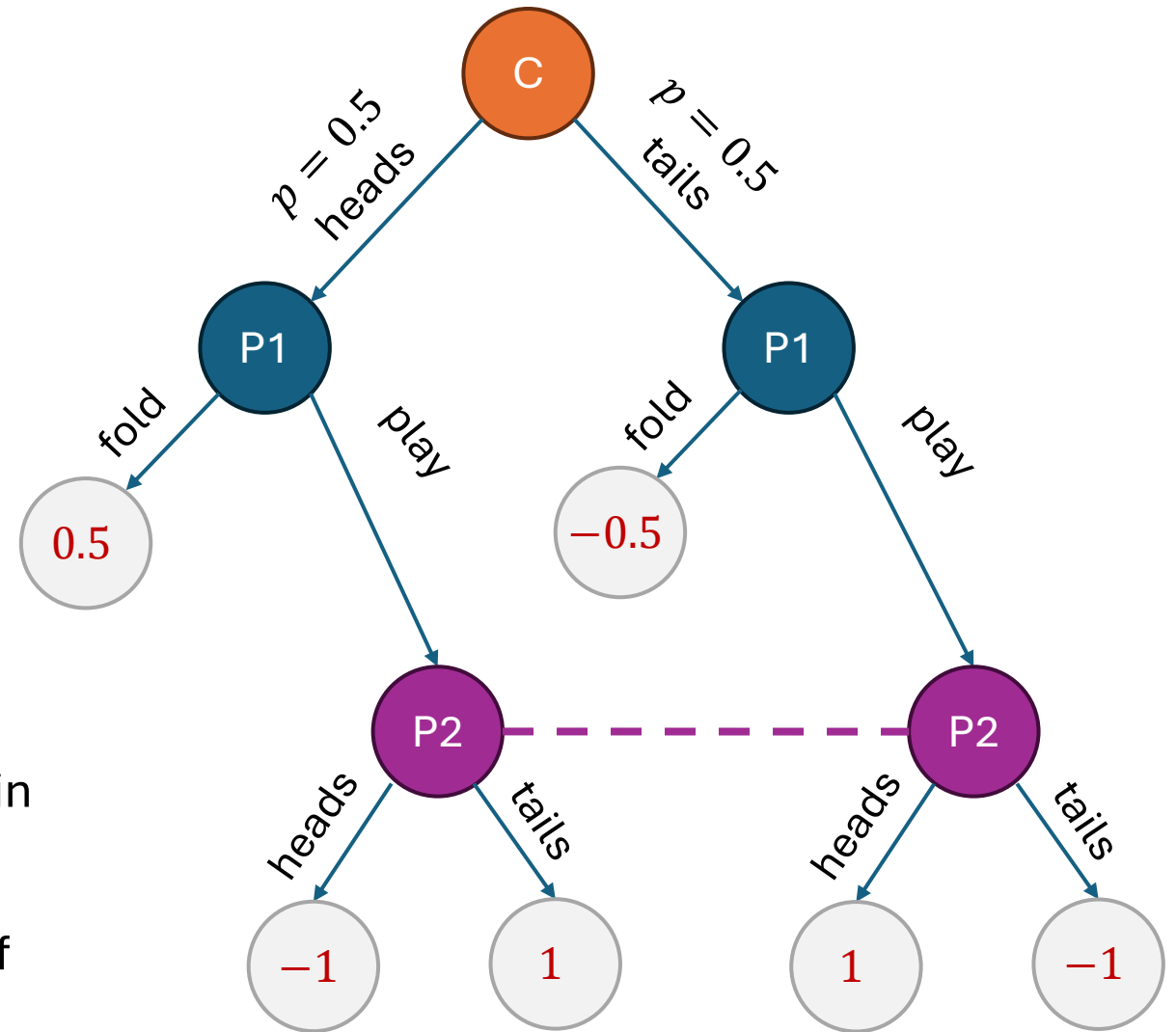
A Simple Game

A Simple Game
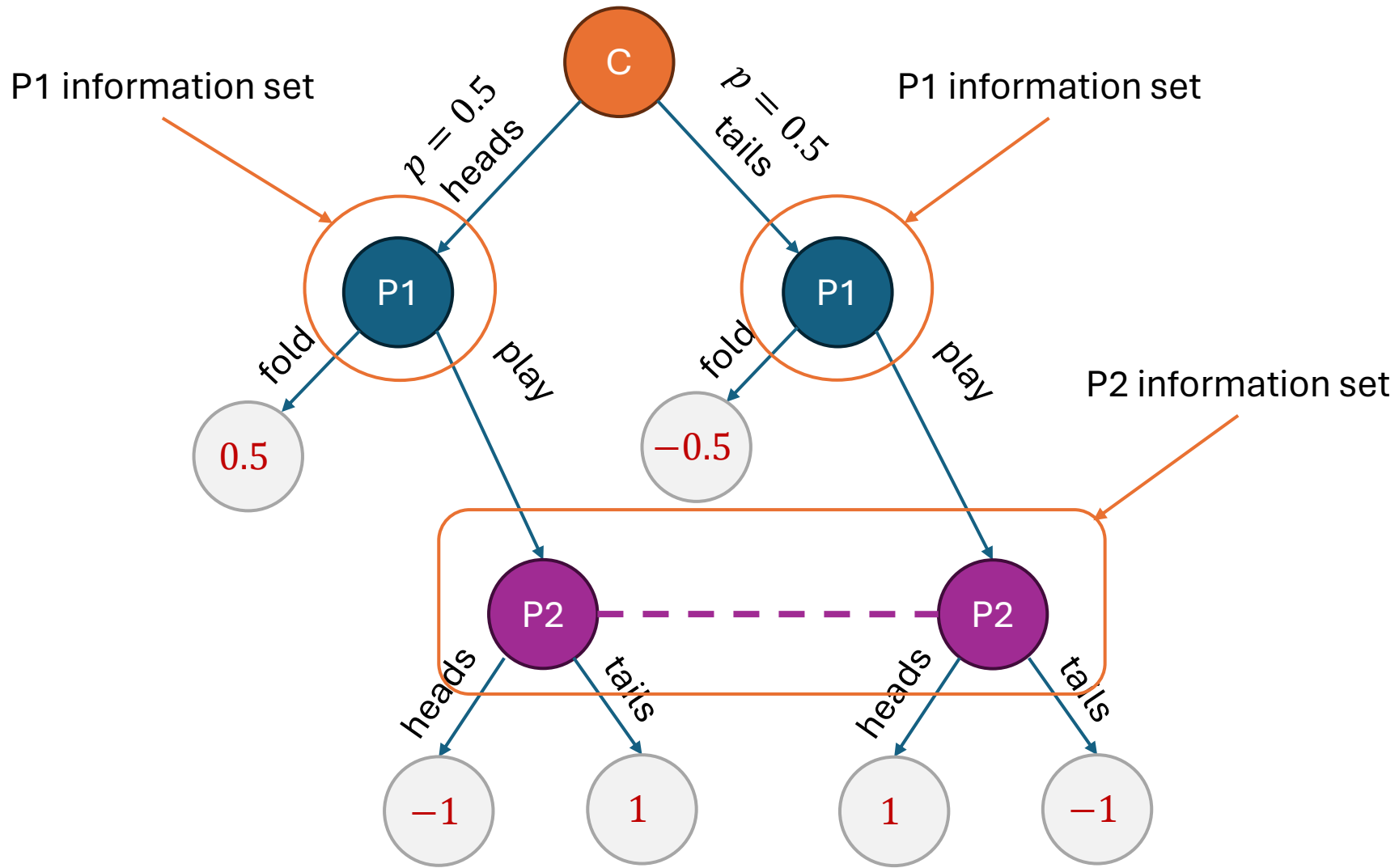
A Simple Game
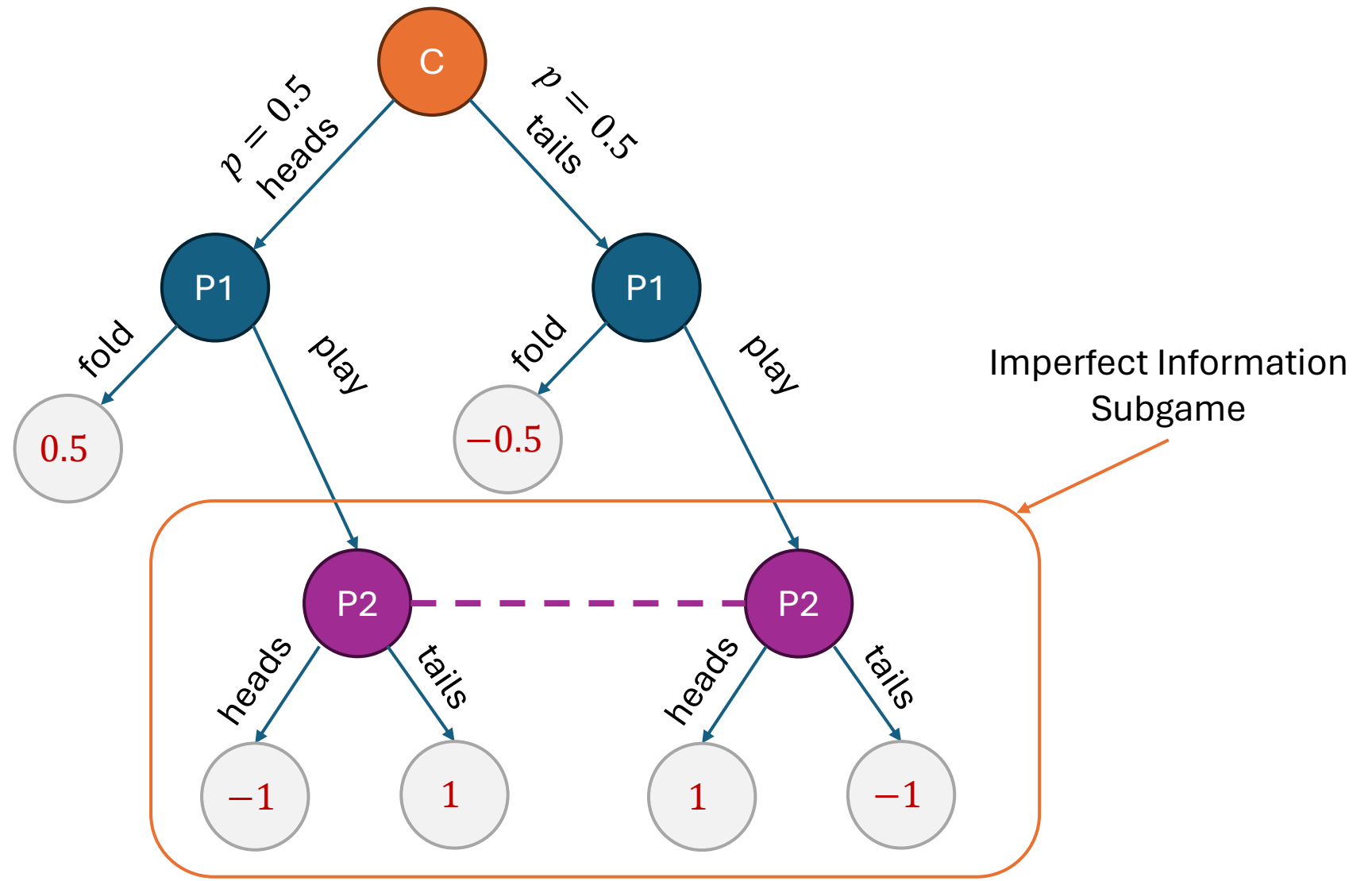
How should P2 play in the sub-game?
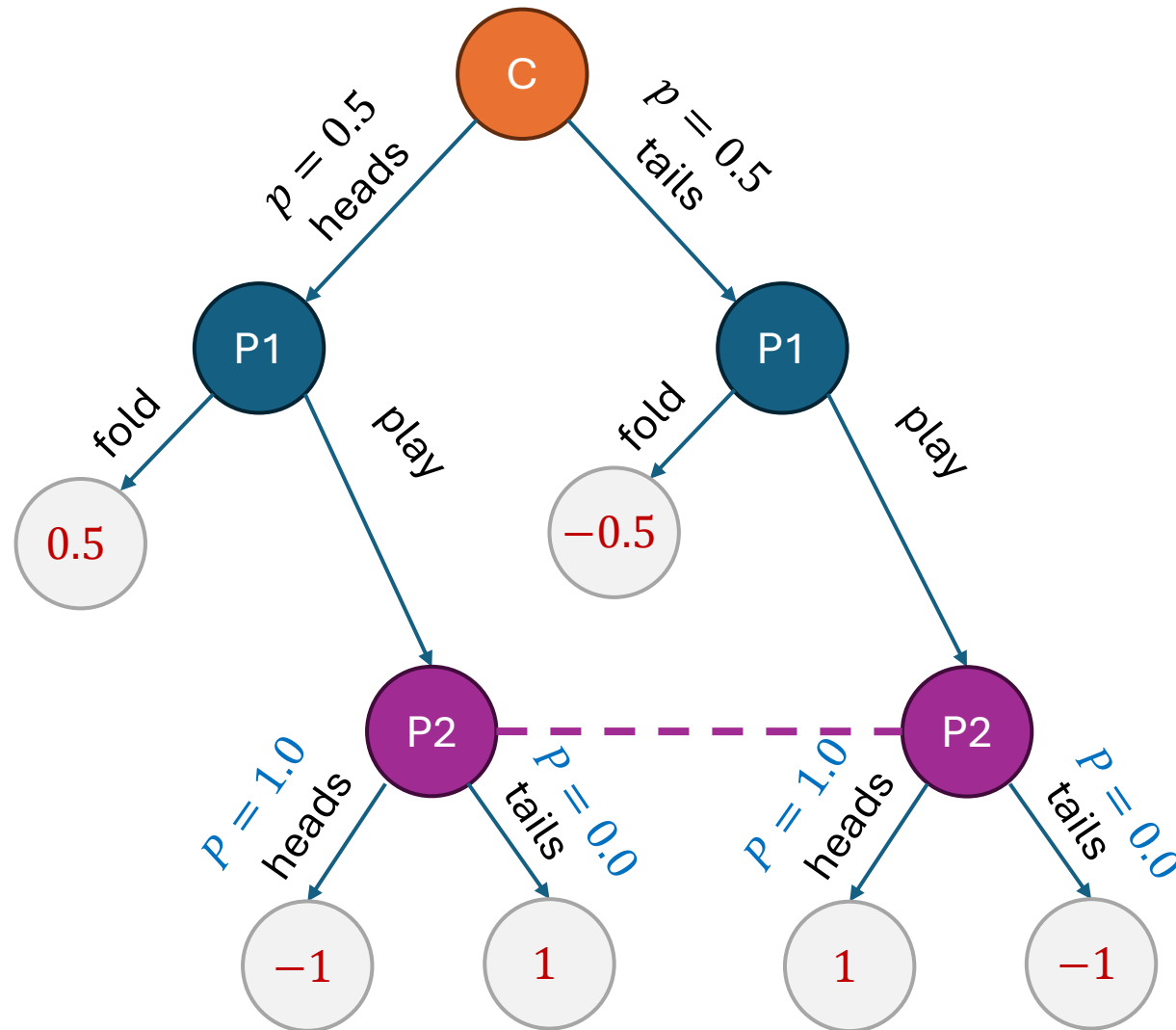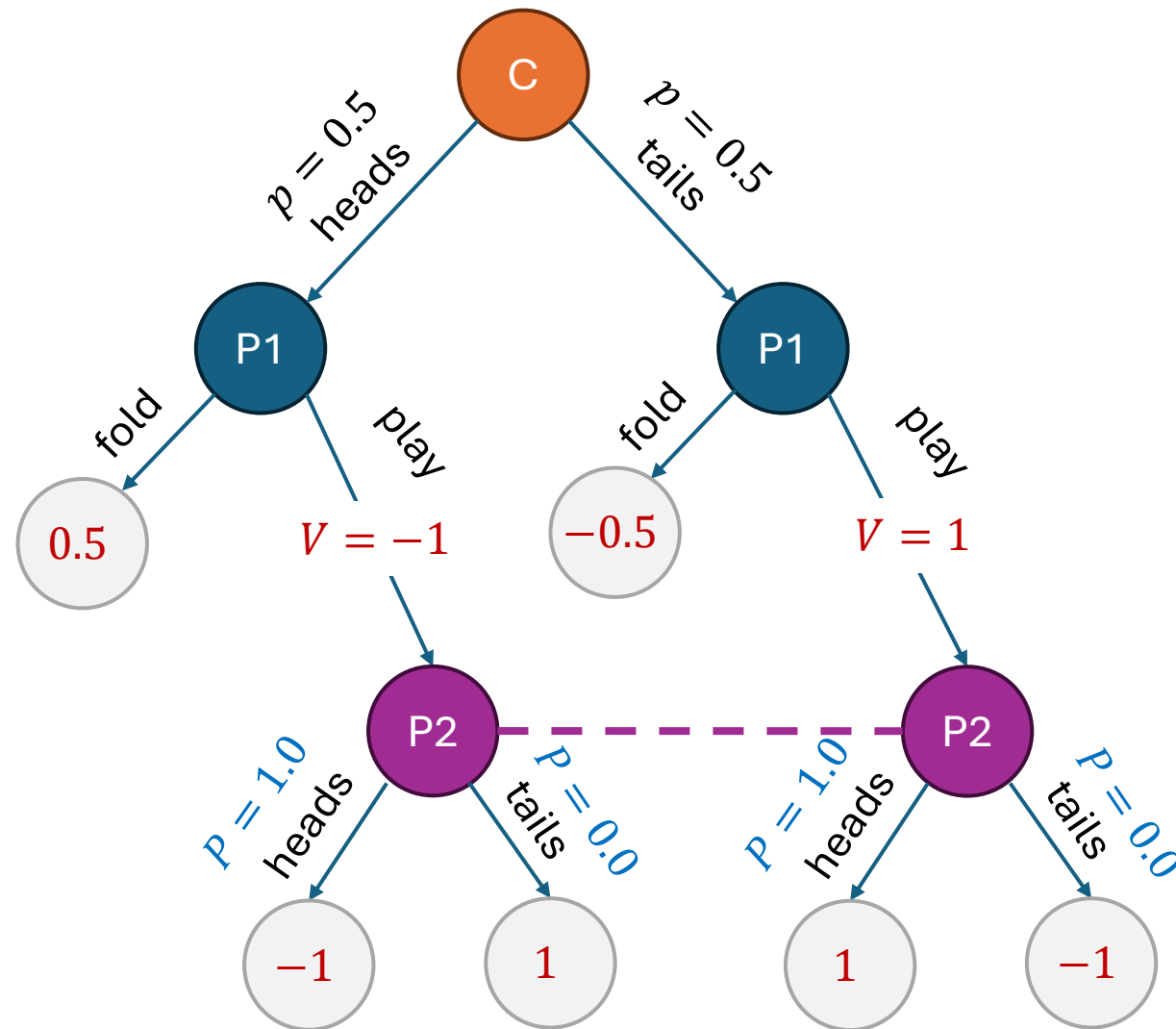
How should P2 play in the sub-game?

How should P2 play in the sub-game?

How should P2 play in the sub-game?

How should P2 play in the sub-game?
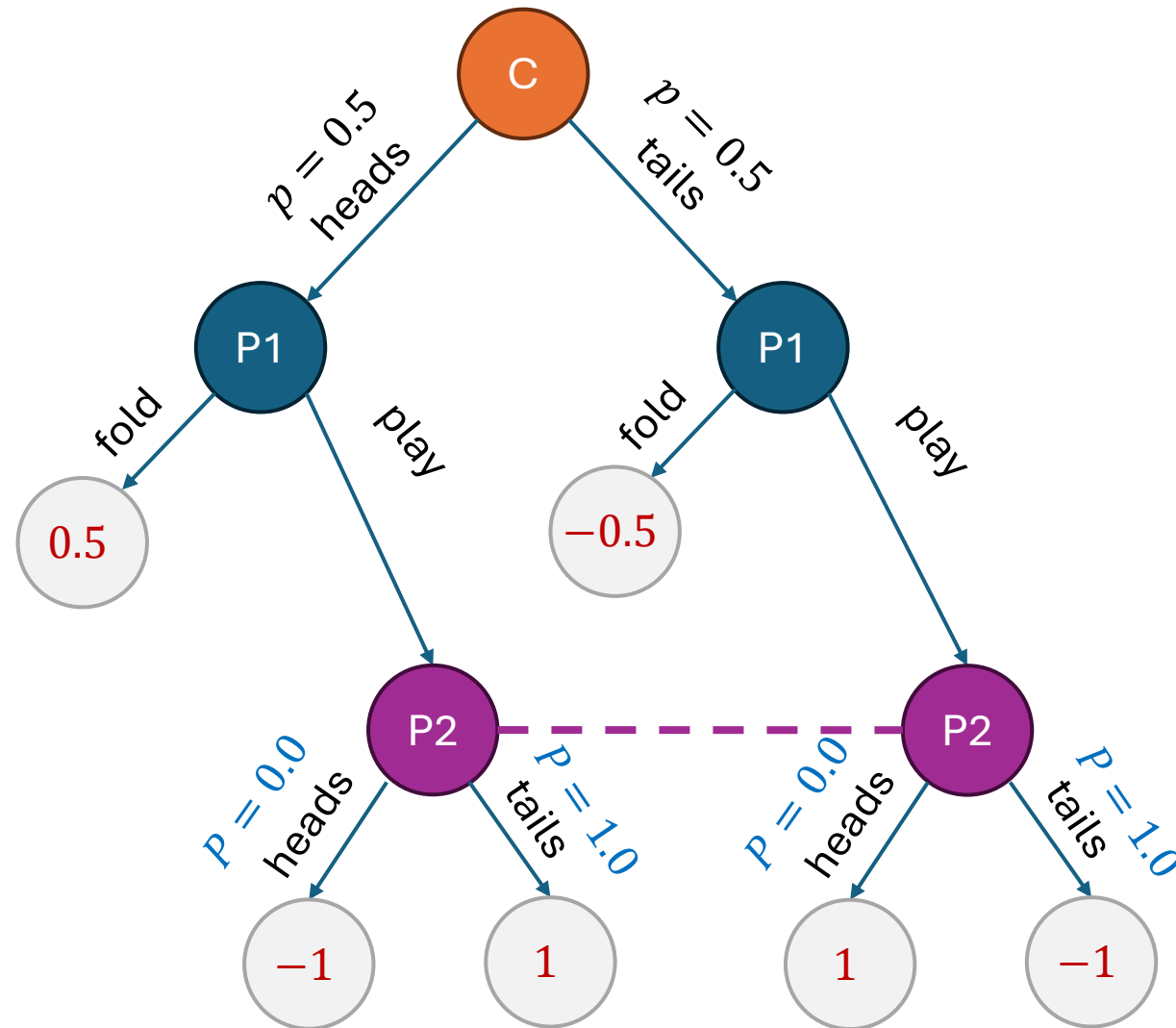
How should P2 play in the sub-game?
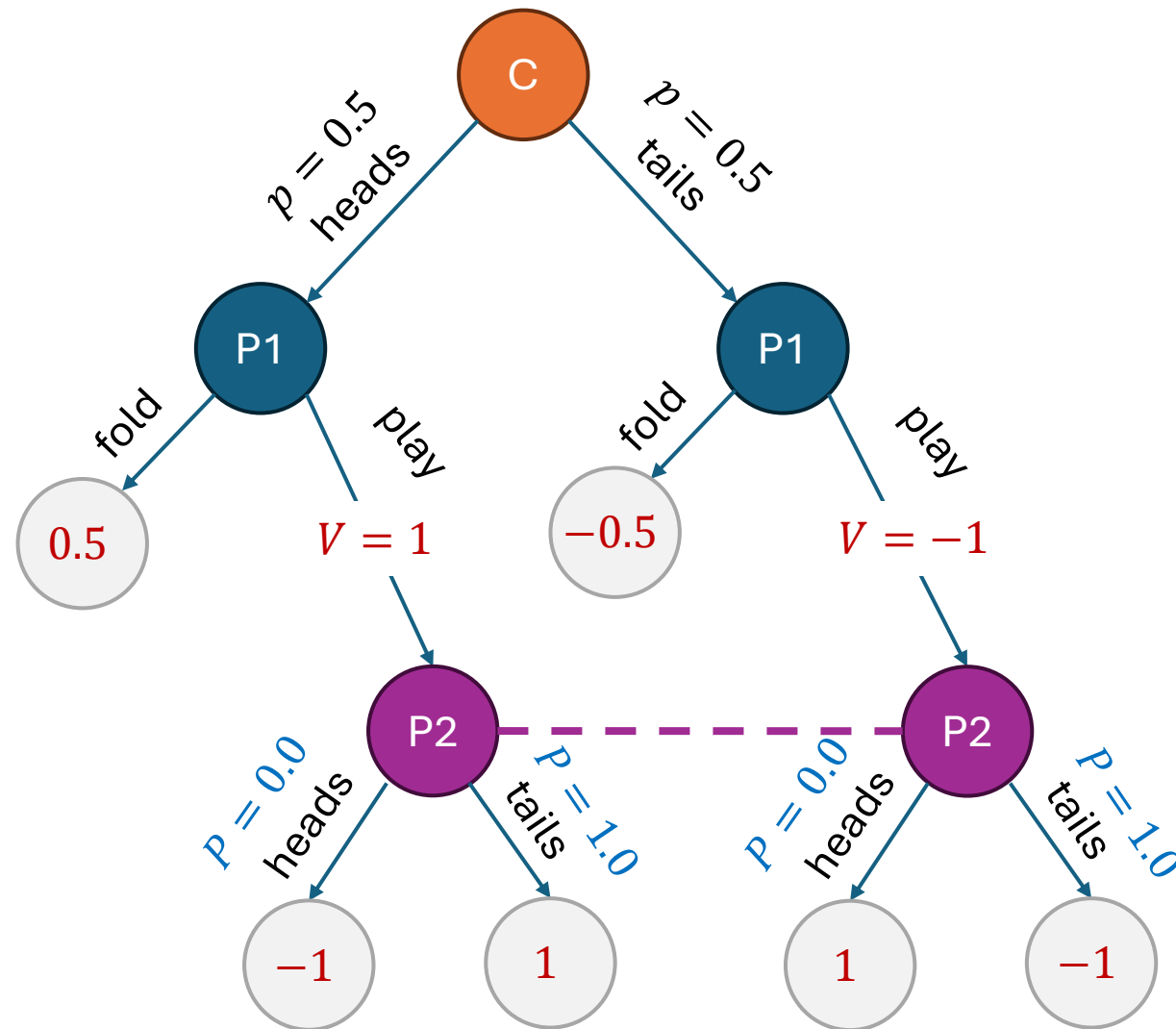
How should P2 play in the sub-game?
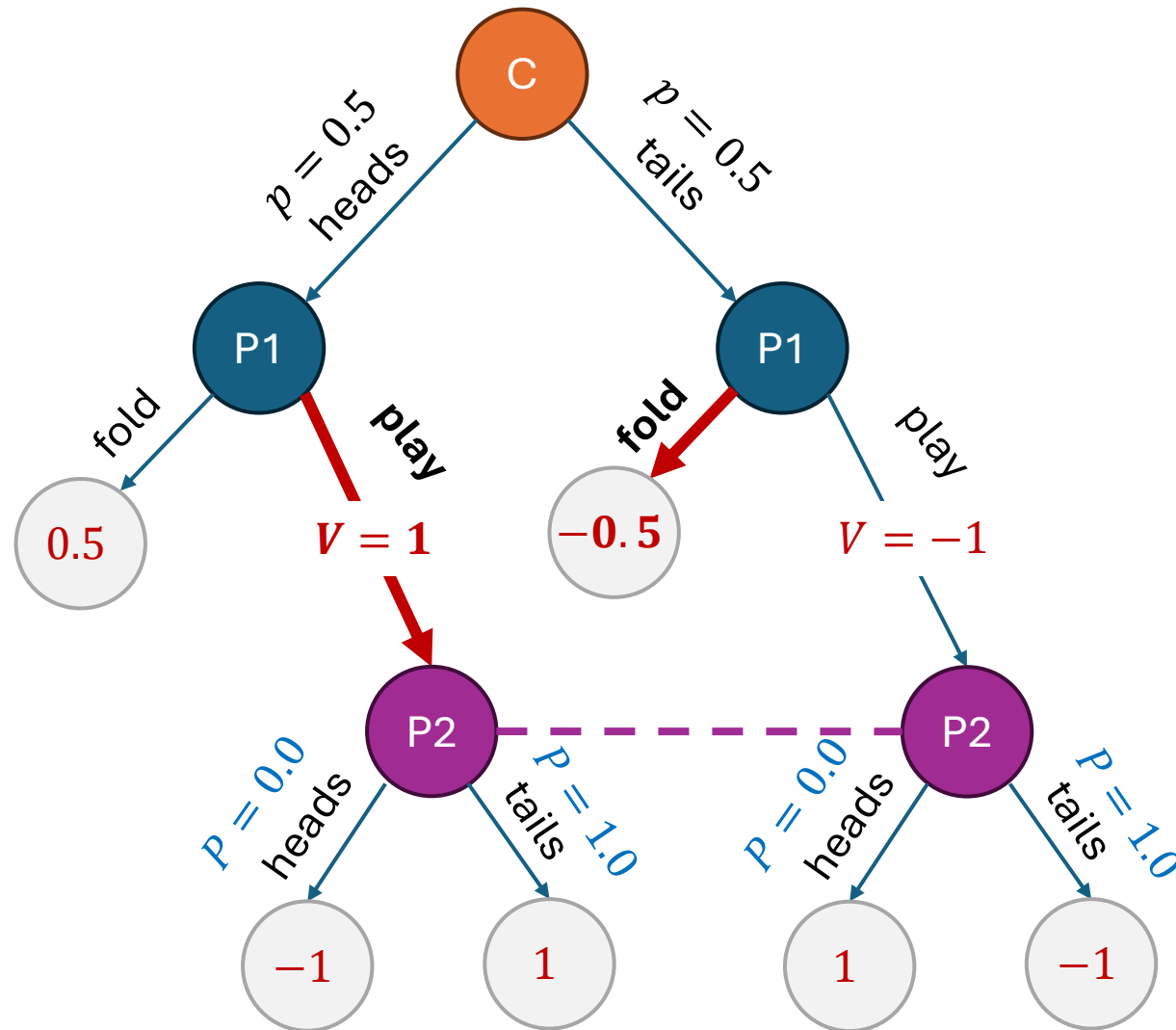
How should P2 play in the sub-game?
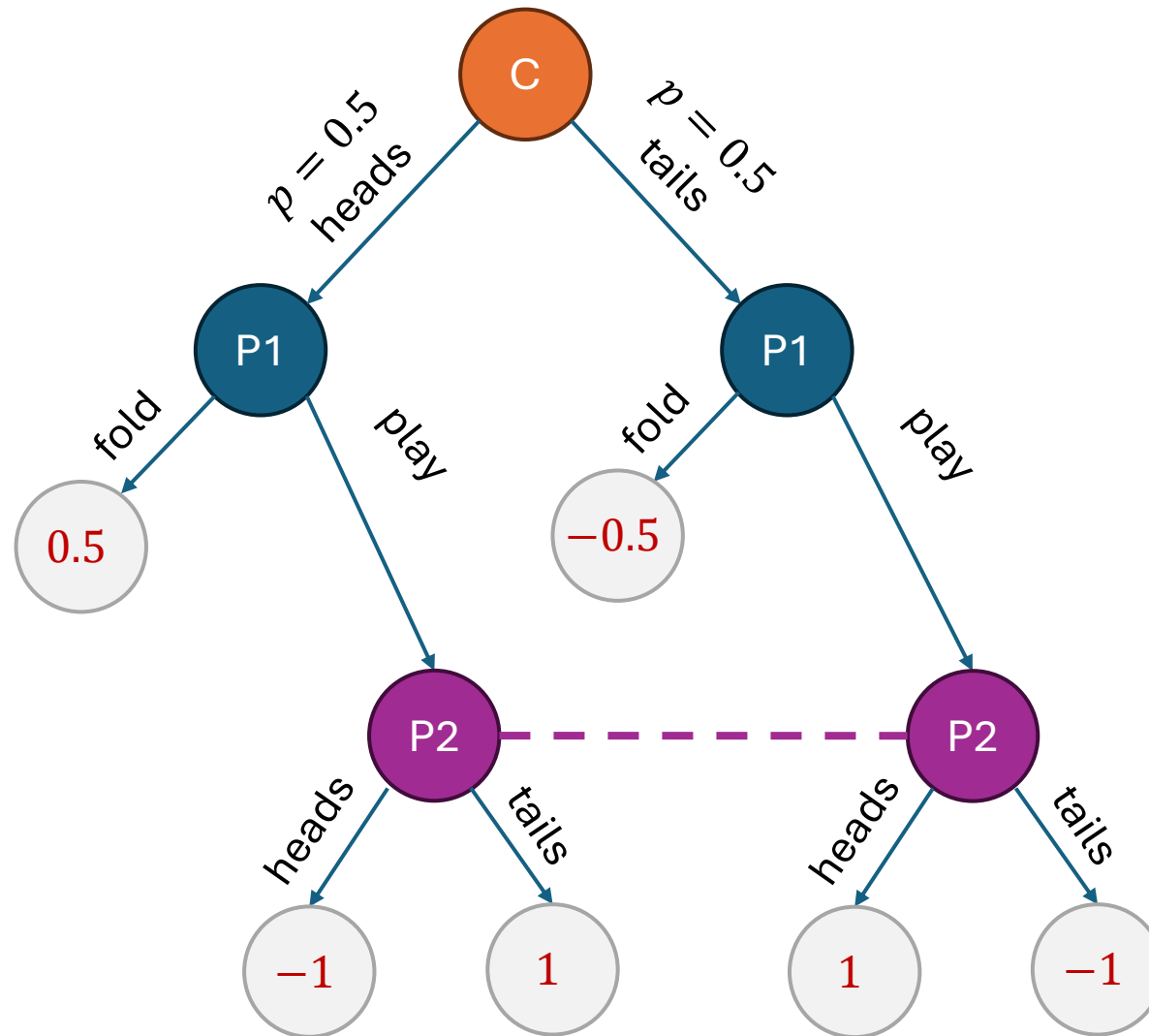
How should P2 play in the sub-game?
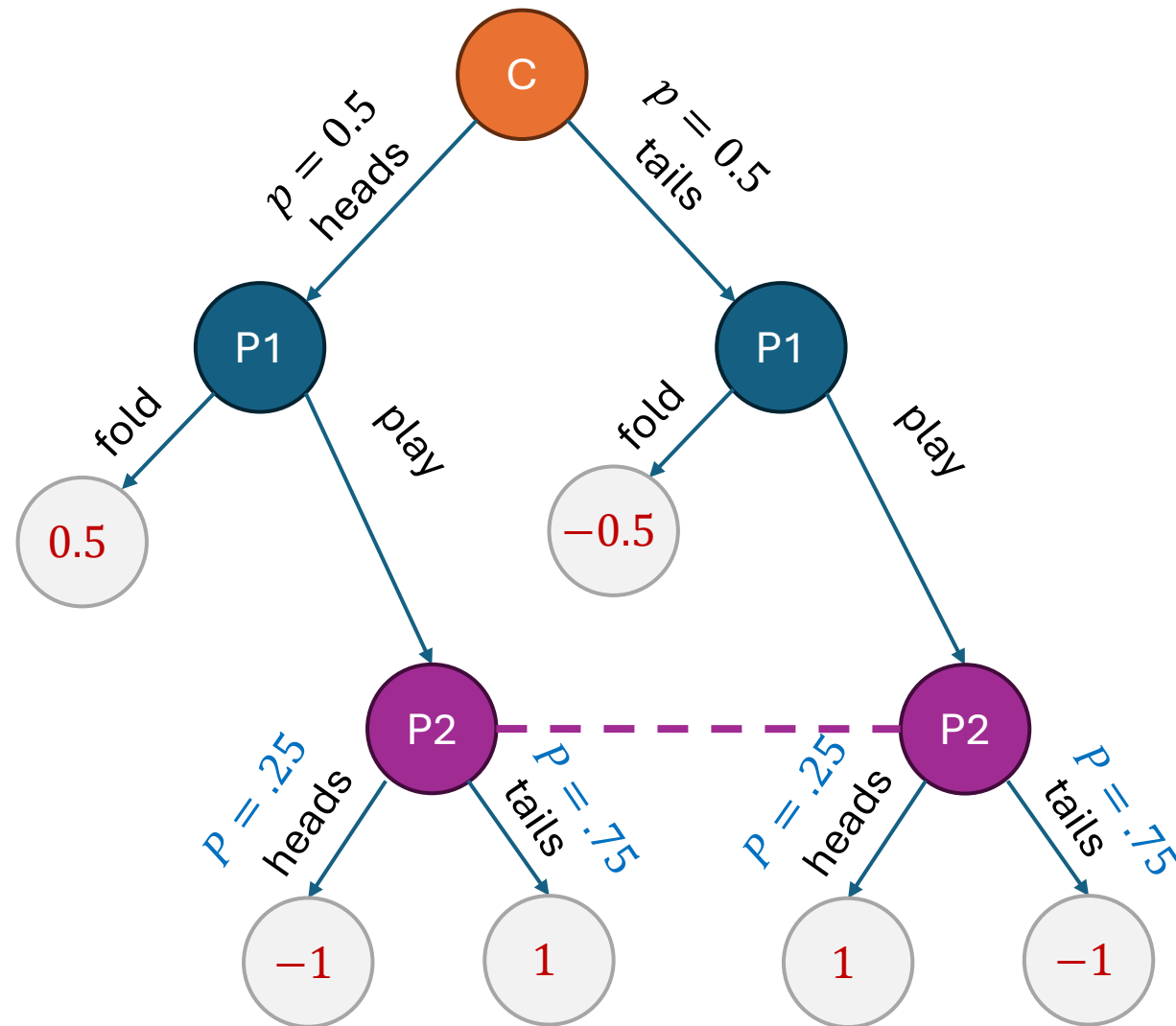
How should P2 play in the sub-game?
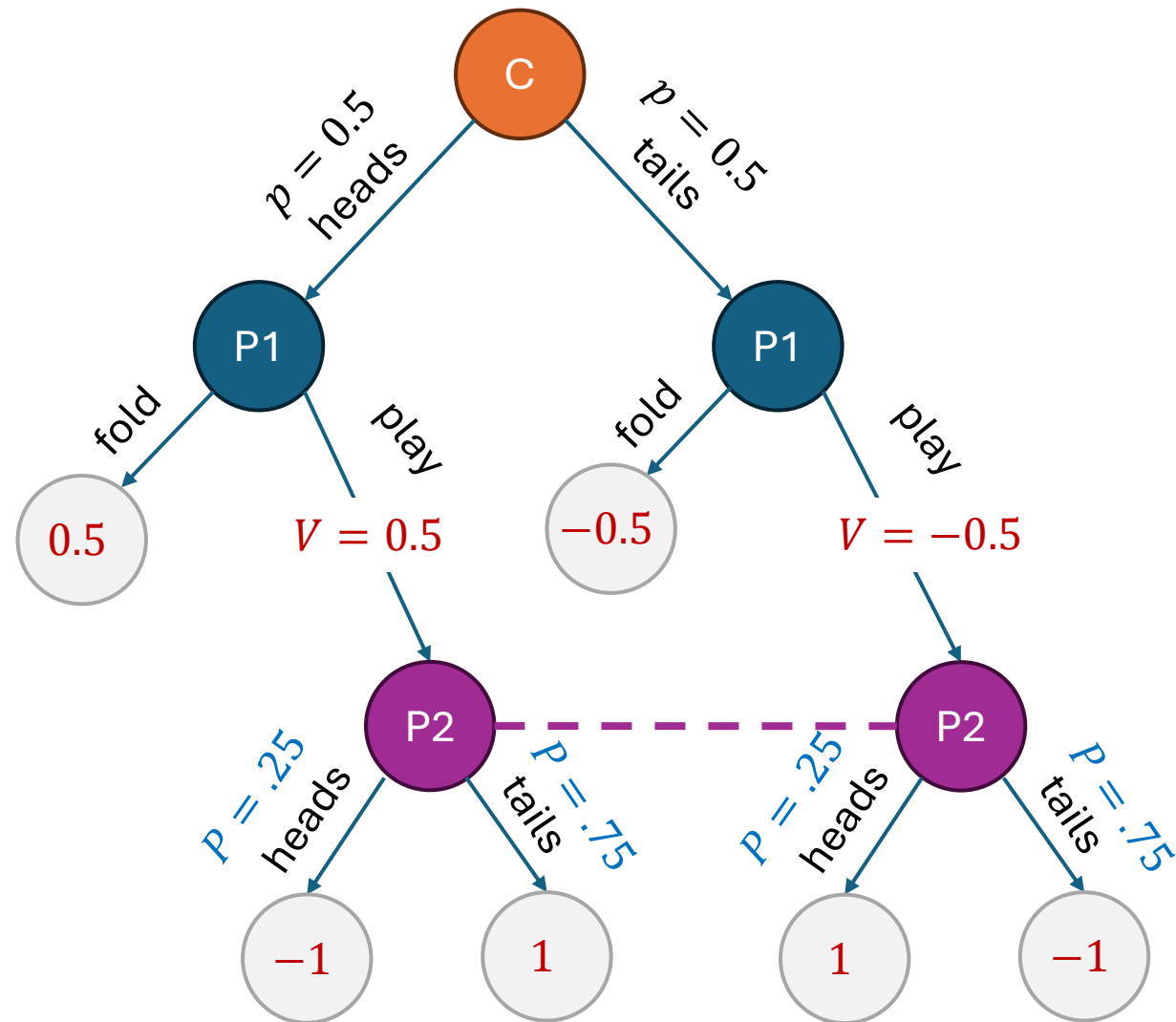
How should P2 play in the sub-game?
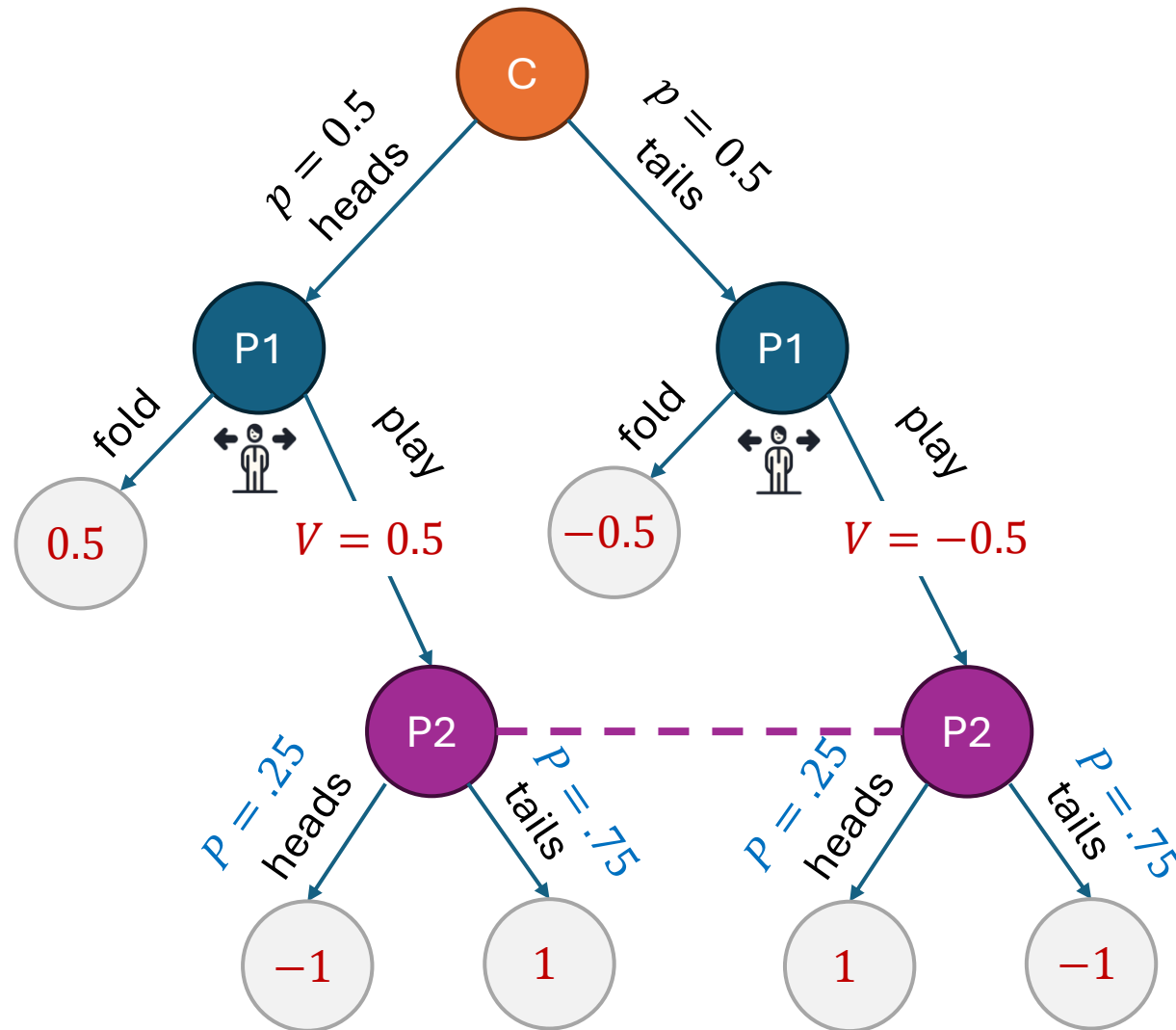
What if we change the value of the fold?
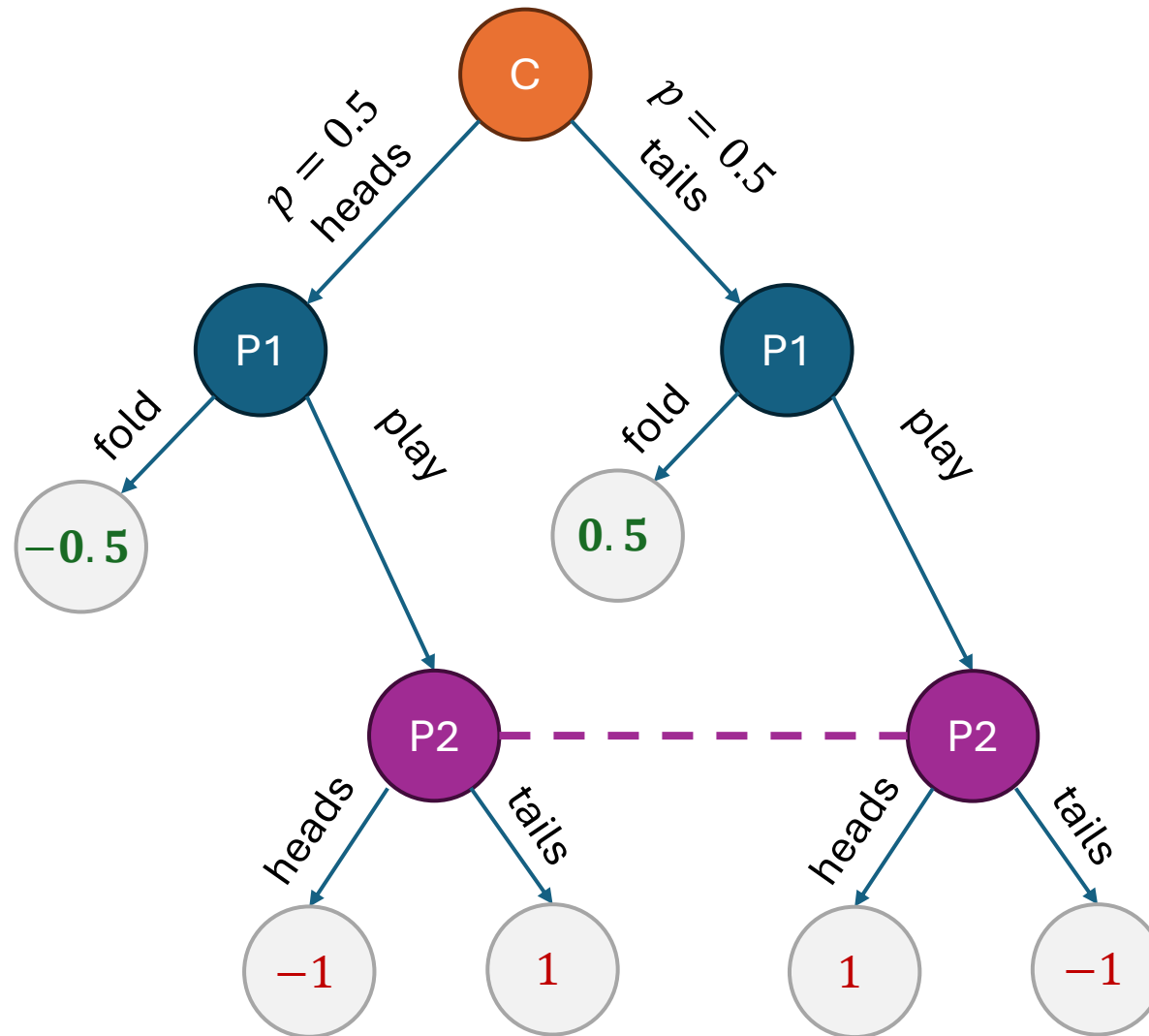
What if we change the value of the fold?
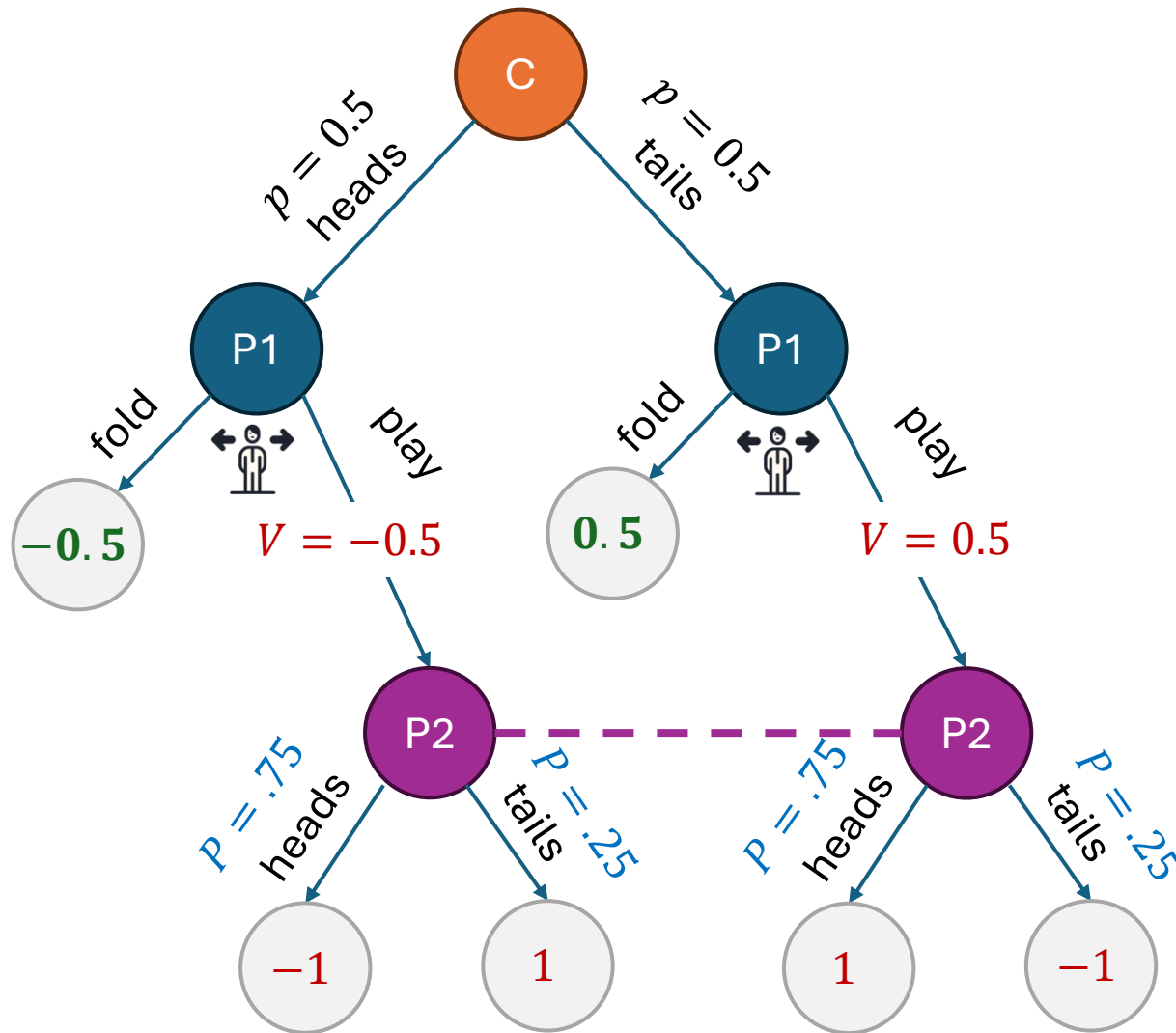
What if we change the value of the fold?
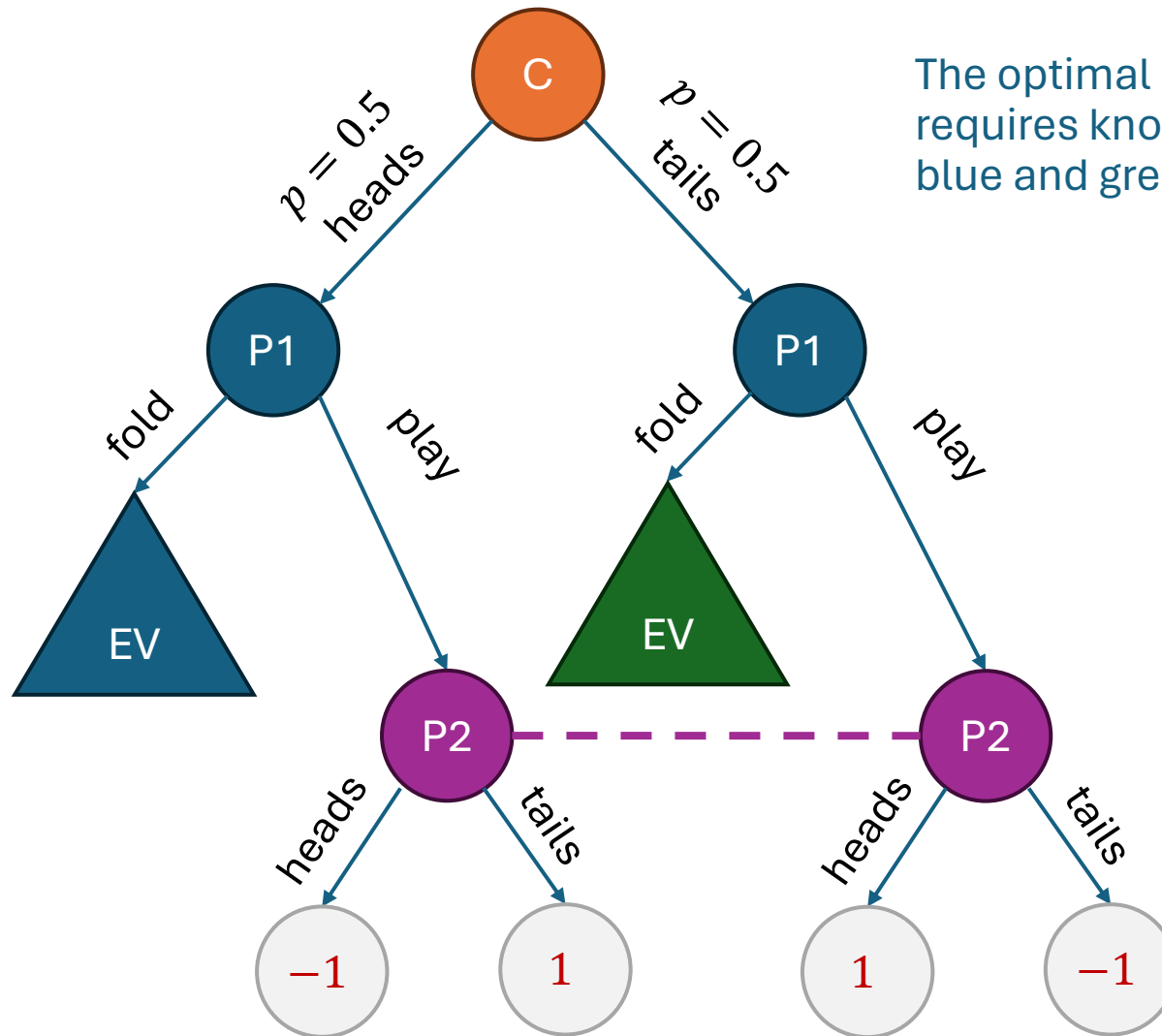
The optimal strategy in the "play" sub-game, requires knowledge of what happens in the blue and green "fold" sub-games

The Elements of an Imperfect Information Game Tree
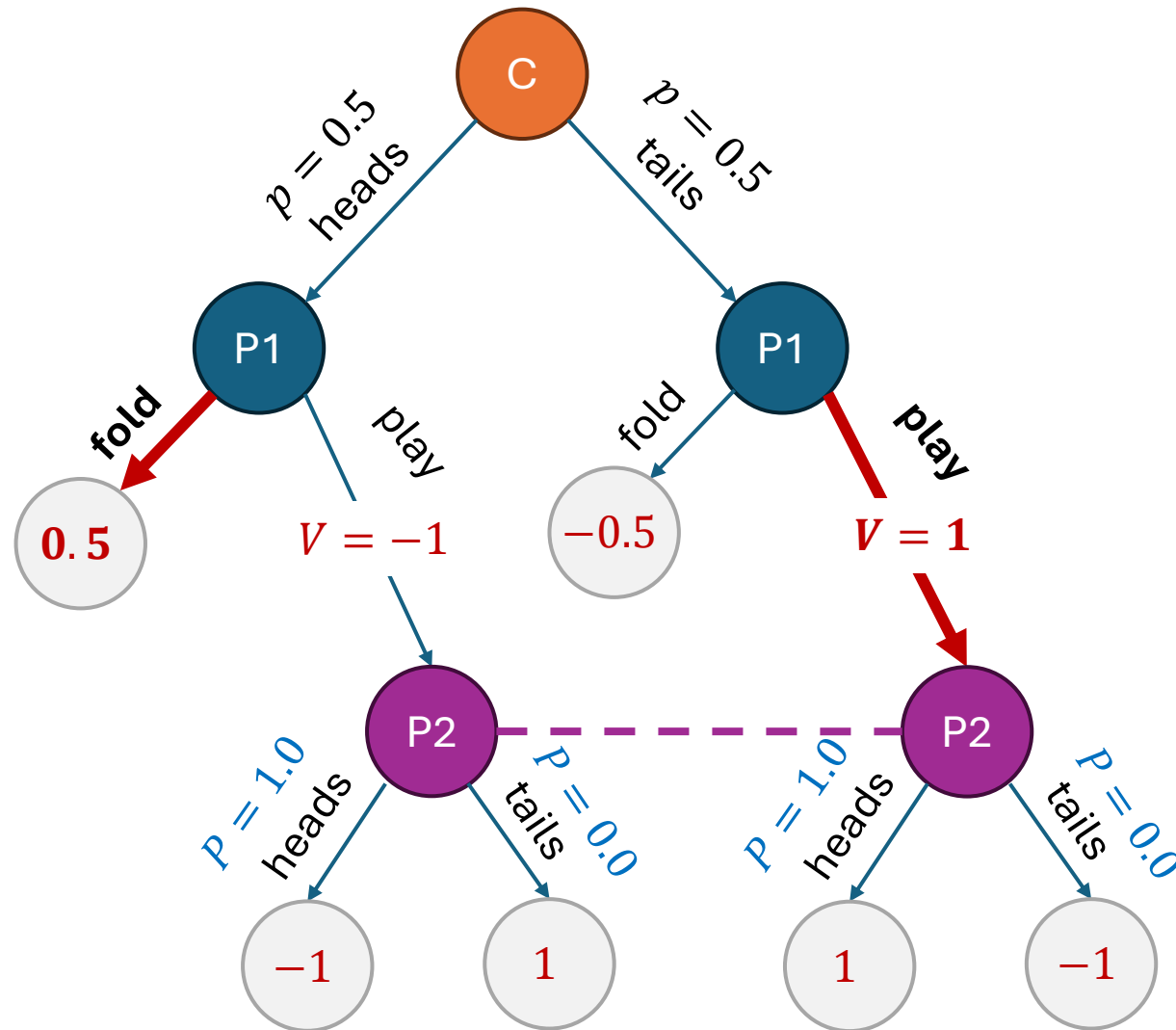
Tree Representation and Information Sets

- **Nodes.** Each node in the tree is a decision point for some player
- **Information sets (infoset).** Nodes that belong to player i are partitioned into information sets $I \in \mathcal{I}_i$, with indices $\mathcal{J}_i = \{j_1, \ldots, j_{K_i}\}$
- Player does not know which node in the information set is chosen
- Must use the same strategy on all nodes in information set
- Each infoset $j \in \mathcal{J}_i$ has a set of actions A_I that the player can take
- **Leaf nodes Z.** The set of terminal states. Player 1 gains utility $u(z)$
- **Chance nodes.** Chance or Nature moves with a fixed distribution

Perfect Recall

- Players remember all the past actions they took

- For each infoset I, there is unique "parent" (infoset, action) pair

- For every node in $I \in \mathcal{I}_i$, the parent pair I', a' was the last infoset visited and action taken by player i before reaching I

- Let p_j the last action player took before reaching infoset indexed j

Strategic Form Representation

- **Mixed strategy.** A distribution over pure strategies
- **Behavioral strategy.** A set of distributions over actions at each information set

Kuhn's Theorem. For every mixed strategy there is an equivalent behavioral strategy that against all profiles of strategies of opponents induces the same distribution over terminal nodes

We will only be talking about behavioral strategies hereafter

A Simple "Weird" Poker Game
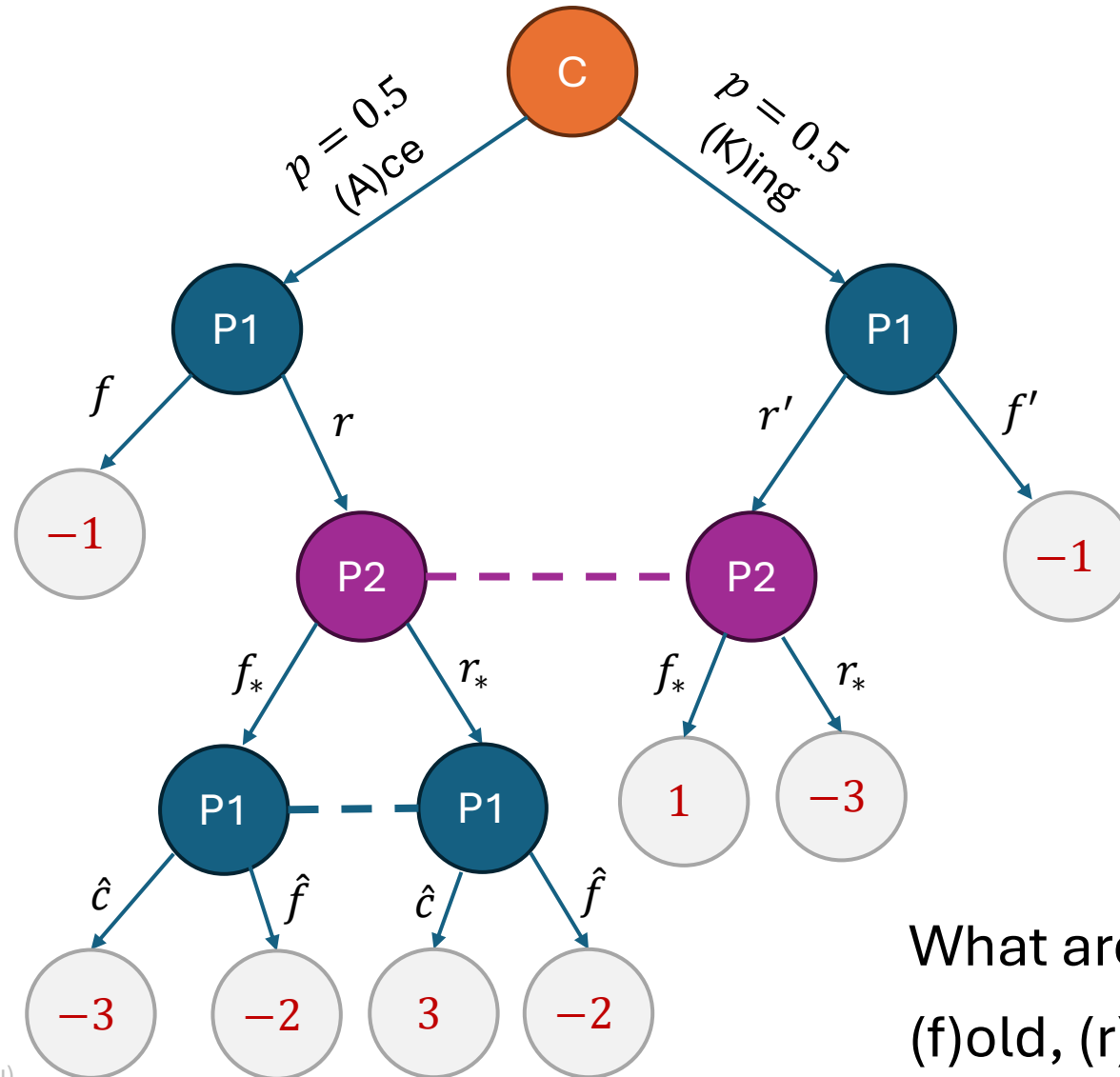
What are the rules of the game?

(f)old, (r)aise, (c)heck

Why is Nash Equilibrium a Good Idea?

- In zero-sum games where no player has an a-priori competitive advantage, Nash Equilibrium guarantees no loss in expectation

- It is a "safe" strategy no matter what the opponent does!

Computing Nash Equilibrium

- We know how to compute equilibria of static zero-sum games
- Can we view the extensive form zero-sum game also as min-max

$$\min_{x \in X} \max_{y \in Y} x^\top A y$$

- What does x and y encode?

- What if $x = \left(x^j\right)_{j \in \mathcal{J}_1}$, where x^j is mixed strategy at infoset $j \in \mathcal{J}_1$
- What if $y = \left(y^j\right)_{j \in \mathcal{J}_2}$, where y^j is mixed strategy at infoset $j \in \mathcal{J}_2$

Behavioral Strategy Representation
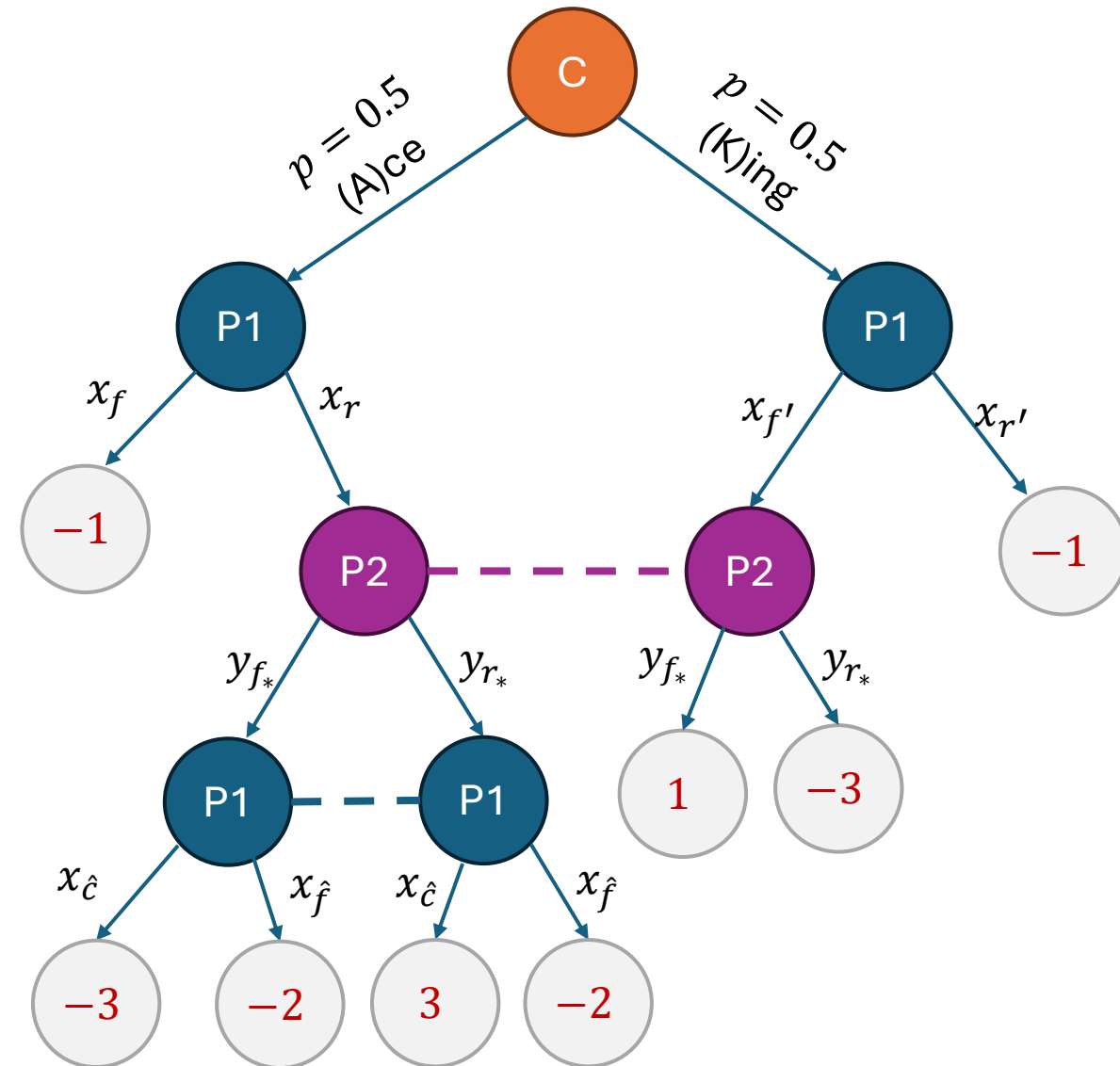
What is the expected payoff of x?

Behavioral Strategy Representation

What is the expected payoff of x?

$$\frac{1}{2}\left(-x_f - 3\,x_r y_{f_*} x_{\hat{c}} - 2\,x_r y_{f_*} x_{\hat{f}} + 3 x_r y_{r_*} x_{\hat{c}} + 2\,x_r y_{r_*} x_{\hat{f}}\right)$$
$$+ \frac{1}{2}\left(x_{f'} y_{f_*} - 3\,x_{f'} y_{r_*} - x_{r'}\right)$$

Behavioral Strategy Representation

What is the expected payoff of x?

$$\frac{1}{2}\left(-x_f - 3\,x_r y_{f_*} x_{\hat{c}} - 2\,x_r y_{f_*} x_{\hat{f}} + 3 x_r y_{r_*} x_{\hat{c}} + 2\,x_r y_{r_*} x_{\hat{f}}\right)$$
$$+\frac{1}{2}\left(x_{f'} y_{f_*} - 3\,x_{f'} y_{r_*} - x_{r'}\right)$$

Is it of the form $x^\top A y$?

Behavioral Strategy Representation

What is the expected payoff of x?

$$\frac{1}{2}\left(-x_f - 3\,x_r y_{f_*} x_{\hat{c}} - 2\,x_r y_{f_*} x_{\hat{f}} + 3 x_r y_{r_*} x_{\hat{c}} + 2\,x_r y_{r_*} x_{\hat{f}}\right)$$
$$+\frac{1}{2}\left(x_{f'} y_{f_*} - 3\,x_{f'} y_{r_*} - x_{r'}\right)$$

Is it of the form $x^\top A y$?
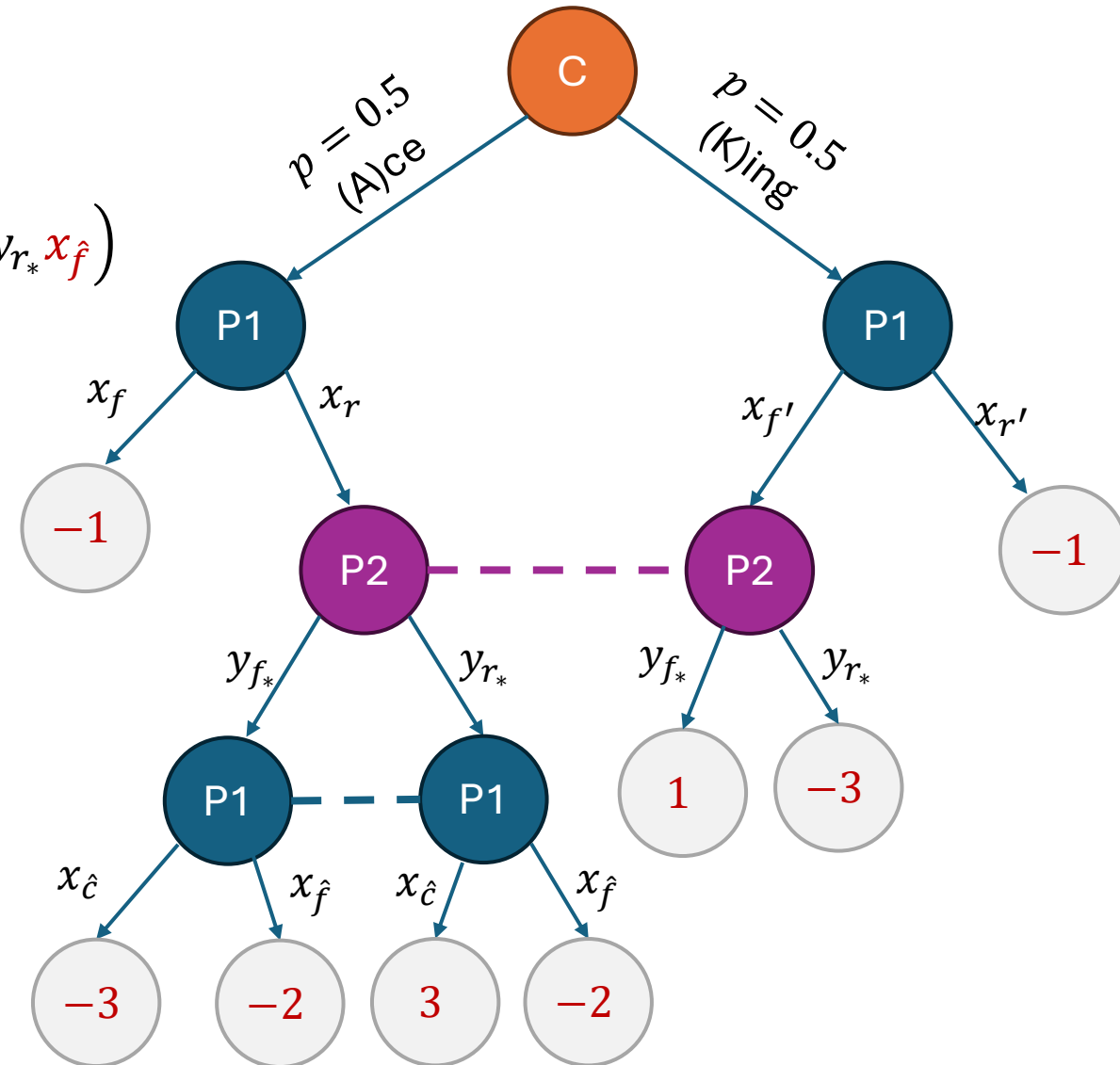
Behavioral Strategy Representation

What is the expected payoff of x?

$$\frac{1}{2}\left(-x_f - 3\,\underbrace{x_r y_{f_*} x_{\hat{c}}}_{\tilde{x}_{r\hat{c}}} - 2\,\underbrace{x_r y_{f_*} x_{\hat{f}}}_{\tilde{x}_{r\hat{f}}} + 3\,\underbrace{x_r y_{r_*} x_{\hat{c}}}_{\tilde{x}_{r\hat{c}}} + 2\,\underbrace{x_r y_{r_*} x_{\hat{f}}}_{\tilde{x}_{r\hat{f}}}\right)$$

$$+ \frac{1}{2}\left(x_{f'} y_{f_*} - 3\,x_{f'} y_{r_*} - x_{r'}\right)$$

IDEA. Group together products that appear into new "variables"

New variables represent the product of the probabilities of the actions chosen by P1 on the path to the last action in the sequence!

We will annotate them just with the last action

$$\tilde{x}_{r\hat{c}} \equiv \tilde{x}_{\hat{c}}, \qquad \tilde{x}_{r\hat{f}} \equiv \tilde{x}_{\hat{f}}$$
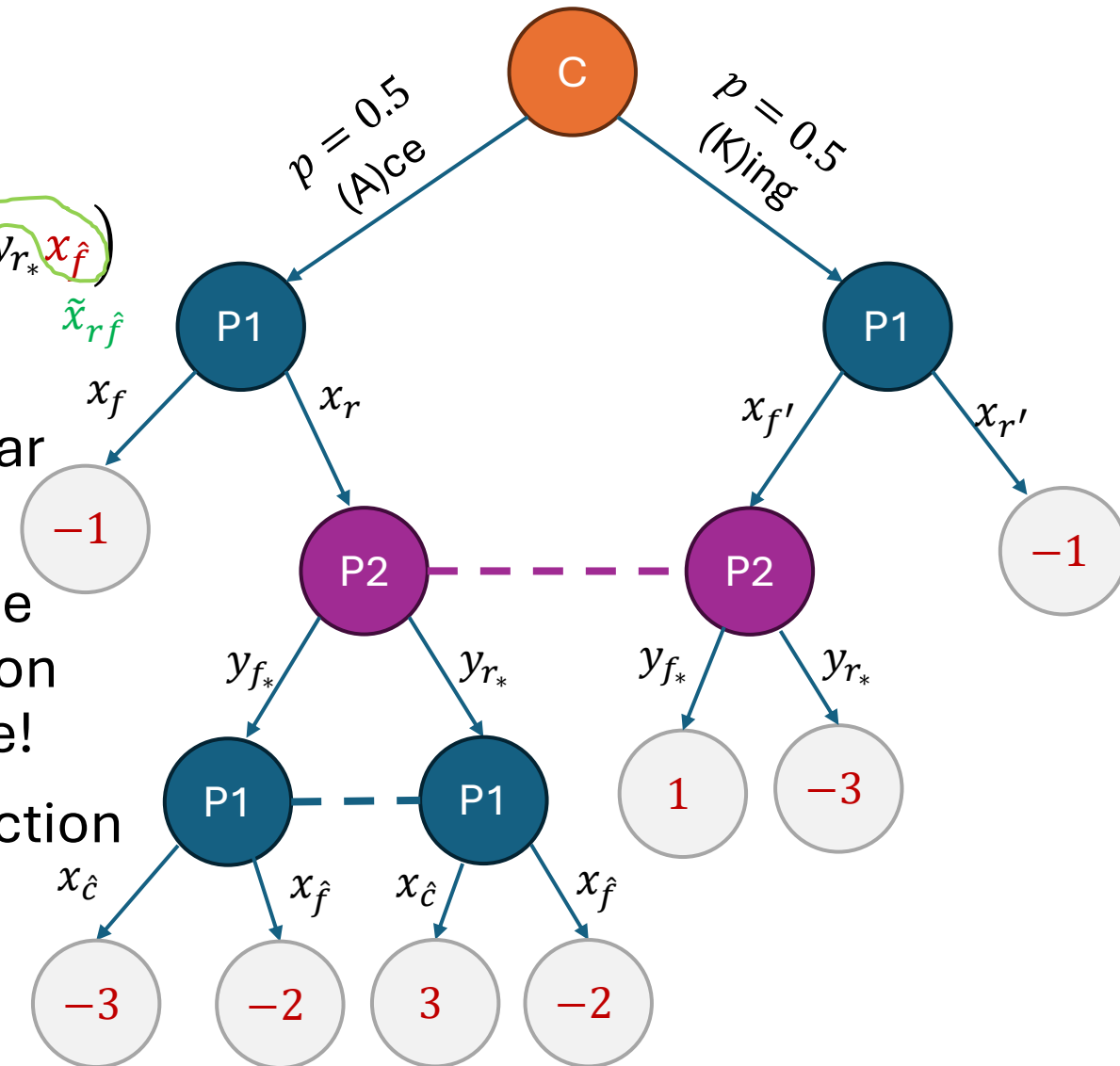
Sequence Form Representation

What is the expected payoff of x?

$$\frac{1}{2}\left(-x_f - 3\,\underbrace{x_r y_{f_*} x_{\hat{c}}}_{\tilde{x}_{r\hat{c}}} - 2\,\underbrace{x_r y_{f_*} x_{\hat{f}}}_{\tilde{x}_{r\hat{f}}} + 3\,\underbrace{x_r y_{r_*} x_{\hat{c}}}_{\tilde{x}_{r\hat{c}}} + 2\,\underbrace{x_r y_{r_*} x_{\hat{f}}}_{\tilde{x}_{r\hat{f}}}\right)$$

$$+ \frac{1}{2}\left(x_{f'} y_{f_*} - 3\,x_{f'} y_{r_*} - x_{r'}\right)$$

Sequence form strategies. We can define these new variables $\tilde{x}_a$ for all actions of P1

$\tilde{x}_a$: represents product of probabilities of all actions of P1 on the path to a
$$\tilde{x}_f, \tilde{x}_r, \tilde{x}_{f'}, \tilde{x}_{r'}, \tilde{x}_{\hat{c}}, \tilde{x}_{\hat{f}}$$

Sequence Form Representation

What is the expected payoff of x?

$$\frac{1}{2}\left(-x_f - 3\,\underbrace{x_r y_{f_*} x_{\hat{c}}}_{\tilde{x}_{r\hat{c}}} - 2\,\underbrace{x_r y_{f_*} x_{\hat{f}}}_{\tilde{x}_{r\hat{f}}} + 3\,\underbrace{x_r y_{r_*} x_{\hat{c}}}_{\tilde{x}_{r\hat{c}}} + 2\,\underbrace{x_r y_{r_*} x_{\hat{f}}}_{\tilde{x}_{r\hat{f}}}\right)$$

$$+ \frac{1}{2}\left(x_{f'} y_{f_*} - 3\,x_{f'} y_{r_*} - x_{r'}\right)$$

Sequence form strategies. We can define these new variables $\tilde{x}_a$ for all actions of P1

$\tilde{x}_a$: represents product of probabilities of all actions of P1 on the path to a

$$\tilde{x}_f, \tilde{x}_r, \tilde{x}_{f'}, \tilde{x}_{r'}, \tilde{x}_{\hat{c}}, \tilde{x}_{\hat{f}}$$

$\tilde{y}_a$: represents product of probabilities of all actions of P2 on the path to a

$$\tilde{y}_{f_*}, \tilde{y}_{r_*}$$

Sequence Form Representation

What is the expected payoff of x?

$$\frac{1}{2}\left(-\tilde{x}_f - 3\, y_{f_*}\tilde{x}_{\hat{c}} - 2\, y_{f_*}\tilde{x}_{\hat{f}} + 3y_{r_*}\tilde{x}_{\hat{c}} + 2\, y_{r_*}\tilde{x}_{\hat{f}}\right)$$
$$+\frac{1}{2}\left(\tilde{x}_{f'}\tilde{y}_{f_*} - 3\, \tilde{x}_{f'}\tilde{y}_{r_*} - \tilde{x}_{r'}\right)$$

Sequence form strategies. We can define these new variables $\tilde{x}_a$ for all actions of P1

$\tilde{x}_a$: represents product of probabilities of all actions of P1 on the path to a
$$\tilde{x}_f, \tilde{x}_r, \tilde{x}_{f'}, \tilde{x}_{r'}, \tilde{x}_{\hat{c}}, \tilde{x}_{\hat{f}}$$

$\tilde{y}_a$: represents product of probabilities of all actions of P2 on the path to a
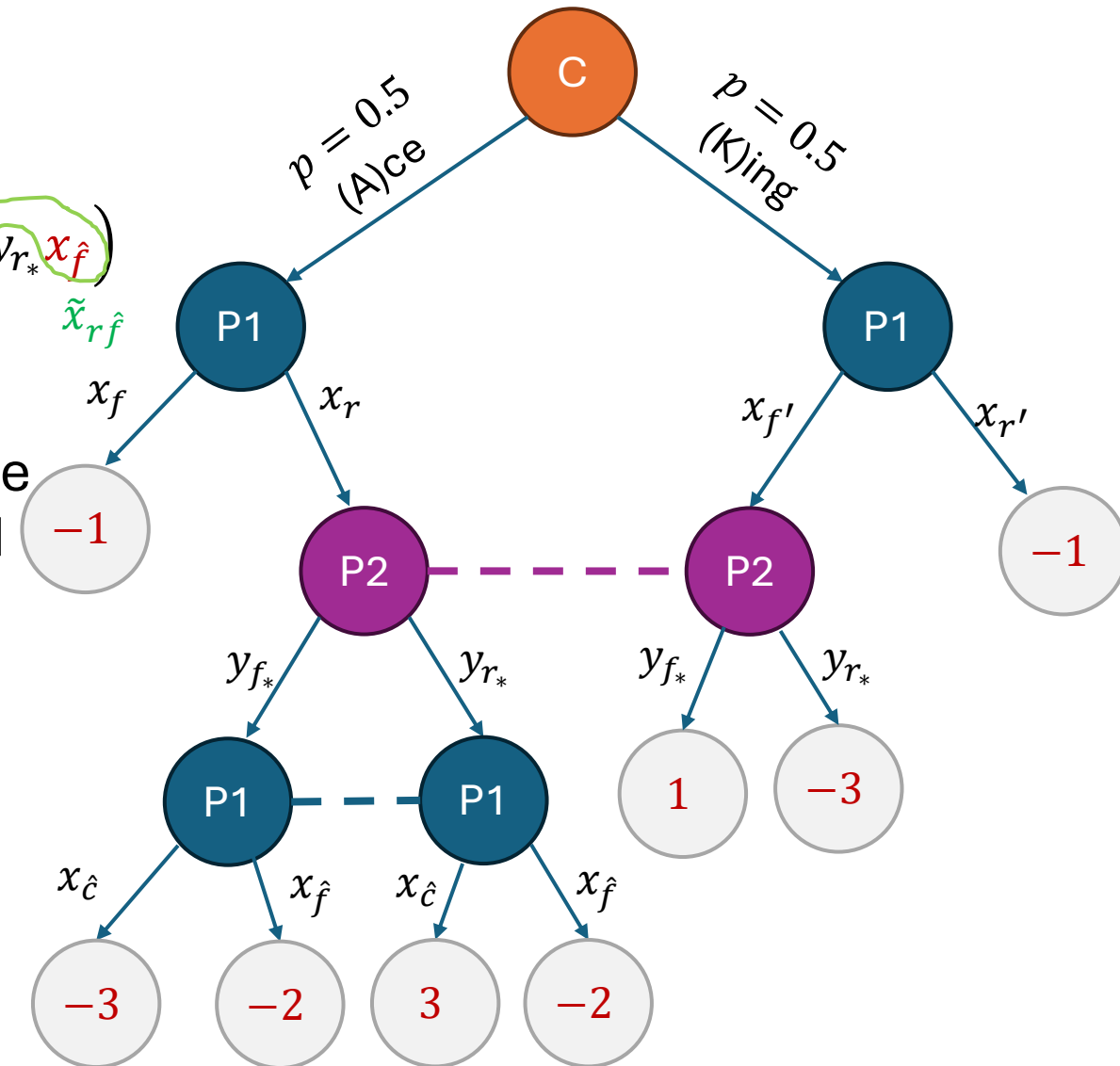$$\tilde{y}_{f_*}, \tilde{y}_{r_*}$$

Sequence Form Representation

What is the expected payoff of x?

$$\frac{1}{2}\left(-\tilde{x}_f - 3\,y_{f_*}\tilde{x}_{\hat{c}} - 2\,y_{f_*}\tilde{x}_{\hat{f}} + 3y_{r_*}\tilde{x}_{\hat{c}} + 2\,y_{r_*}\tilde{x}_{\hat{f}}\right)$$

$$+\frac{1}{2}\left(\tilde{x}_{f'}\tilde{y}_{f_*} - 3\,\tilde{x}_{f'}\tilde{y}_{r_*} - \tilde{x}_{r'}\right)$$

Observation. This is of the form $x^\top A y$.

What is the dimension of A?

Sequence Form Representation

What is the expected payoff of x?

$$\frac{1}{2}\left(-\tilde{x}_f - 3\,y_{f_*}\tilde{x}_{\hat{c}} - 2\,y_{f_*}\tilde{x}_{\hat{f}} + 3y_{r_*}\tilde{x}_{\hat{c}} + 2\,y_{r_*}\tilde{x}_{\hat{f}}\right)$$
$$+\frac{1}{2}\left(\tilde{x}_{f'}\tilde{y}_{f_*} - 3\,\tilde{x}_{f'}\tilde{y}_{r_*} - \tilde{x}_{r'}\right)$$

Observation. This is of the form $x^\top A y$.

What is the dimension of A?

One row for each possible action a of P

One column for each possible action a' of P2

Sequence Form Representation

What is the expected payoff of x?

$$\frac{1}{2}\left(-\tilde{x}_f - 3\,y_{f_*}\tilde{x}_{\hat{c}} - 2\,y_{f_*}\tilde{x}_{\hat{f}} + 3y_{r_*}\tilde{x}_{\hat{c}} + 2\,y_{r_*}\tilde{x}_{\hat{f}}\right)$$
$$+ \frac{1}{2}\left(\tilde{x}_{f'}\tilde{y}_{f_*} - 3\,\tilde{x}_{f'}\tilde{y}_{r_*} - \tilde{x}_{r'}\right)$$

Observation. This is of the form $x^\top A y$.

What is the dimension of A?

One row for each possible action a of P
One column for each possible action a' of P2

What is the value $A_{a,a'}$?

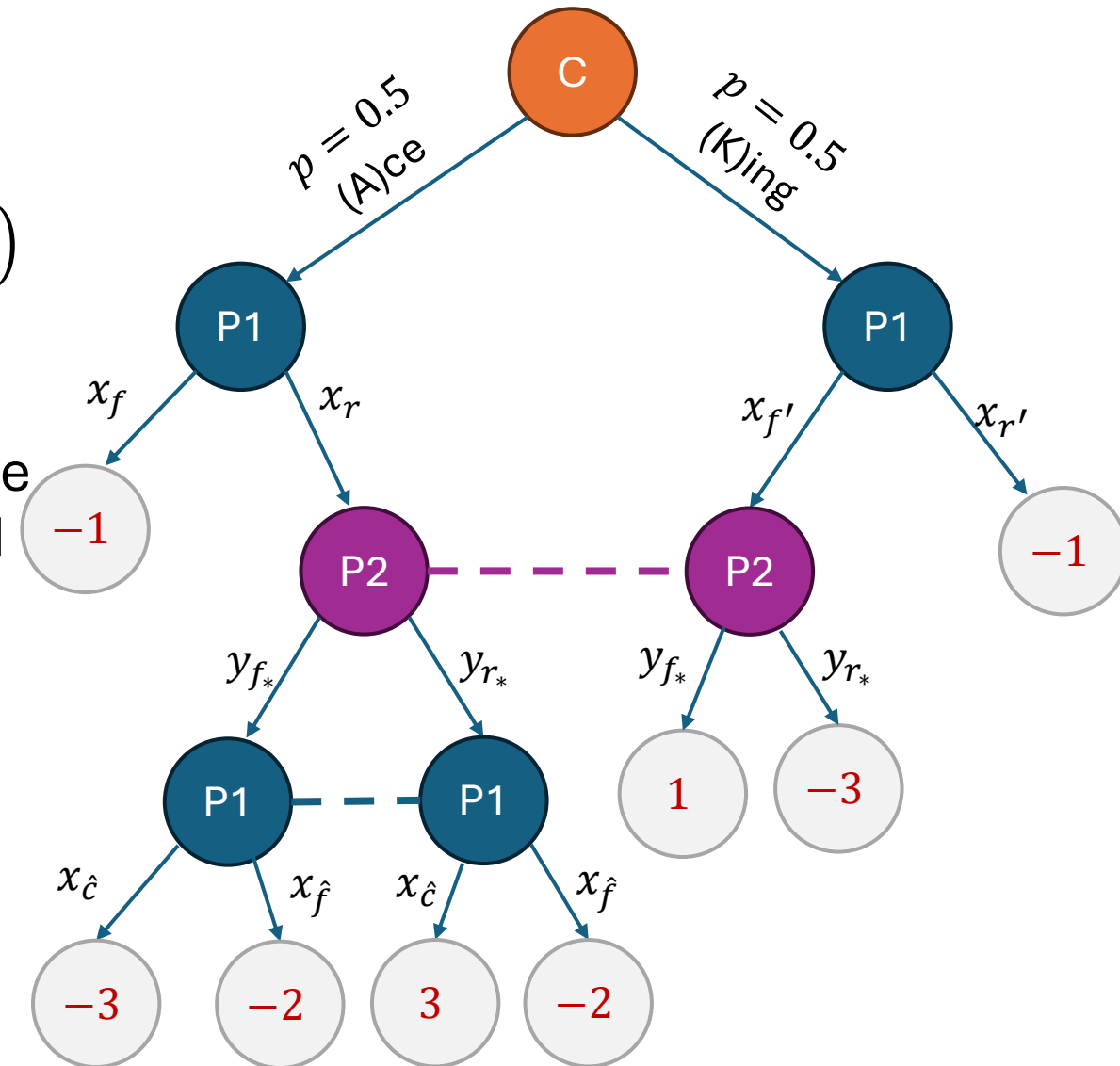# Sequence Form Representation

What is the expected payoff of x?

$$\frac{1}{2}\left(-\tilde{x}_f - 3\,y_{f_*}\tilde{x}_{\hat{c}} - 2\,y_{f_*}\tilde{x}_{\hat{f}} + 3y_{r_*}\tilde{x}_{\hat{c}} + 2\,y_{r_*}\tilde{x}_{\hat{f}}\right)$$
$$+\frac{1}{2}\left(\tilde{x}_{f'}\tilde{y}_{f_*} - 3\,\tilde{x}_{f'}\tilde{y}_{r_*} - \tilde{x}_{r'}\right)$$

Observation. This is of the form $x^\top A y$.

What is the dimension of A?

One row for each possible action a of P
One column for each possible action a' of P2

What is the value $A_{a,a'}$?

If there exists a terminal node, such that a was the last action chosen by P1 and a' was the last action chosen by P2 then it is the value of the leaf multiplied by all "chance" probabilities on the path to the leaf
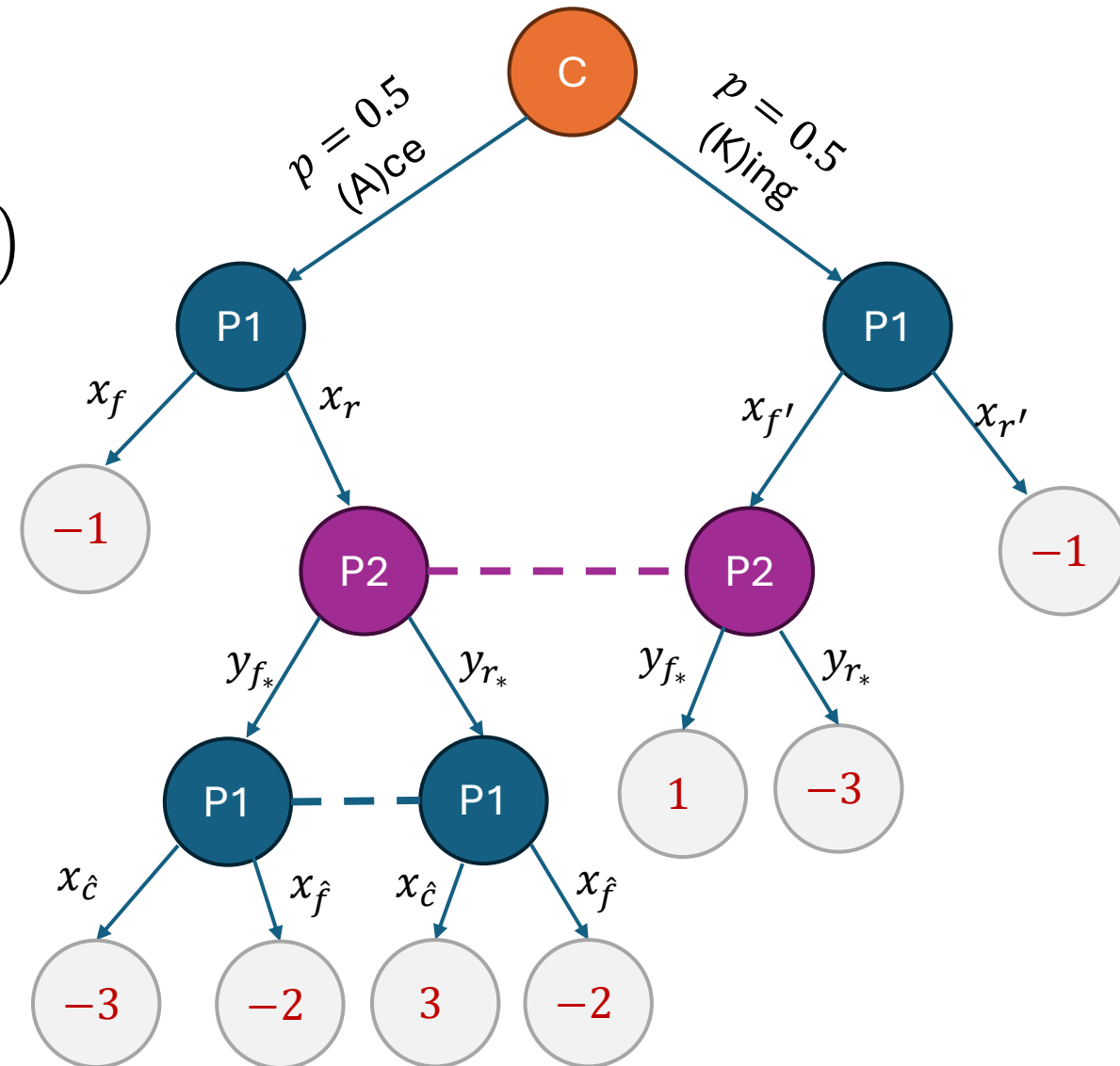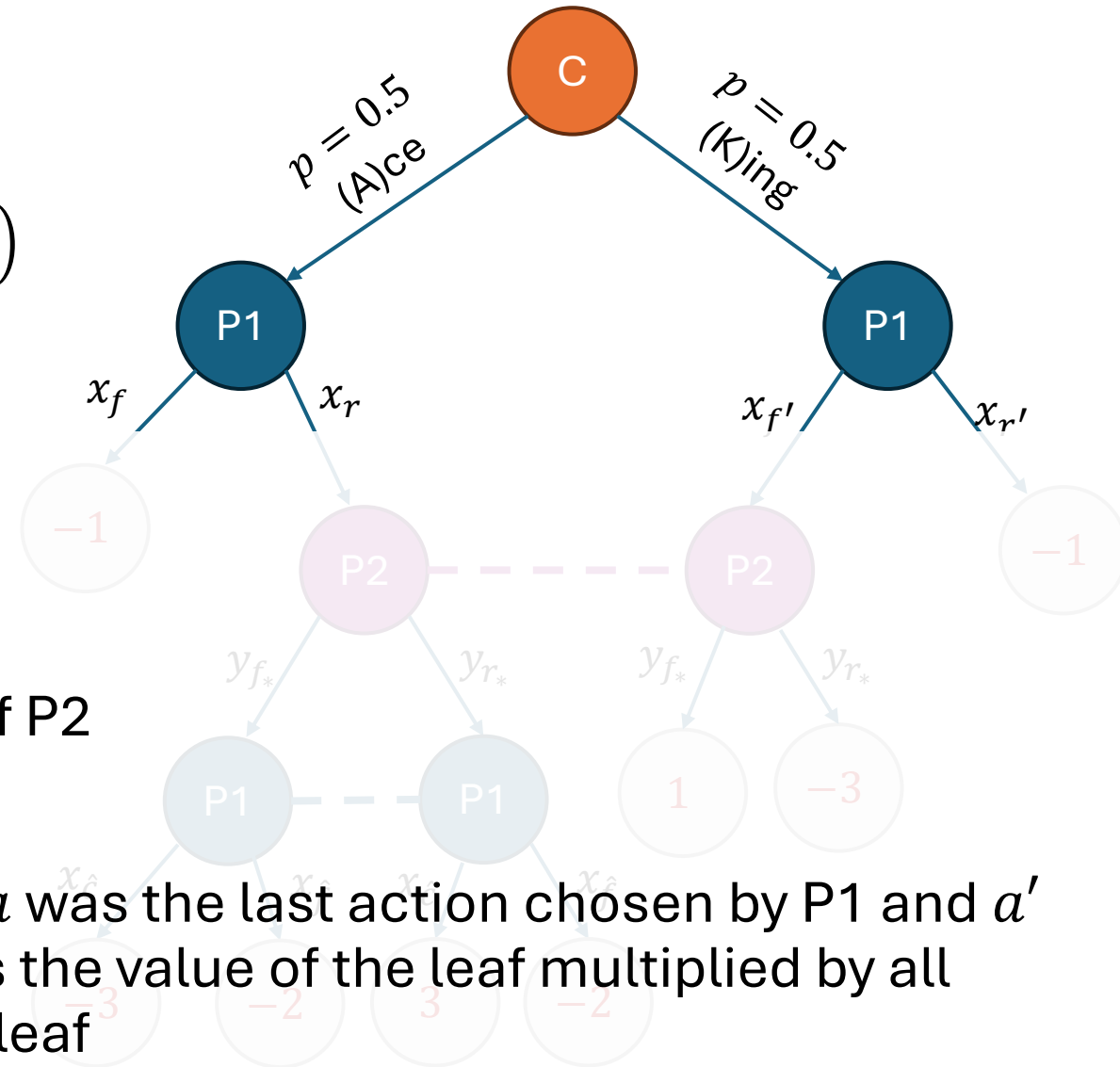
Sequence Form Representation
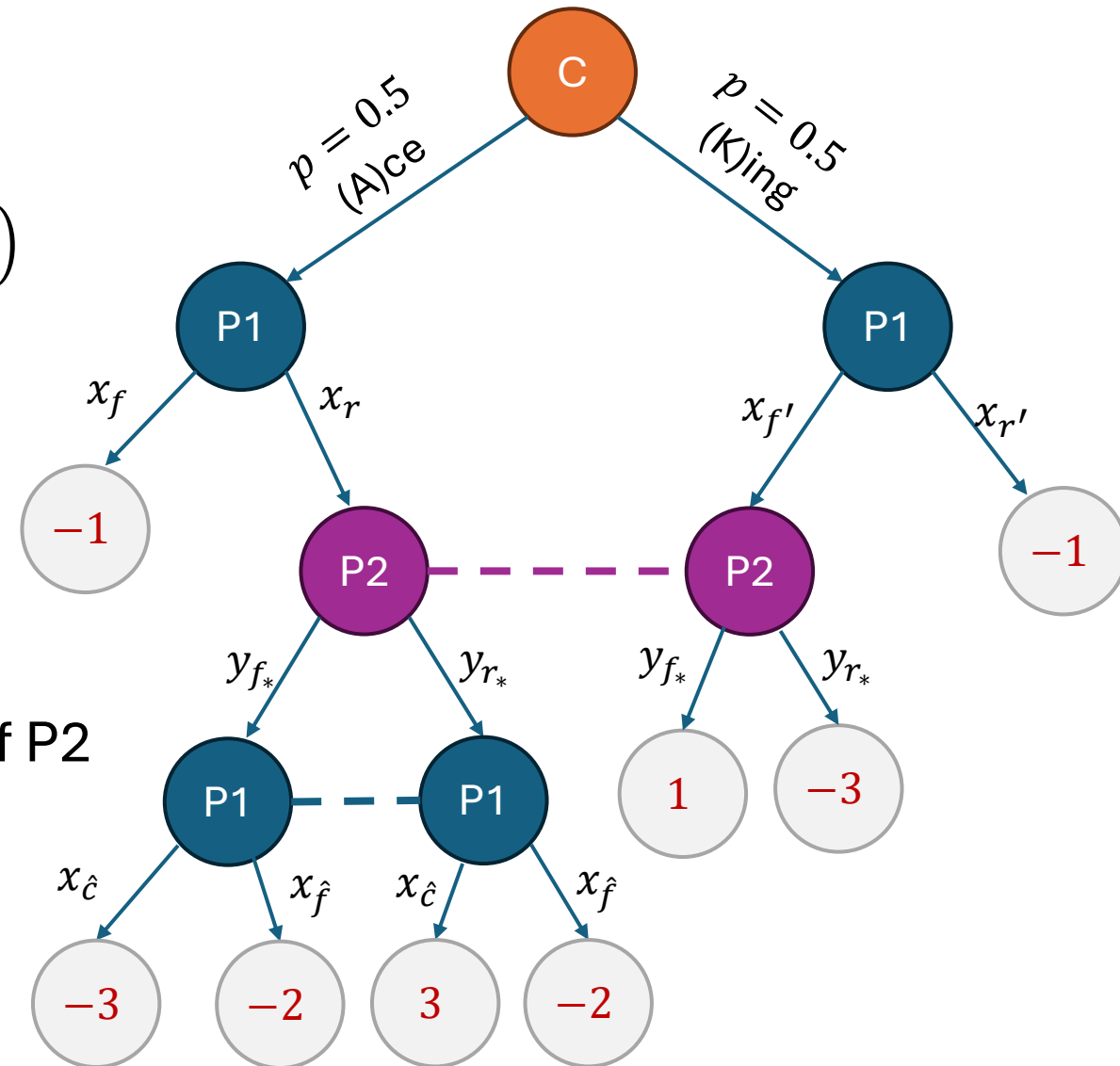
What is the expected payoff of x?

$$\frac{1}{2}\left(-\tilde{x}_f - 3\,y_{f_*}\tilde{x}_{\hat{c}} - 2\,y_{f_*}\tilde{x}_{\hat{f}} + 3y_{r_*}\tilde{x}_{\hat{c}} + 2\,y_{r_*}\tilde{x}_{\hat{f}}\right)$$
$$+ \frac{1}{2}\left(\tilde{x}_{f'}\tilde{y}_{f_*} - 3\,\tilde{x}_{f'}\tilde{y}_{r_*} - \tilde{x}_{r'}\right)$$

What constraints does $\tilde{x}$ need to respect?

Sequence Form Representation

What is the expected payoff of x?

$$\frac{1}{2}\left(-\tilde{x}_f - 3\,y_{f_*}\tilde{x}_{\hat{c}} - 2\,y_{f_*}\tilde{x}_{\hat{f}} + 3y_{r_*}\tilde{x}_{\hat{c}} + 2\,y_{r_*}\tilde{x}_{\hat{f}}\right)$$
$$+\frac{1}{2}\left(\tilde{x}_{f'}\tilde{y}_{f_*} - 3\,\tilde{x}_{f'}\tilde{y}_{r_*} - \tilde{x}_{r'}\right)$$

What constraints does $\widetilde{x}$ need to respect?

Since $\tilde{x}_{\hat{c}}$ is supposed to represent $x_r x_{\hat{c}}$ and $\tilde{x}_{\hat{f}}$ is supposed to represent $x_r x_{\hat{f}}$

$$\tilde{x}_{\hat{c}} + \tilde{x}_{\hat{f}} = x_r\left(x_{\hat{c}} + x_{\hat{f}}\right) = x_r$$

Sequence Form Representation

What is the expected payoff of x?

$$\frac{1}{2}\left(-\tilde{x}_f - 3\,y_{f_*}\tilde{x}_{\hat{c}} - 2\,y_{f_*}\tilde{x}_{\hat{f}} + 3y_{r_*}\tilde{x}_{\hat{c}} + 2\,y_{r_*}\tilde{x}_{\hat{f}}\right)$$

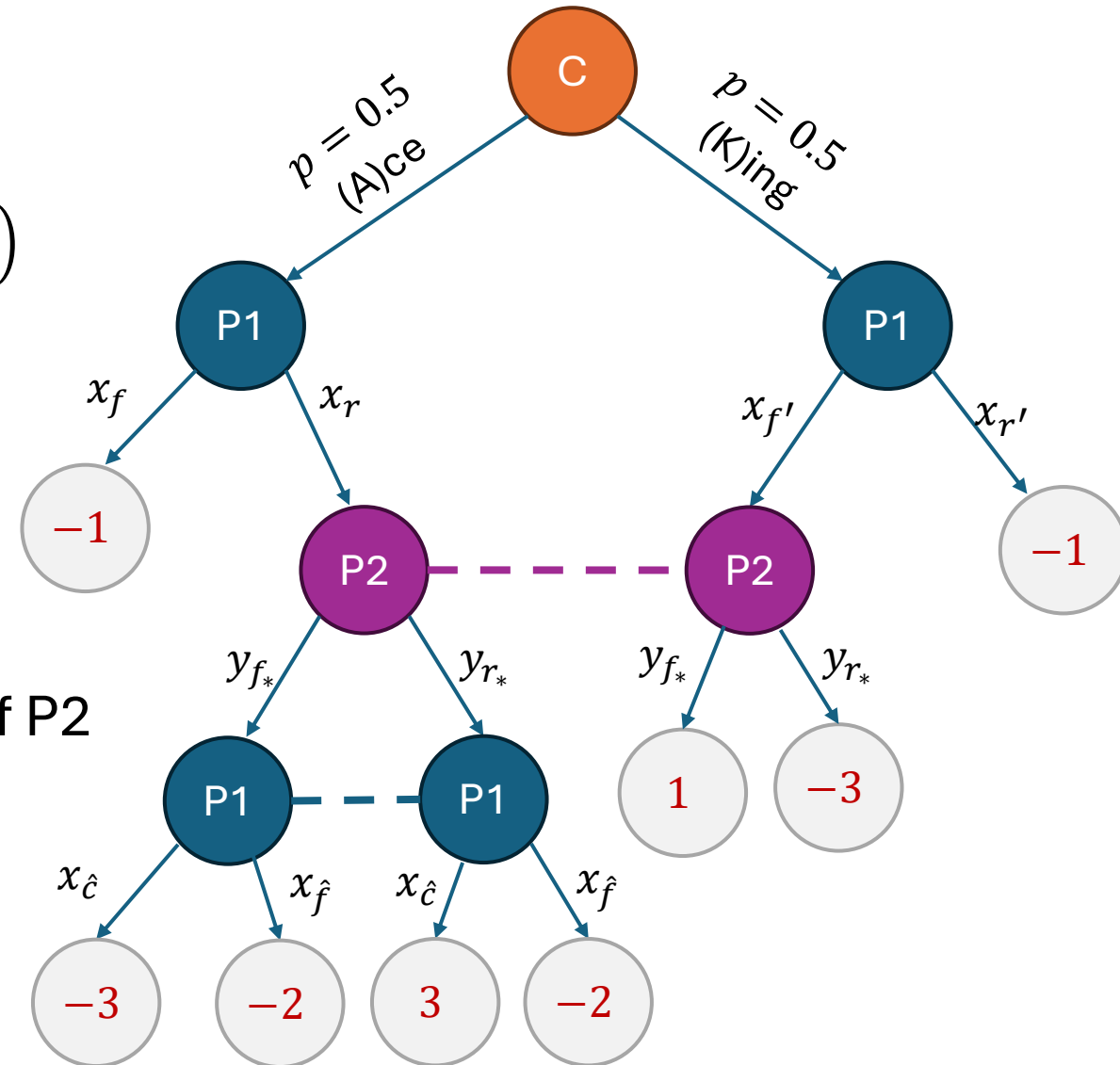$$+\frac{1}{2}\left(\tilde{x}_{f'}\tilde{y}_{f_*} - 3\,\tilde{x}_{f'}\tilde{y}_{r_*} - \tilde{x}_{r'}\right)$$

What constraints does $\widetilde{x}$ need to respect?

Since $\tilde{x}_{\hat{c}}$ is supposed to represent $x_r x_{\hat{c}}$ and $\tilde{x}_{\hat{f}}$ is supposed to represent $x_r x_{\hat{f}}$

$$\tilde{x}_{\hat{c}} + \tilde{x}_{\hat{f}} = x_r\left(x_{\hat{c}} + x_{\hat{f}}\right) = x_r$$

The sum of $\tilde{x}_a$ for all actions at an info set $j \in \mathcal{J}_i$ must be matching $\tilde{x}_{p_j}$, i.e., variable associated with the last action chosen before j
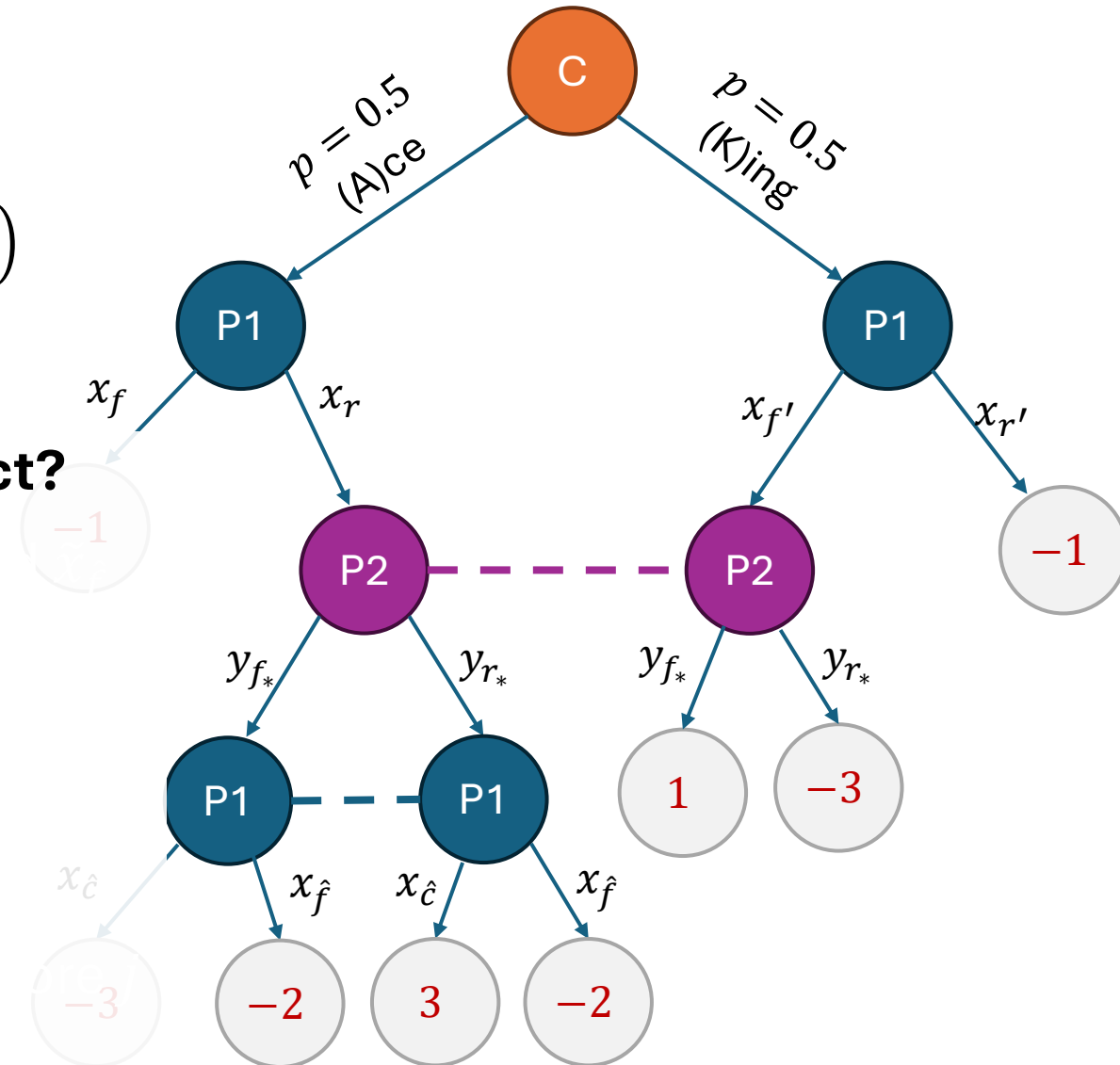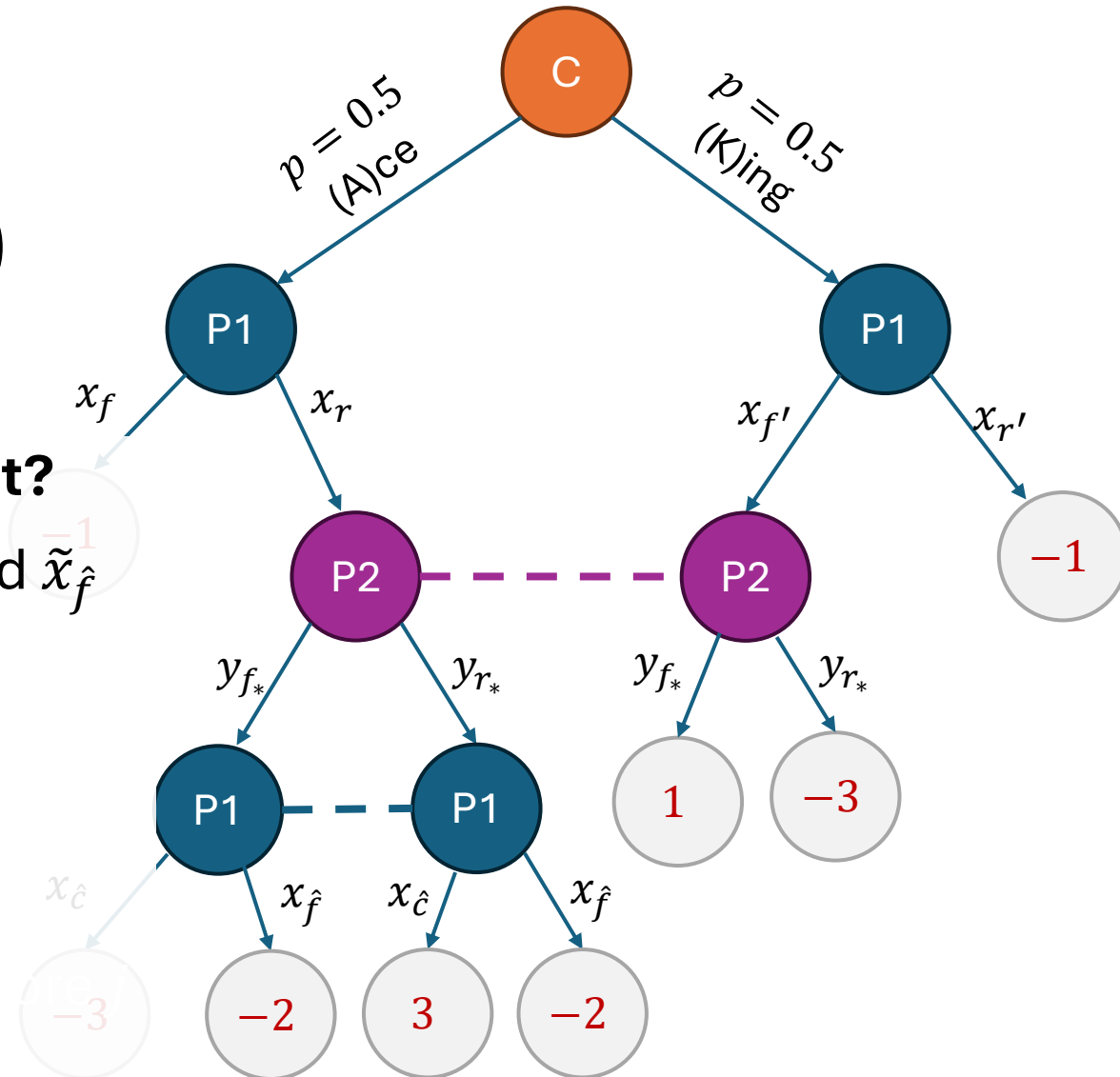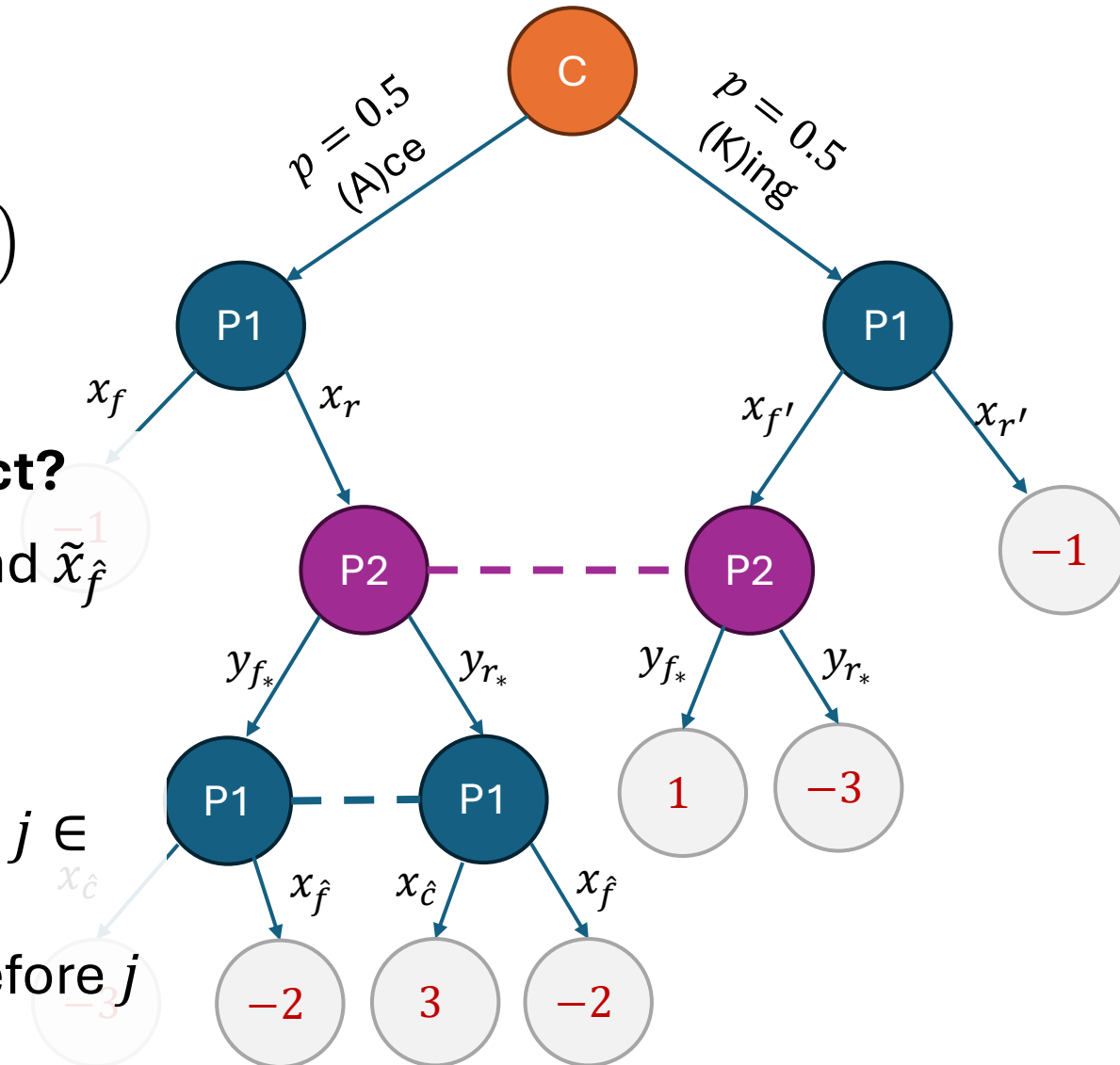
Sequence Form Representation

What is the expected payoff of x?

$$\frac{1}{2}\left(-\tilde{x}_f - 3\, y_{f_*}\tilde{x}_{\hat{c}} - 2\, y_{f_*}\tilde{x}_{\hat{f}} + 3 y_{r_*}\tilde{x}_{\hat{c}} + 2\, y_{r_*}\tilde{x}_{\hat{f}}\right)$$
$$+ \frac{1}{2}\left(\tilde{x}_{f'}\tilde{y}_{f_*} - 3\, \tilde{x}_{f'}\tilde{y}_{r_*} - \tilde{x}_{r'}\right)$$

Are these all?

Since $\tilde{x}_{\hat{c}}$ is supposed to represent $x_r x_{\hat{c}}$ and $\tilde{x}_{\hat{f}}$ is supposed to represent $x_r x_{\hat{f}}$

$$\tilde{x}_{\hat{c}} + \tilde{x}_{\hat{f}} = x_r\left(x_{\hat{c}} + x_{\hat{f}}\right) = x_r = \tilde{x}_r$$

For every, $\tilde{x}$, we can find a valid behavioral x

$$\frac{\tilde{x}_c}{\tilde{x}_r} = x_{\hat{c}}, \qquad \frac{\tilde{x}_{\hat{f}}}{\tilde{x}_r} = x_{\hat{f}}$$
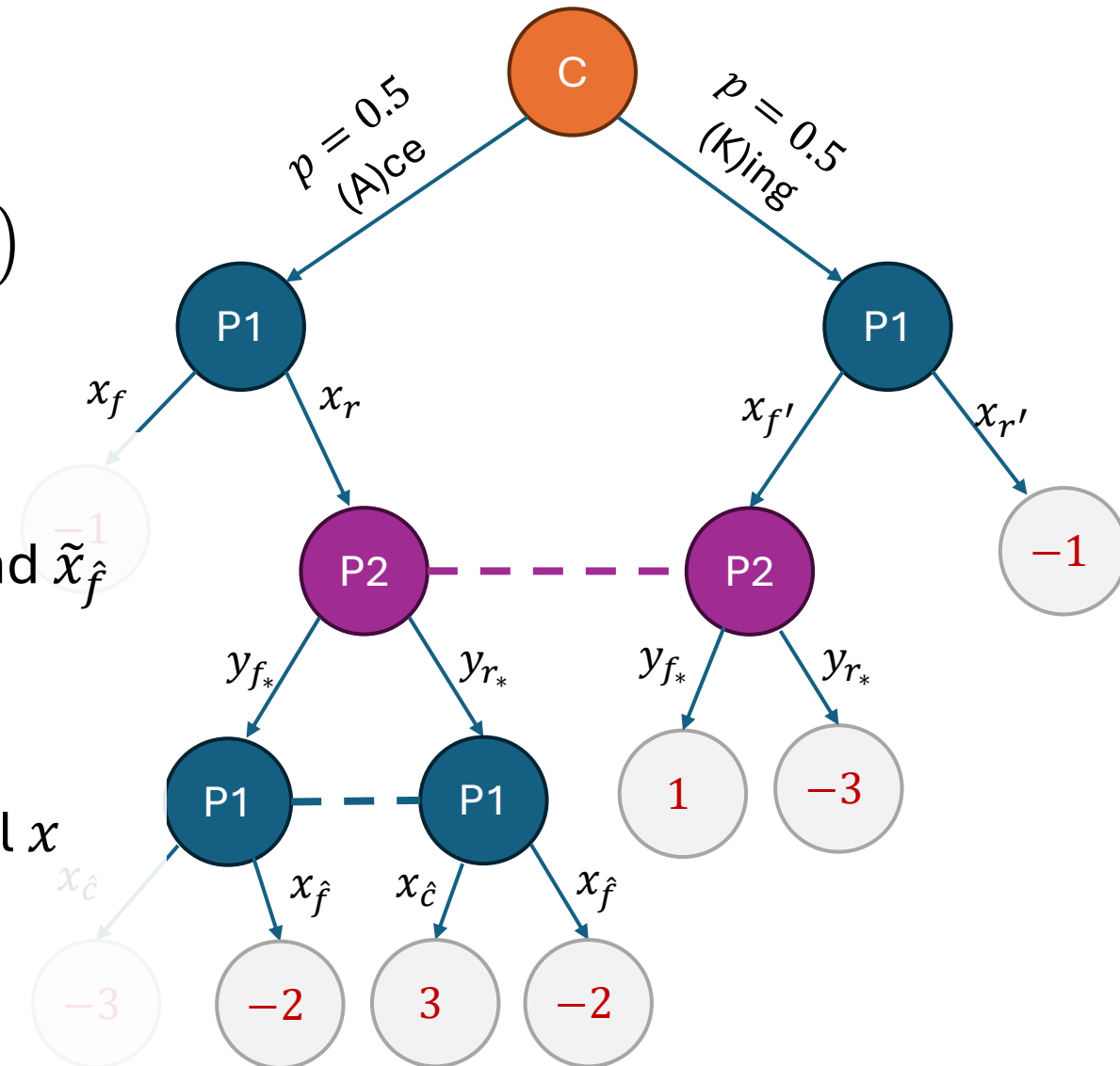
Recap: Sequence Form Representation

- The strategies of the player can be represented as $\tilde{x} \in X, \tilde{y} \in Y$

- $\tilde{x}_a$: product of probabilities of all actions of P1 on the path to a

- $\tilde{y}_a$: product of probabilities of all actions of P2 on the path to a

$$X := \left\{ \forall j \in \mathcal{J}_1 : \sum_{a \in A_j} \tilde{x}_a = \tilde{x}_{p_j} \right\}, \qquad Y := \left\{ \forall j \in \mathcal{J}_2 : \sum_{a \in A_j} \tilde{y}_a = \tilde{y}_{p_j} \right\}$$

- The payoff to P1 under sequence strategies $\tilde{x} \in X, \tilde{y} \in Y$ is
$$\tilde{x}^\top A \tilde{y}$$

- $A_{a,a'} = $ **if** a was the last action of P1 and a' the last action of P2 before some leaf z, **then** payoff to P1 at z times product of chance probabilities on path to z **else** zero

Recap: From Sequence to Behavioral

- Every sequence form strategy $\tilde{x}$ can be transformed into a behavioral form strategy as (recursively bottom up):

$$\forall a \in A_j: x_a = \frac{\tilde{x}_a}{\tilde{x}_{p_j}}$$

  if info-set is un-reachable, i.e. $\tilde{x}_{p_j} = 0$, then use any behavioral

- Every behavioral strategy x can be transformed into a sequence form strategy as (recursively top down):

$$\forall a \in A_j: \tilde{x}_a = \tilde{x}_{p_j} \cdot x_a$$

Sequence Form Representation

	$\emptyset$	f_*	r_*
f			
r			
f'			
r'			
$\hat{c}$			
$\hat{f}$			

Let's fill it in!
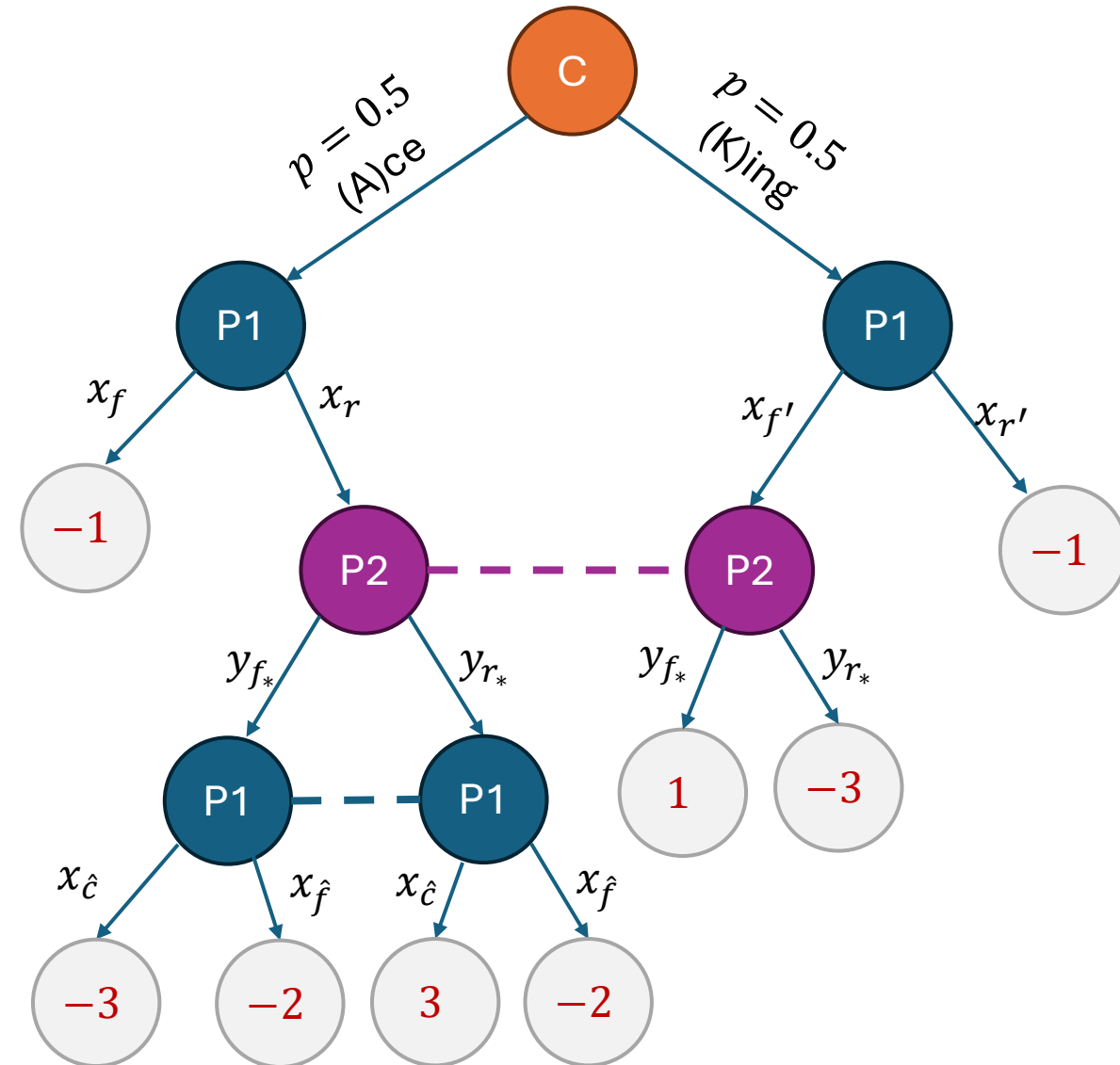
TreePlex Representation of Strategy Space

The strategy space of each player corresponds to a set of interconnected "scaled" simplices

$$\sum_{a \in A_j} \tilde{x}_a = \tilde{x}_{p_j}$$

To generate $\tilde{x}_a$ we can simply generate an element of the simplex (i.e. a behavioral strategy x_a) and then scale all of its coordinates by $\tilde{x}_{p_j}$, i.e. $\tilde{x}_a = \tilde{x}_{p_j} \cdot x_a$
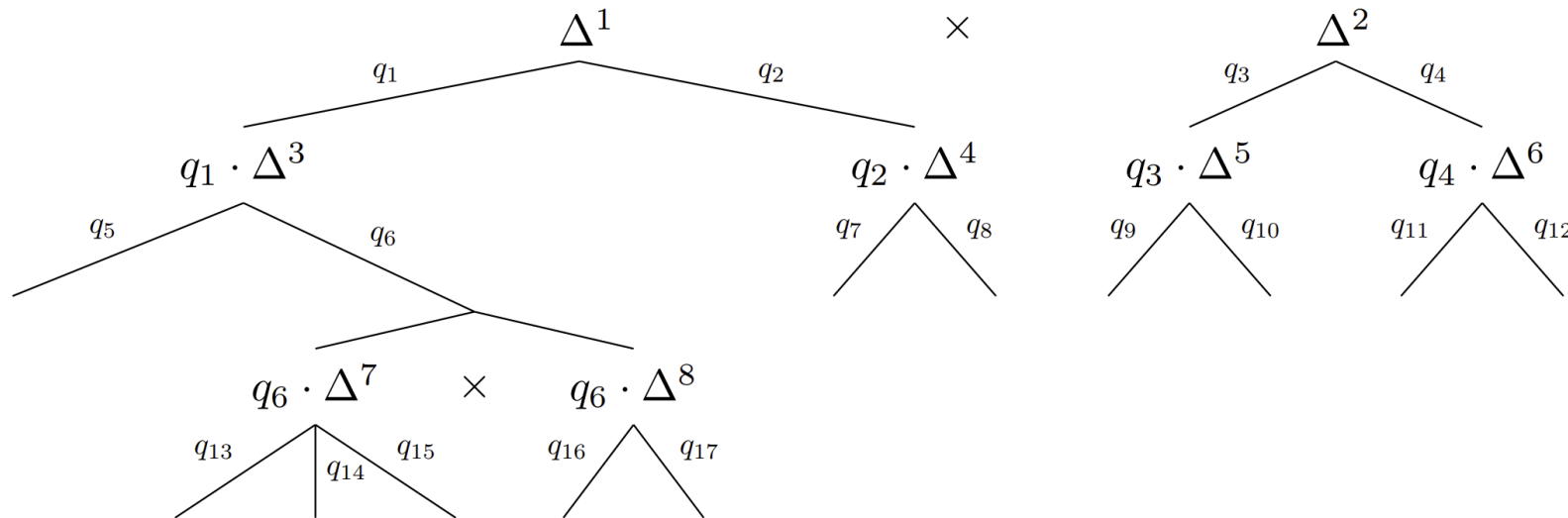
TreePlex Representation
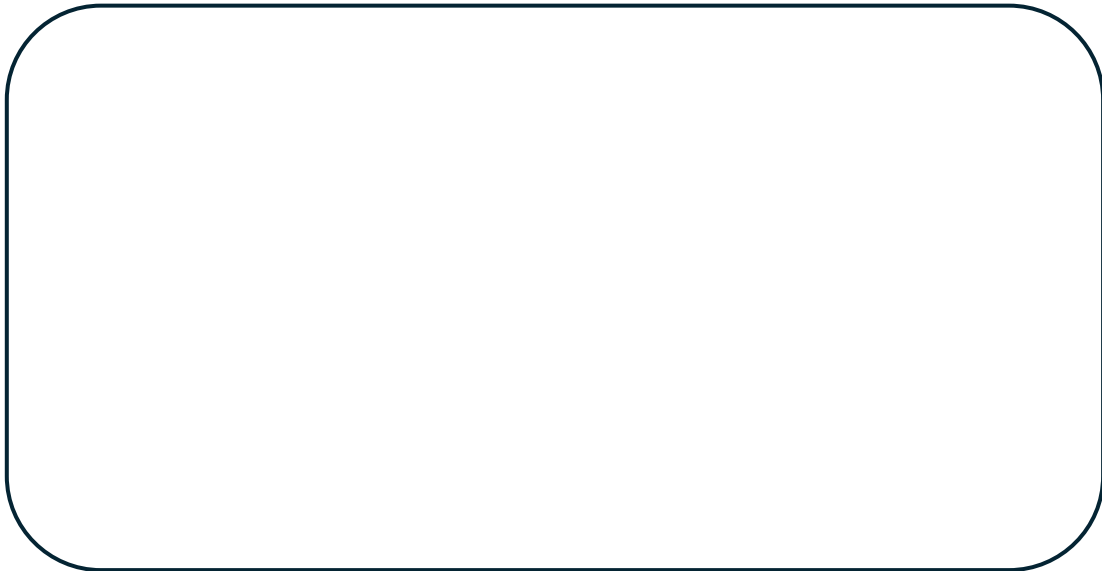
Let's fill it in!
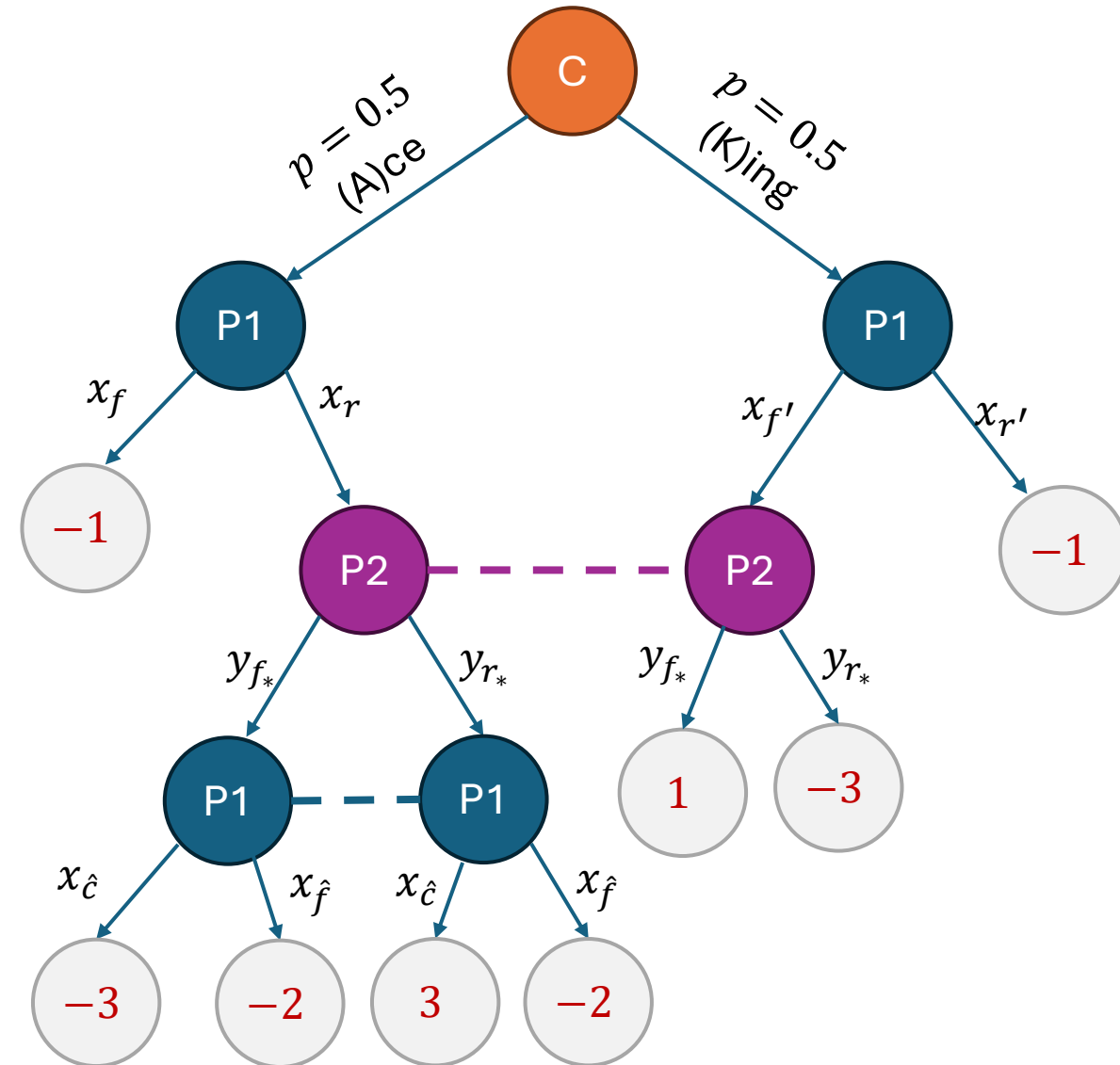
No-Regret Learning in Sequence Form

- We have successfully turned imperfect information extensive form zero-sum games into a familiar object

$$\min_{\tilde{x} \in X} \max_{\tilde{y} \in Y} \tilde{x}^\top A \tilde{y}$$

- X, Y are convex sets, i.e., sequence-form strategies

- We can invoke minimax theorem to prove existence of equilibria

- We can calculate equilibria via LP duality

- We can calculate equilibria via no-regret learning!

Recap from Lecture 2: Regret of FTRL

(FTRL)
$$x_t = \underset{x \in X}{\operatorname{argmin}} \left[\sum_{\tau < t} \langle x, \ell_\tau \rangle \right] + \left[\frac{1}{\eta} \mathcal{R}(x) \right]$$

1-strongly convex function of x that stabilizes the minimizer

Historical performance of always choosing strategy x

Theorem. Assuming the loss function at each period
$$f_t(x) = \langle x, \ell_t \rangle$$

is L-Lipschitz with respect to some norm $\|\cdot\|$ and the regularizer is 1-strongly convex with respect to the same norm then

$$\mathrm{Regret} - \mathrm{FTRL}(T) \leq \boxed{\eta L} + \boxed{\frac{1}{\eta T} \left(\max_{x \in X} \mathcal{R}(x) - \min_{x \in X} \mathcal{R}(x) \right)}$$

Average stability induced by regularizer

Average loss distortion caused by regularizer

Same for utilities

(FTRL)
$$x_t = \underset{x \in X}{\operatorname{argmax}} \boxed{\sum_{\tau < t} \langle x, u_\tau \rangle} - \boxed{\frac{1}{\eta} \mathcal{R}(x)}$$

1-strongly convex function of x that stabilizes the maximizer

Historical performance of always choosing strategy x

Theorem. Assuming the utility function at each period
$$f_t(x) = \langle x, u_t \rangle$$

is L-Lipschitz with respect to some norm $\|\cdot\|$ and the regularizer is 1-strongly convex with respect to the same norm then
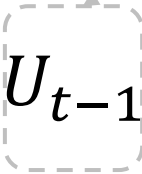
$$\operatorname{Regret} - \operatorname{FTRL}(T) \leq \boxed{\eta L} + \boxed{\frac{1}{\eta T} \left(\max_{x \in X} \mathcal{R}(x) - \min_{x \in X} \mathcal{R}(x) \right)}$$

Average stability induced by regularizer

Average loss distortion caused by regularizer

Regularizer for the Treeplex Space X

- The only thing we are missing is a good Regularizer for X

$$U_{t-1} = \sum_{\tau < t} u_\tau$$

- **Desiderata.** Be strongly convex in x within X and for the optimization problem to be fast to solve

$$\tilde{x}_t = \operatorname*{argmax}_{\tilde{x} \in X} \sum_{\tau < t} \langle \tilde{x}, u_t \rangle - \frac{1}{\eta} \mathcal{R}(\tilde{x}) = \operatorname*{argmax}_{\tilde{x} \in X} \langle \tilde{x}, U_{t-1} \rangle - \frac{1}{\eta} \mathcal{R}(\tilde{x})$$

- X is no longer a "simplex", so entropy is not a good Regularizer

Dilated Entropy

- X is a combination of *scaled simplices*, i.e., $\tilde{x} = \left(\tilde{x}^j\right)_{j \in \mathcal{J}_1}$
- $\tilde{x}^j = (\tilde{x}_a)_{a \in A_j}$: sequence-form strategies for actions in infoset $j \in \mathcal{J}_1$

$$\tilde{x}^j \in \tilde{x}_{p_j} \cdot \Delta_j \qquad \Leftrightarrow \qquad \tilde{x}^j / \tilde{x}_{p_j} \in \Delta_j$$

- Consider a *weighted combination of local negative entropies*

$$\mathcal{R}(\tilde{x}) := \sum_j \beta_j \, \tilde{x}_{p_j} \, \mathrm{H}\left(\tilde{x}^j / \tilde{x}_{p_j}\right), \qquad \mathrm{H}(u) = \sum_i u_i \log(u_i)$$

Lies in a simplex
Equivalent to the behavioral strategy x^j

Negative Entropy

- $\mathcal{R}(\tilde{x})$ is $1/M$ strongly convex w.r.t. ℓ_1 norm, where $M = \max_{\tilde{x} \in X} \|x\|_1$, for appropriate choice of β_j based on game tree structure

Solving the Optimization Problem

- Optimization problem decomposes into local simplex problems

$$\sum_{j \in \mathcal{J}_1} \left\langle \tilde{x}^j, U_{t-1}^j \right\rangle - \frac{1}{\eta} \beta_j \, \tilde{x}_{p_j} \, \mathrm{H}\left(\frac{\tilde{x}^j}{\tilde{x}_{p_j}}\right) = \sum_{j \in \mathcal{J}_1} \tilde{x}_{p_j} \left\{ \left\langle \frac{\tilde{x}^j}{\tilde{x}_{p_j}}, U_{t-1}^j \right\rangle - \frac{1}{\eta} \beta_j \, \mathrm{H}\left(\frac{\tilde{x}^j}{\tilde{x}_{p_j}}\right) \right\}$$

- Max of quantity $\dfrac{\tilde{x}^j}{\tilde{x}_{p_j}}$ over simplex Δ_j is independent of solution x_a for all ancestral actions

- Note quantity $\dfrac{\tilde{x}^j}{\tilde{x}_{p_j}}$ is essentially the behavioral strategy x^j at infoset j

Solving the Optimization Problem

- Decomposes in local max over behavioral strategies x^j solved recursively bottom up

$$V^j = \max_{x^j \in \Delta_j} \left\langle x^j, U^j_{t-1} \right\rangle - \frac{1}{\eta} \beta_j \, H(x^j) \Rightarrow x^j \propto \exp\left(\frac{\eta}{\beta_j} U^j_{t-1}\right)$$

- Value V^j at the maximum multiplies x_{p_j} and is appended to the "utility vector" associated with p_j in the parent infoset

$$V^j = \log \sum_{a \in A_j} \exp\left(\frac{\eta}{\beta_j} U^a_{t-1}\right) \approx \mathrm{softmax}\left(\frac{\eta}{\beta_j} U^j_{t-1}\right)$$

Recap: Nash via FTRL with Dilated Entropy

Each player chooses $\tilde{x}_t, \tilde{y}_t$ based on FTRL with dilated entropy

- For x-player $u_t = A\tilde{y}_t$ and $U_t = U_{t-1} + u_t$ and initialize $Q = U_t$
- Traverse the tree bottom-up; for each infoset $j \in \mathcal{J}_1$

$$x_{t+1}^j \propto \exp\left(\eta_j Q^j\right), \qquad V^j = \mathrm{softmax}\left(\eta_j Q^j\right), \qquad Q_{p_j} \leftarrow Q_{p_j} + V^j$$

- Define sequence-form strategies top-down: $\tilde{x}_{t+1}^j = \tilde{x}_{p_j} \cdot x_{t+1}^j$

- Similarly, for y player

Return average of sequence-form strategies as equilibrium

Fast Rates

Theorem. If we use Optimistic FTRL instead of FTRL then we get faster convergence to a Nash equilibrium at rate $1/T$ instead of $1/\sqrt{T}$. Plus, we get last-iterate convergence instead of only average iterate convergence.

Interpreting utility vector

$$u_{t,a} = A\tilde{y}_t = \sum_{a' \in A_{P2}} A_{a,a'} \tilde{y}_{t,a'}$$

$A_{a,a'}$ is zero if the combination of a, a' does not lead to a leaf node

$$u_{t,a} = \sum_{\substack{\text{Leafs } z: \ a \text{ was last P1 action} \\ a' \text{ was last P2 action}}} u(z) \Pr \begin{pmatrix} \text{Chance chooses} \\ \text{sequence on} \\ \text{path to } z \end{pmatrix} \Pr \begin{pmatrix} \text{P2 plays} \\ \text{sequence} \\ \text{leading to } a' \end{pmatrix}$$

Illustration: First Step of Dynamics

- Go to **Infoset 3**

$$Q^3 = \left(u_{\hat{c}}, u_{\hat{f}} \right) = \left(\right)$$

Illustration: First Step of Dynamics

- Go to **Infoset 3**

$$Q^3 = \left(u_{\hat{c}}, u_{\hat{f}} \right) = \left(-3\frac{1}{2} y_{f_*} + 3\frac{1}{2} y_{r_*}, -2\frac{1}{2} y_{f_*} - 2\frac{1}{2} y_{r_*} \right)$$

$$x^3 = \left(x_{\hat{c}}, x_{\hat{f}} \right) \propto \exp(\eta_3 \, Q^3)$$
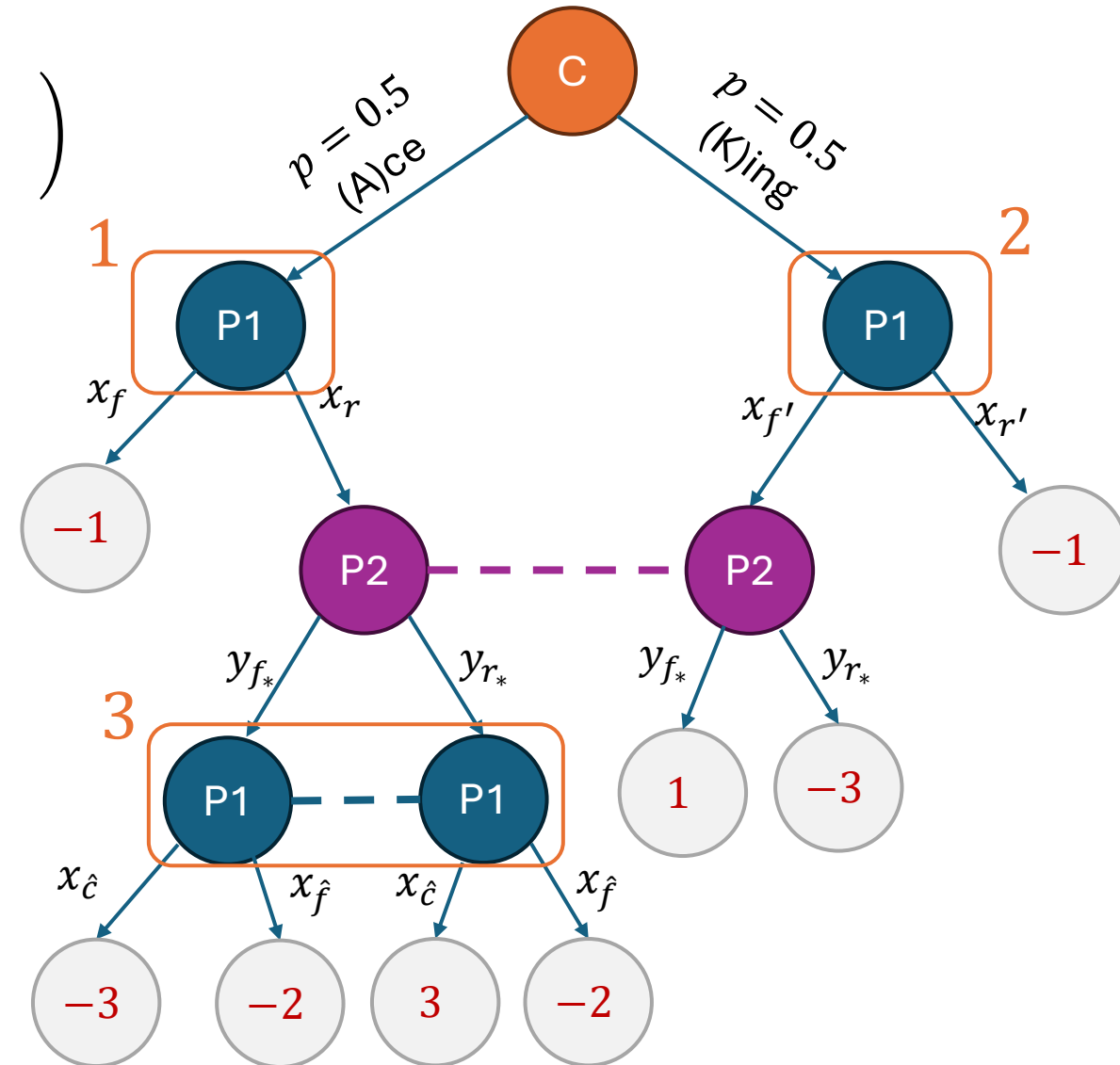
Illustration: First Step of Dynamics

- Go to **Infoset 3**

$$Q^3 = \left(u_{\hat{c}}, u_{\hat{f}}\right) = \left(-3\frac{1}{2}y_{f_*} + 3\frac{1}{2}y_{r_*}, -2\frac{1}{2}y_{f_*} - 2\frac{1}{2}y_{r_*}\right)$$

$$x^3 = \left(x_{\hat{c}}, x_{\hat{f}}\right) \propto \exp(\eta_3 \, Q^3)$$

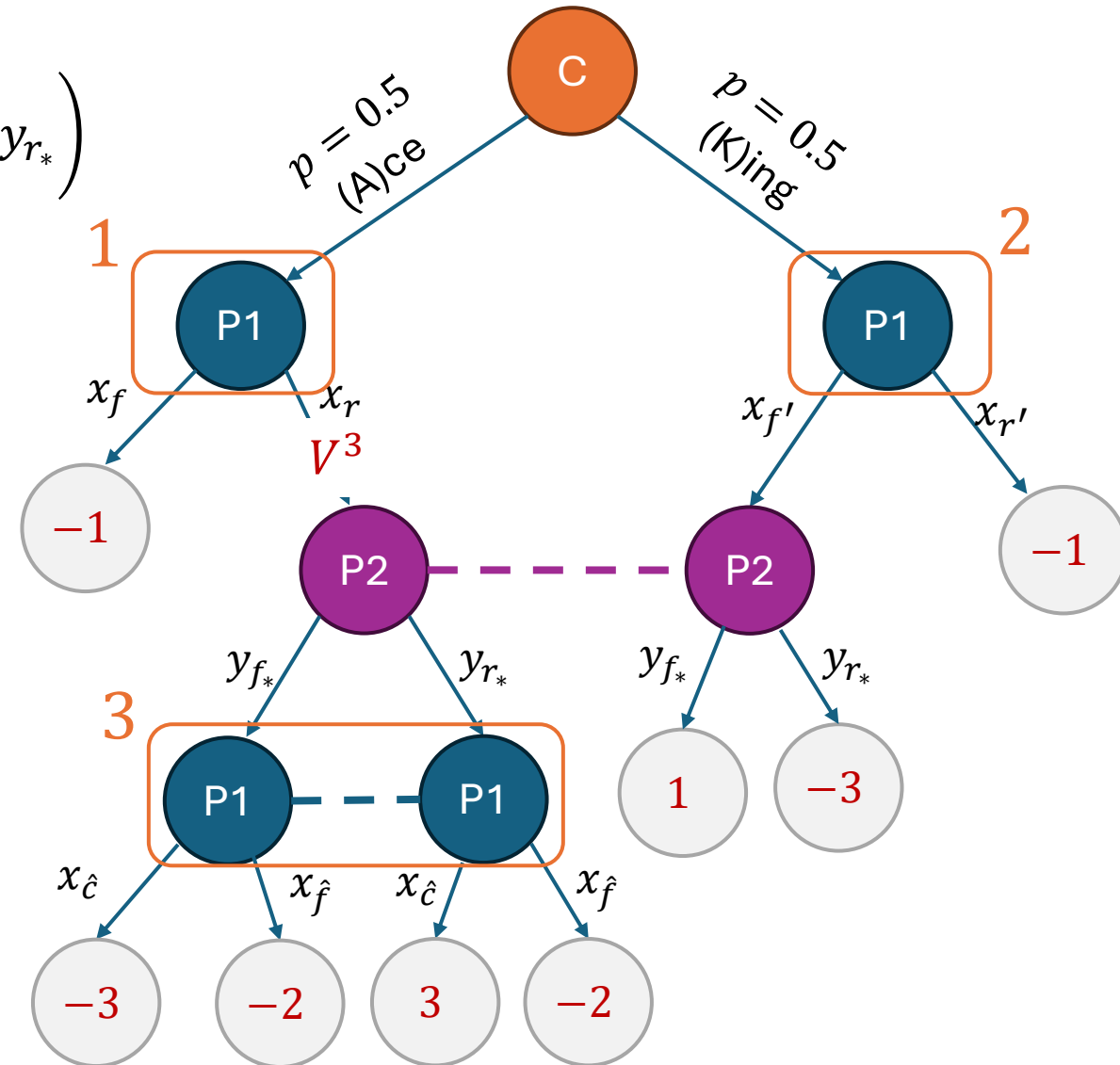$$V^3 = \mathrm{softmax}(\eta_3 Q^3)$$

Illustration: First Step of Dynamics

- Go to **Infoset 3**

$$Q^3 = \left(u_{\hat{c}}, u_{\hat{f}}\right) = \left(-3\frac{1}{2}y_{f_*} - 2\frac{1}{2}y_{r_*}, 3\frac{1}{2}y_{f_*} - 2\frac{1}{2}y_{r_*}\right)$$

$$x^3 = \left(x_{\hat{c}}, x_{\hat{f}}\right) \propto \exp(\eta_3\, Q^3)$$

$$V^3 = \text{softmax}(\eta_3 Q^3)$$

- Go to **Infoset 1**

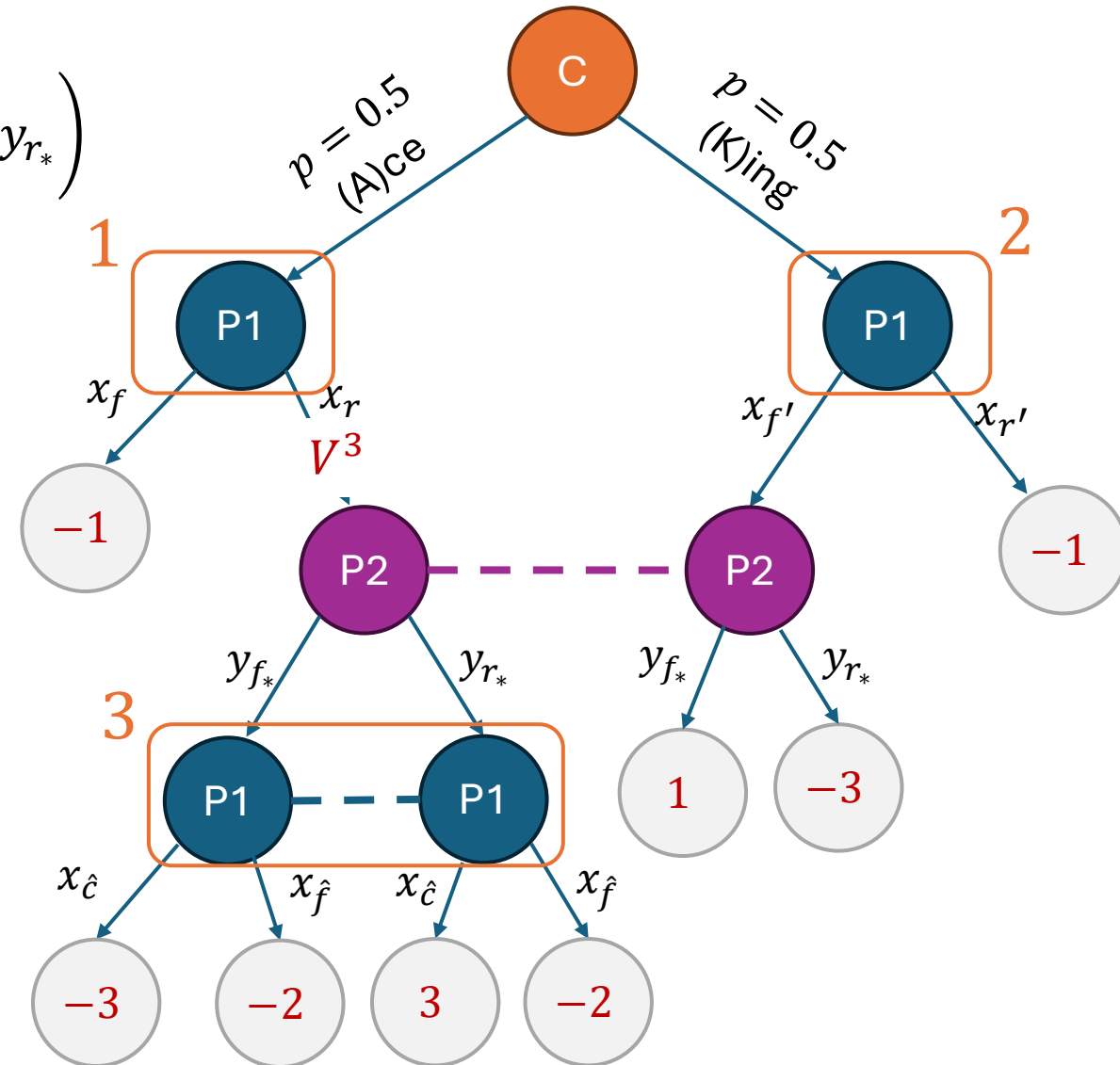$$Q^1 = \left(u_f, u_r + V^3\right) = \begin{pmatrix} & \\ & \end{pmatrix}$$

Illustration: First Step of Dynamics

- Go to **Infoset 3**

$$Q^3 = \left(u_{\hat{c}}, u_{\hat{f}}\right) = \left(-3\frac{1}{2}y_{f_*} - 2\frac{1}{2}y_{r_*}, 3\frac{1}{2}y_{f_*} - 2\frac{1}{2}y_{r_*}\right)$$

$$x^3 = \left(x_{\hat{c}}, x_{\hat{f}}\right) \propto \exp(\eta_3 \, Q^3)$$

$$V^3 = \text{softmax}(\eta_3 Q^3)$$

- Go to **Infoset 1**

$$Q^1 = \left(u_f, u_r + V^3\right) = \left(-1\frac{1}{2}, V^3\right)$$

$$x^1 = \left(x_f, x_r\right) \propto \exp(\eta_1 \, Q^1)$$

- Go to **Infoset 2**

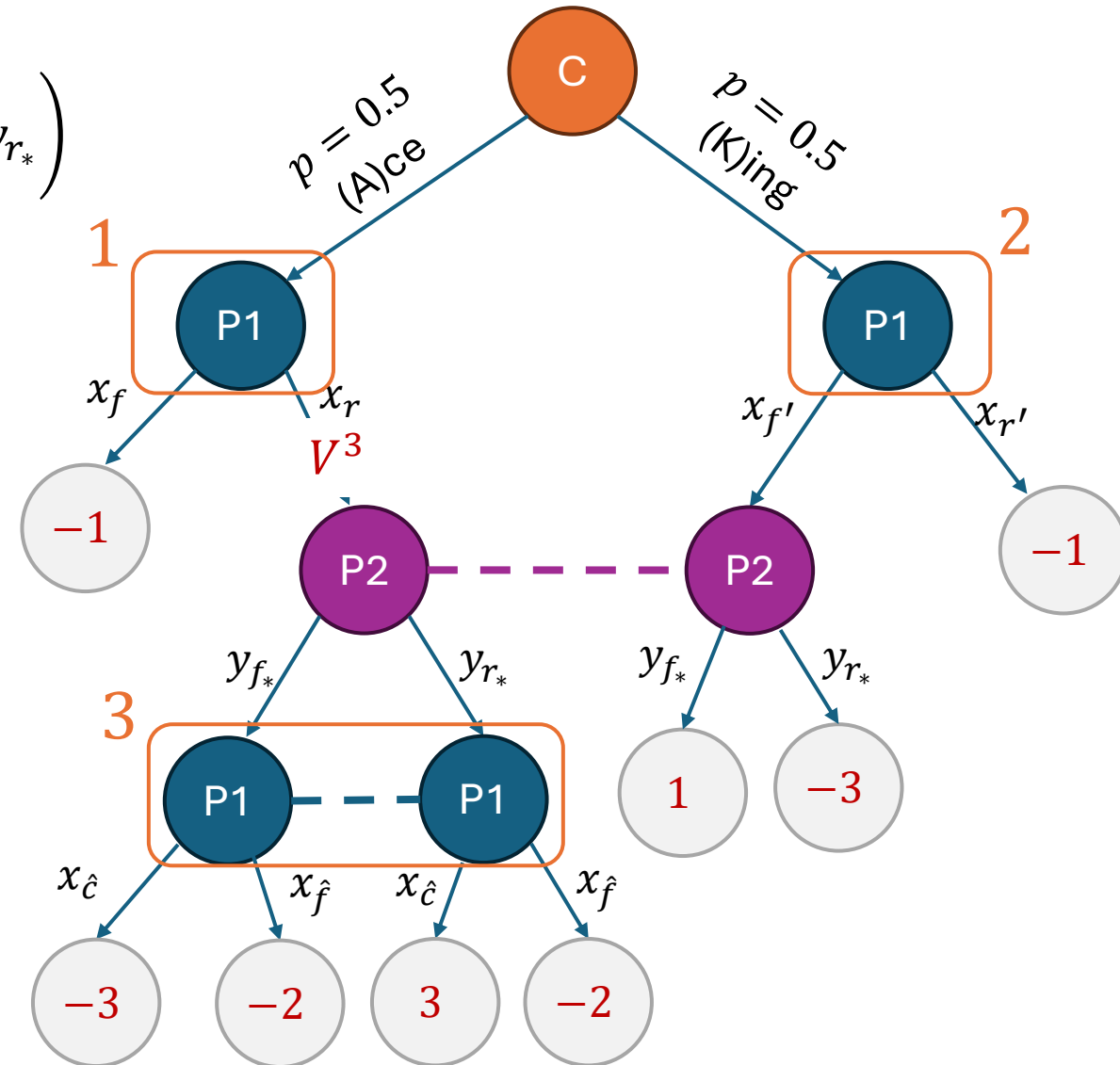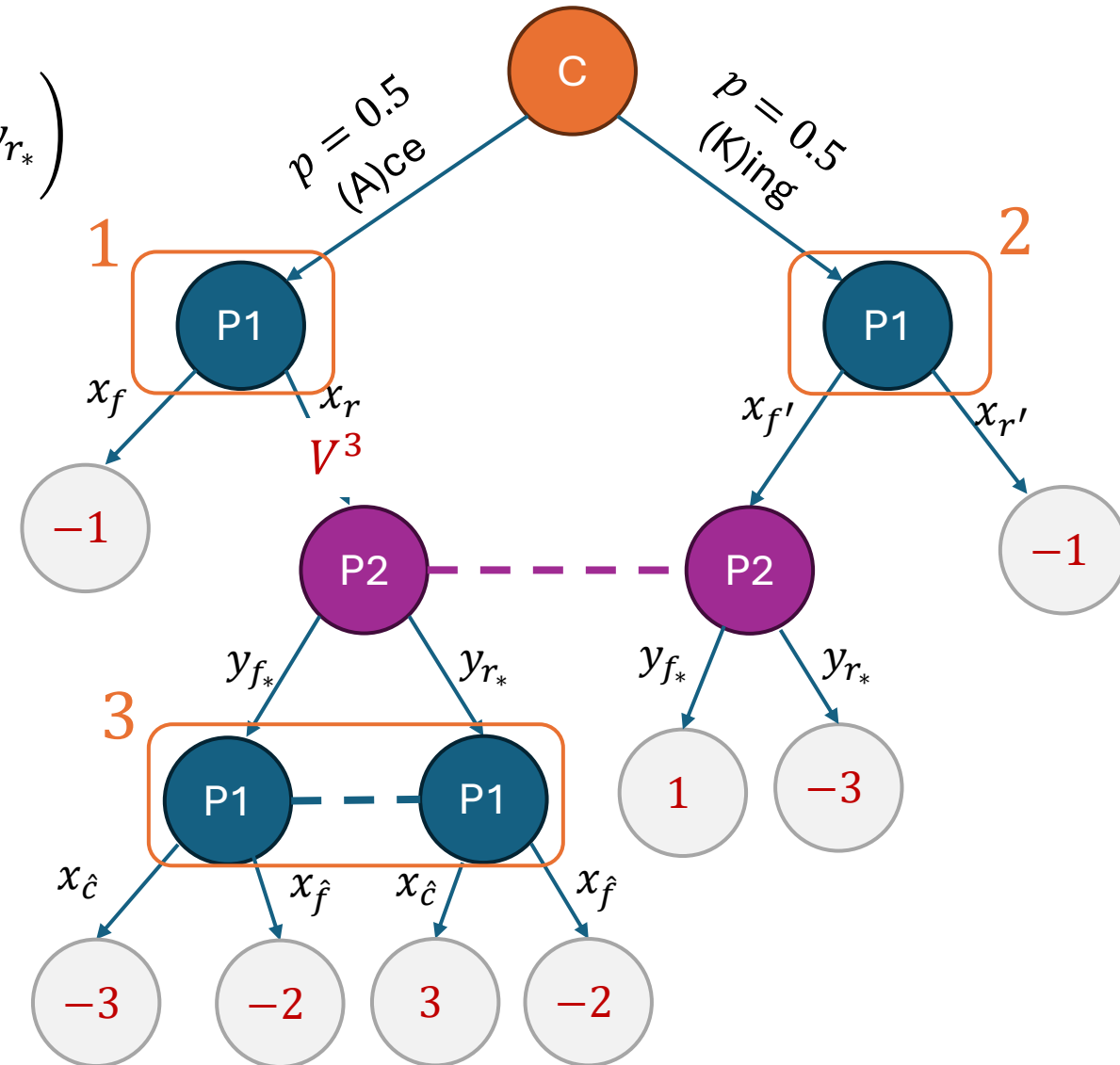$$Q^2 = \left(u_{f'}, u_{r'}\right) = \left(\right)$$

Illustration: First Step of Dynamics

- Go to **Infoset 3**

$$Q^3 = \left(u_{\hat{c}}, u_{\hat{f}}\right) = \left(-3\frac{1}{2}y_{f_*} - 2\frac{1}{2}y_{r_*}, 3\frac{1}{2}y_{f_*} - 2\frac{1}{2}y_{r_*}\right)$$

$$x^3 = \left(x_{\hat{c}}, x_{\hat{f}}\right) \propto \exp(\eta_3\, Q^3)$$

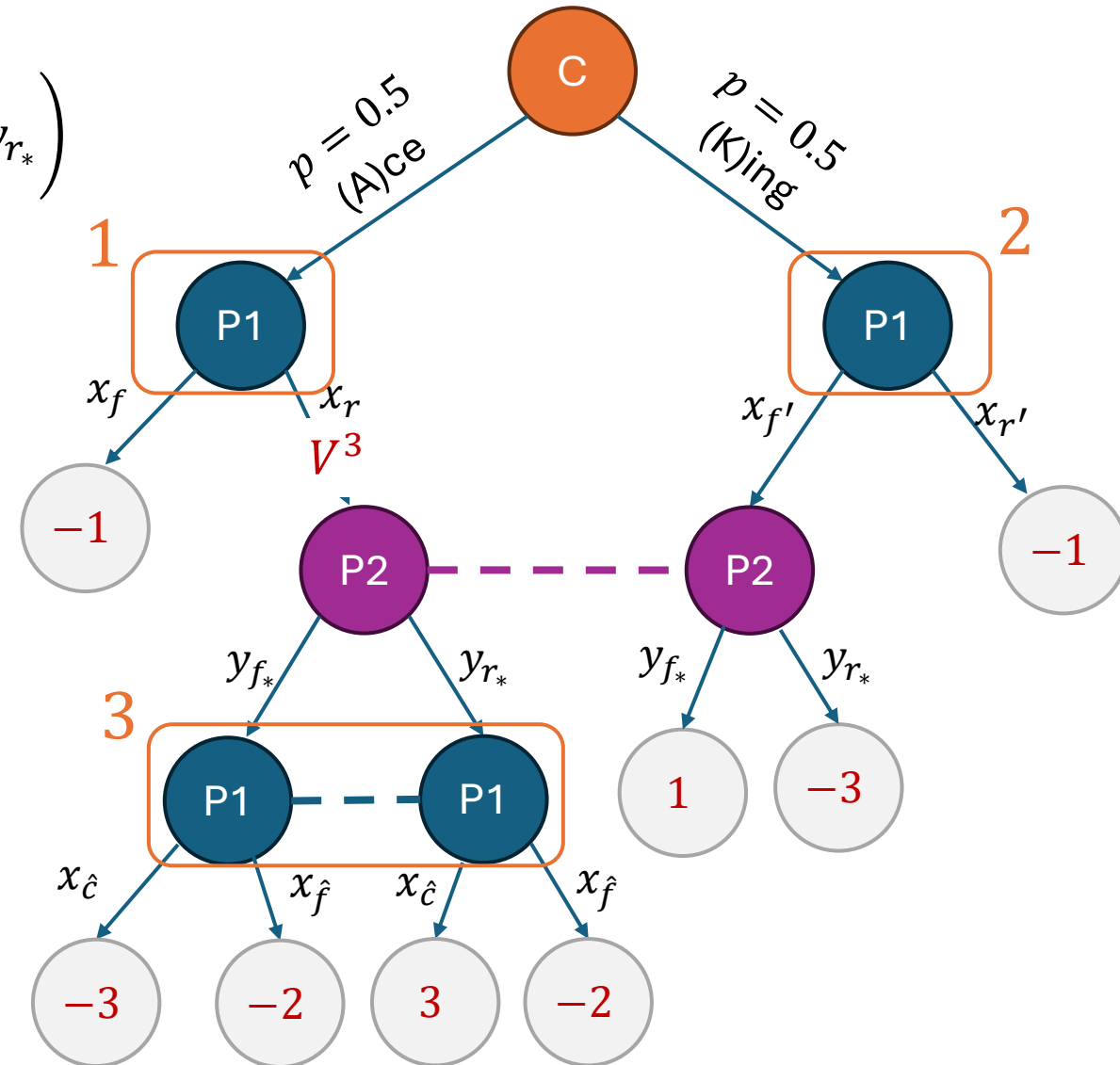$$V^3 = \mathrm{softmax}(\eta_3 Q^3)$$

- Go to **Infoset 1**

$$Q^1 = \left(u_f, u_r + V^3\right) = \left(-1\frac{1}{2}, V^3\right)$$

$$x^1 = \left(x_f, x_r\right) \propto \exp(\eta_1\, Q^1)$$

- Go to **Infoset 2**

$$Q^2 = \left(u_{f'}, u_{r'}\right) = \left(1\frac{1}{2}y_{f_*} - 3\frac{1}{2}y_{r_*}, -1\frac{1}{2}\right)$$

$$x^2 = \left(x_{f'}, x_{r'}\right) \propto \exp(\eta_2 Q^2)$$

Monte-Carlo Stochastic Approximation of Utilities

- Calculating utilities on all nodes of the tree can be very expensive
- In linear online learning it suffices that we use an unbiased estimate of the utility vector

$$\tilde{x}_t = \operatorname*{argmax}_{x \in X} \sum_{\tau < t} \langle x, \hat{u}_\tau \rangle - \frac{1}{\eta} \mathcal{R}(x), \qquad E[\, \hat{u}_\tau \mid F_\tau \,] = u_\tau$$

All random variables observed before period τ

- By standard martingale concentration inequality arguments, the error vanishes with the number of iterations (*we will see later*)
- In this setting, it suffices that we "sample a path for opponent" and that we "sample chance moves"

Illustration: First Step of Dynamics

- Sample chance moves based on fixed distribution and opponent moves based on y_t

- Suppose, we sampled A and f_*

- Go to **Infoset 3**

$$\hat{Q}^3 = \left(\hat{u}_{\hat{c}}, \hat{u}_{\hat{f}}\right) = (-3, -2)$$

$$x^3 = \left(x_{\hat{c}}, x_{\hat{f}}\right) \propto \exp(\eta_3 \, \hat{Q}^3)$$

$$\hat{V}^3 = \text{softmax}(\eta_3 \hat{Q}^3)$$

- Go to **Infoset 1**

$$\hat{Q}^1 = \left(\hat{u}_f, \hat{u}_r + \hat{V}^3\right) = (-1, V^3)$$

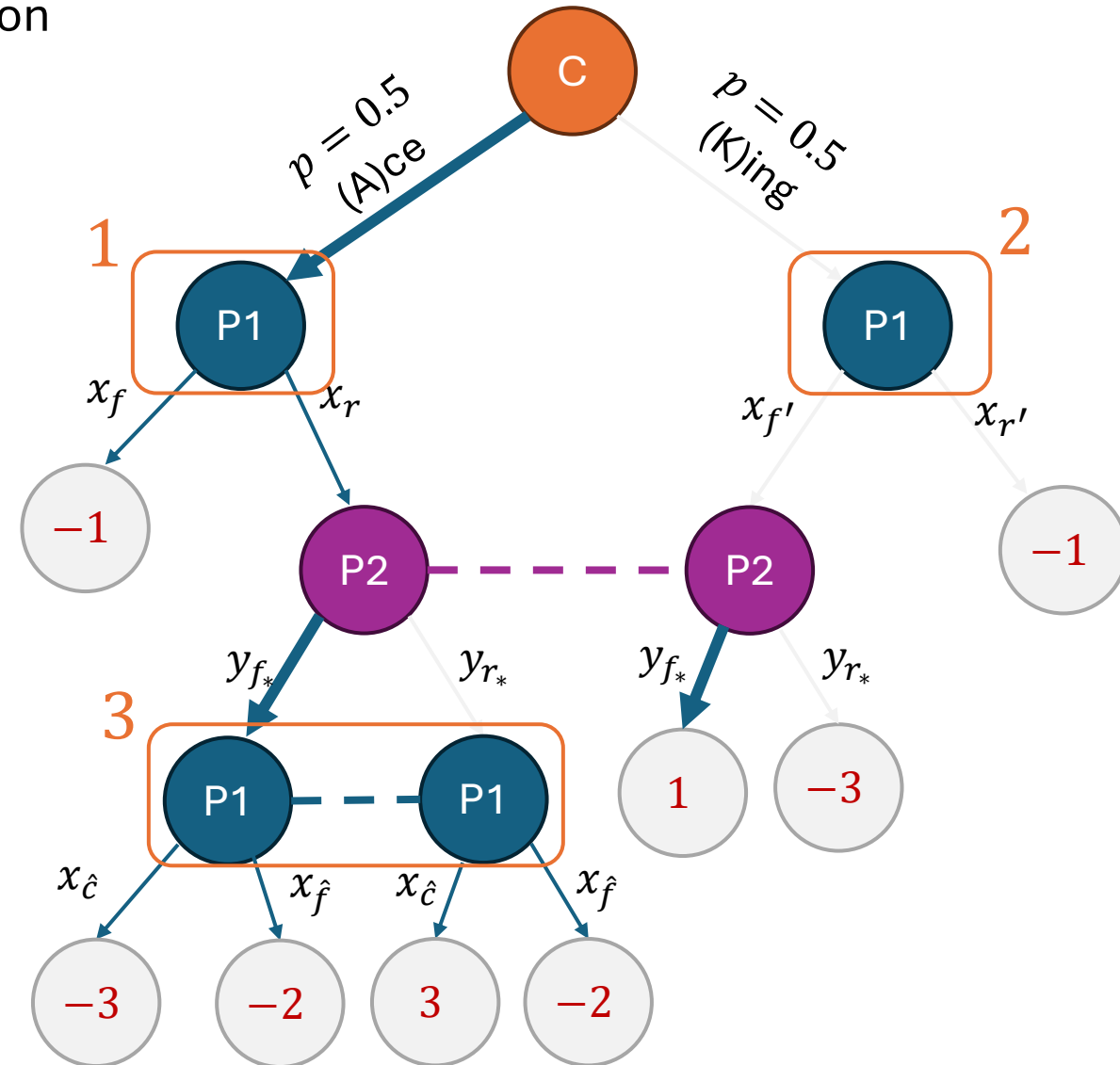$$x^1 = \left(x_f, x_r\right) \propto \exp(\eta_1 \, Q^1)$$
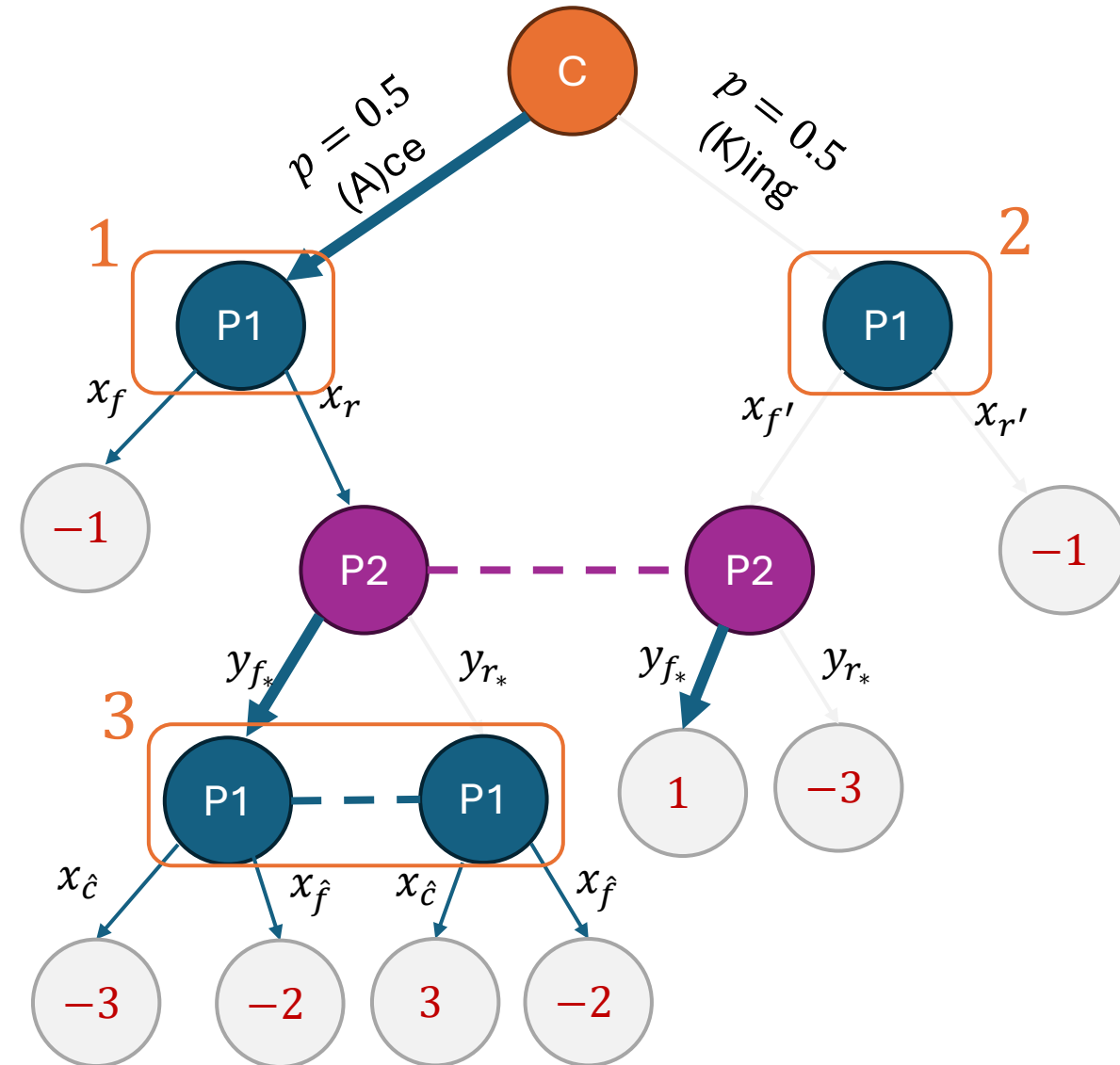
Illustration: First Step of Dynamics

- Equivalently we can do top down and evaluate recursively, sampling when needed

- Sample chance move (e.g. sampled A)

- Go to **Infoset 1**
$$\hat{Q}_f = -1$$

- For $\widehat{Q}_r$ recursively go down tree

- Sample P2 move (e.g. sampled y_{f_*})

- Go down to **Infoset 3**
$$\hat{Q}^3 = (-3, -2)$$
$$x^3 = \left(x_{\hat{c}}, x_{\hat{f}}\right) \propto \exp\left(\eta_3 \, \hat{Q}^3\right)$$
$$V^3 = \text{softmax}(\eta_3 \hat{Q}^3)$$

- Go back up to **Infoset 1;** set $Q_r = V^3$
$$\hat{Q}^1 = \left(\hat{Q}_f, \hat{Q}_r\right) = \left(-1, V^3\right)$$
$$x^1 = (x_f, x_r) \propto \exp\left(\eta_1 \, Q^1\right)$$

Local Dynamics

- These dynamics seem to be doing "local updates" at each node

- They came out of a specific algorithm FTRL with Dilated Entropy

- Is this a general paradigm?

- Can we decompose the no-regret learning problem into local no-regret learners at each node?

- What feedback should each node receive from the learners in nodes below?

- What loss should each learner be optimizing?

Next class: Counterfactual Regret Minimization (CRM)