Counterfactual Regret Minimization and Domination in Extensive-Form Games

Richard Gibson

University of Alberta

Edmonton, Alberta, Canada

University of Alberta Computer Poker Research Group

Alberta Innovates

NSERC CRSNG

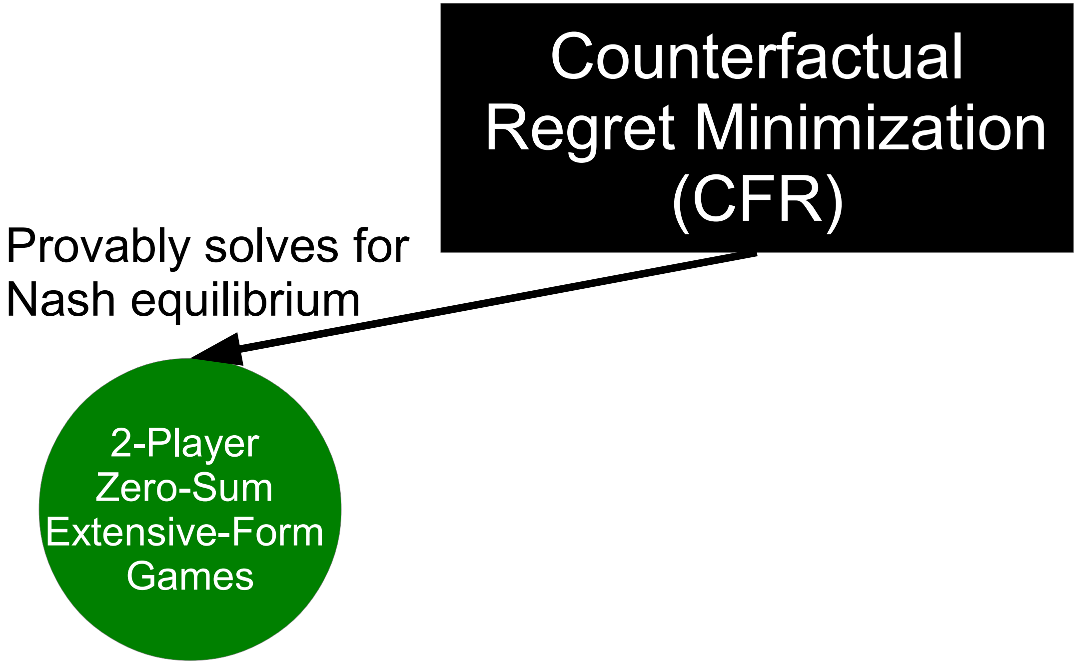
compute • calcul CANADA

WestGrid

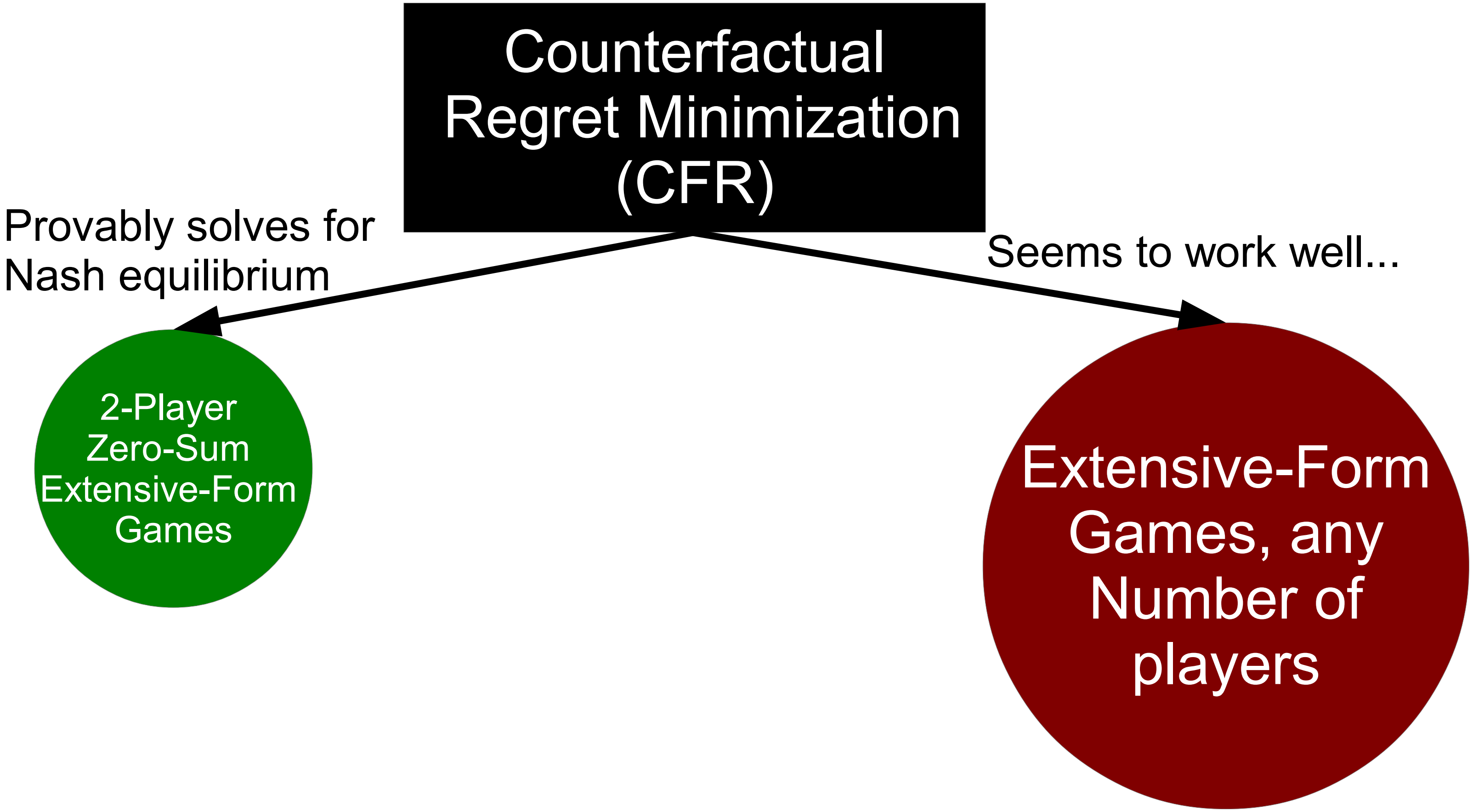
Overview

Counterfactual Regret Minimization (CFR)

Overview

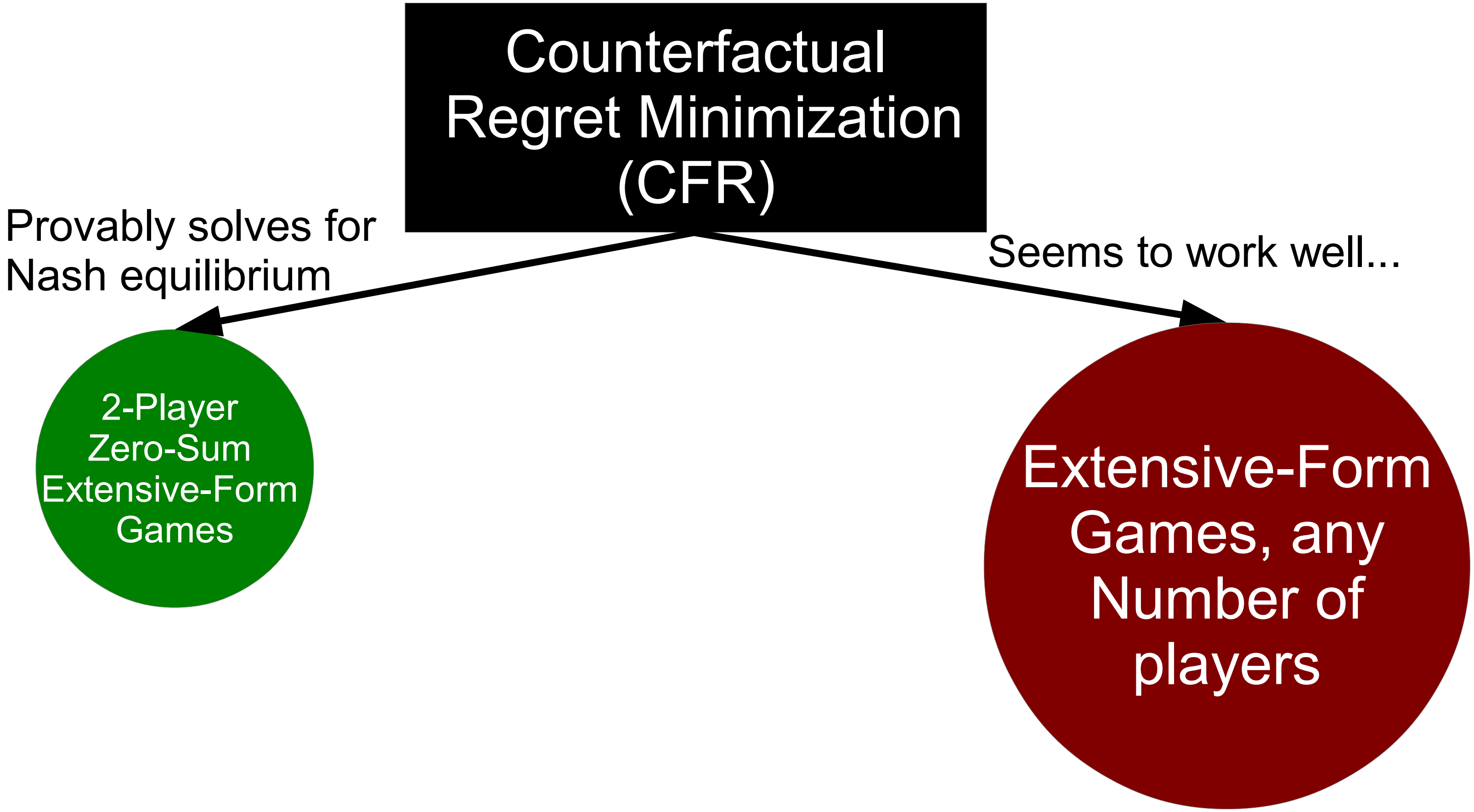

Counterfactual Regret Minimization (CFR)
Provably solves for Nash equilibrium
2-Player Zero-Sum Extensive-Form Games

Overview



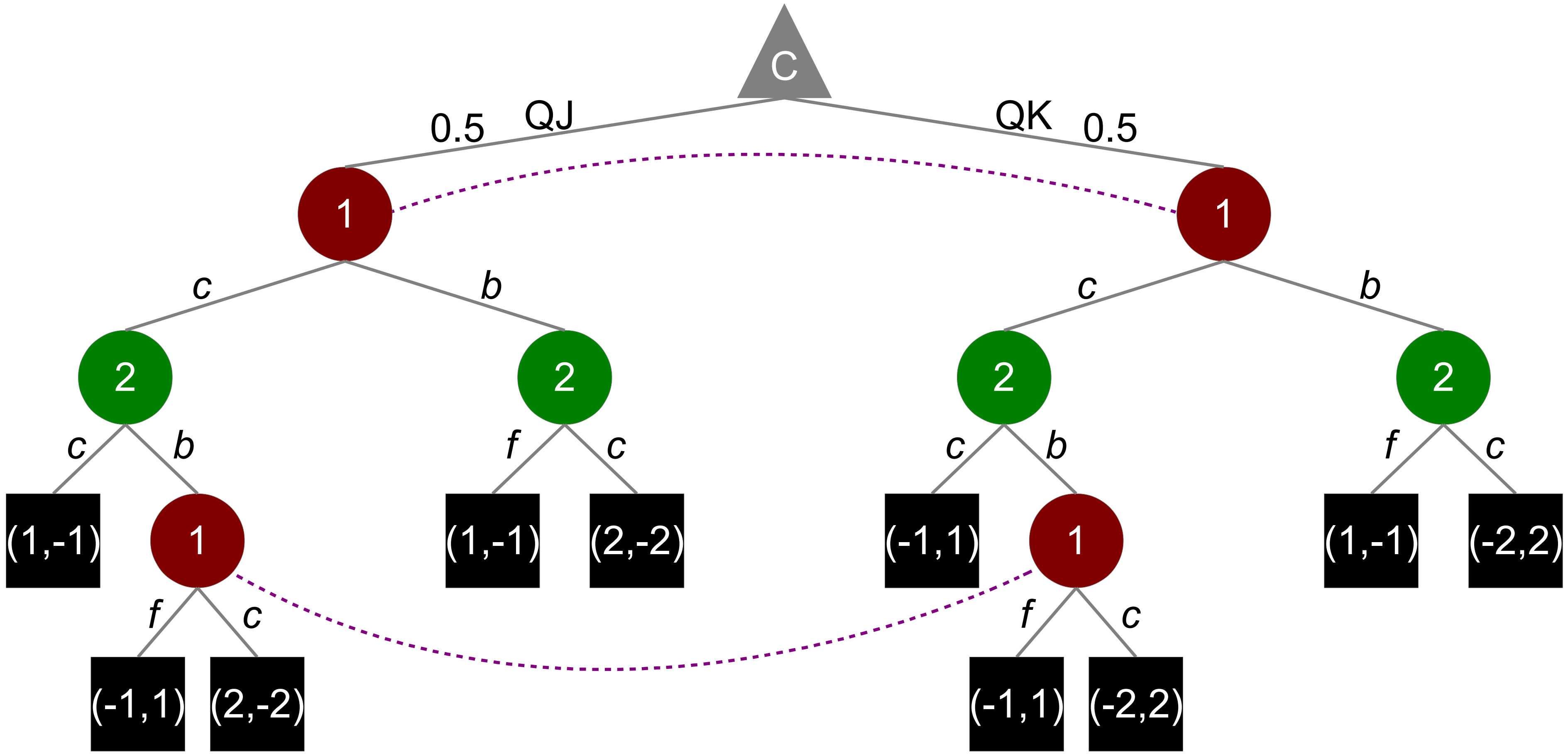
Counterfactual Regret Minimization (CFR)
Provably solves for Nash equilibrium
Seems to work well...
2-Player Zero-Sum Extensive-Form Games
Extensive-Form Games, any Number of players

Overview

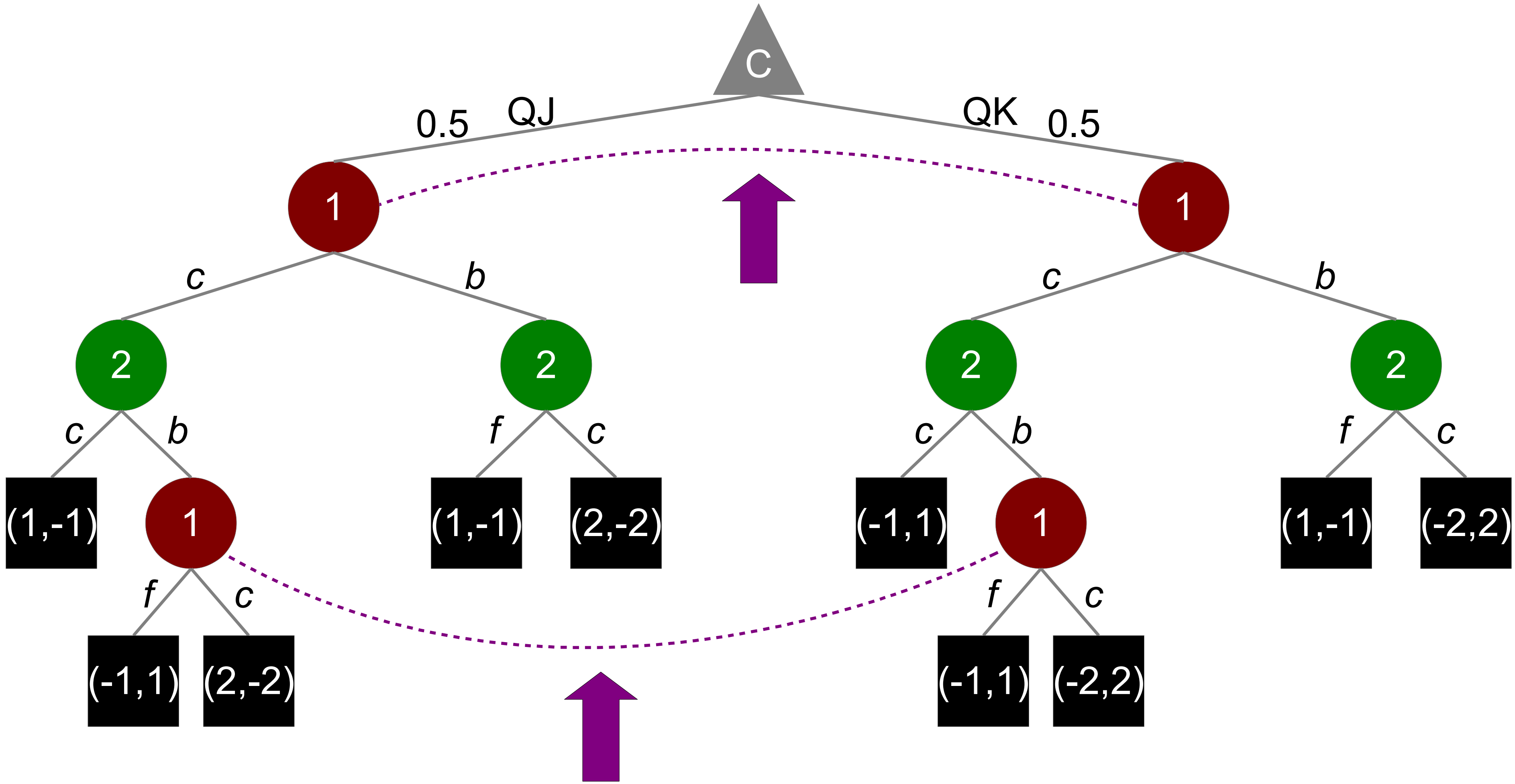


Question: Why do CFR strategies work well in extensive-form games outside of the 2-player zero-sum case?

Extensive-Form Games

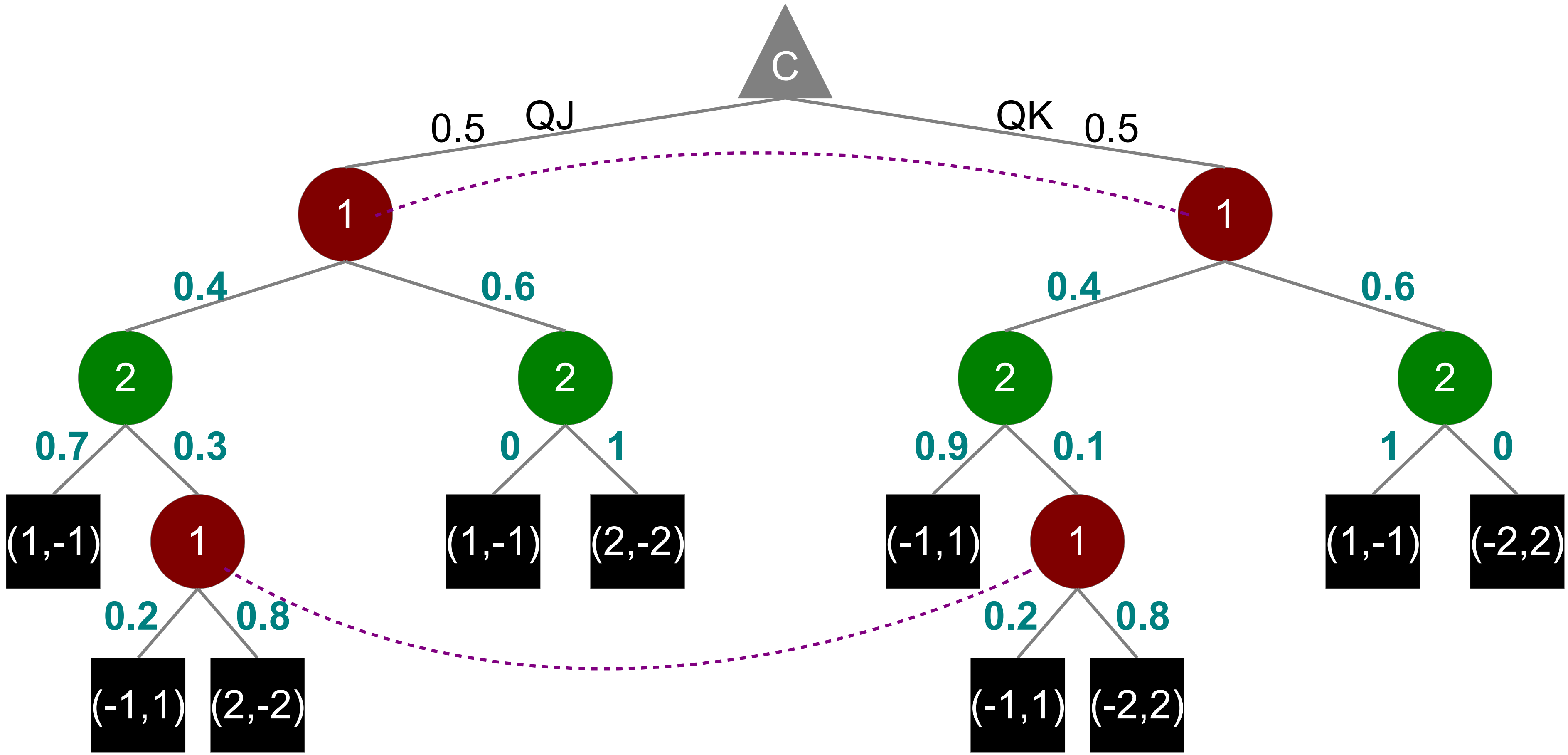


Extensive-Form Games



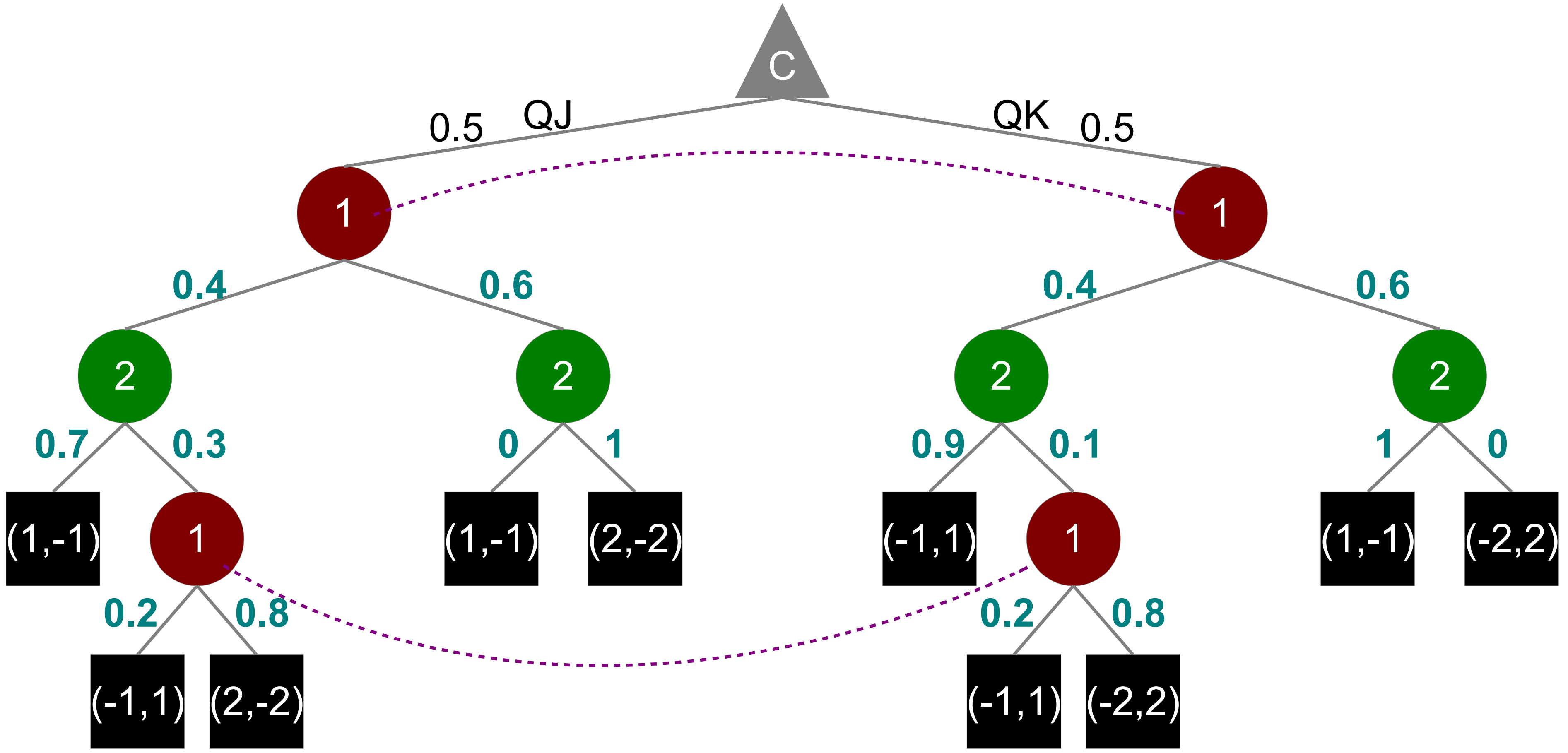
Information sets group states that are indistinguishable to the player.

Extensive-Form Games



A **strategy profile $\sigma = (\sigma_1, \sigma_2)$** assigns a probability distribution over actions at each information set. Example: Probability player 1 checks is $\sigma_1($ Q?, c $)$ = **0.4**.

Extensive-Form Games



$u_i(\sigma)$ is the expected utility for player i, assuming players play according to σ.

Counterfactual Regret Minimization (CFR)

[Zinkevich *et al.*, NIPS 2007]

- CFR is an iterative algorithm that generates strategy profiles (σ^1, σ^2, ... , σ^T) over many iterations T.
- Final output of CFR: σ^{AVG} = Average(σ^1, σ^2, ... , σ^T).
- For 2-player zero-sum games, σ^{AVG} is an ϵ-Nash equilbrium, with $\epsilon \rightarrow 0$ as $T \rightarrow \infty$:

$$u_1(\sigma_1^{AVG}, \sigma_2^{AVG}) \geq \max_{\sigma_1^*} u_1(\sigma_1^*, \sigma_2^{AVG}) - \epsilon$$

$$u_2(\sigma_1^{AVG}, \sigma_2^{AVG}) \geq \max_{\sigma_2^*} u_2(\sigma_1^{AVG}, \sigma_2^*) - \epsilon$$

Counterfactual Regret Minimization (CFR)

- Outside of 2-player zero-sum games, σ^{AVG} is not necessarily an approximate Nash equilibrium [Abou Risk and Szafron, AAMAS 2010].
  - A player may gain by deviating from σ^{AVG}.
- In these games, a Nash equilibrium might not be the most appropriate solution concept anyways.
- On the other hand, σ^{AVG} performs very well in practice...

Annual Computer Poker Competition

3-Player Limit Hold'em - 2009

Agent	Instant Run-off: Round 0
Hyperborean-Eqm	319 ± 2
Hyperborean-BR	299 ± 2
akuma	151 ± 2
dpp	171 ± 2
CMURingLimit	-37 ± 2
dcu3pl	-63 ± 2
Bluechip	-548 ± 2

3-Player Limit Hold'em - 2010

Agent	Instant Run-off: Round 0
Hyperborean.iro	144 ± 32
dcu3pl.tbr	98 ± 30
LittleRock	65 ± 35
Arnold3	-135 ± 39
Bender	-172 ± 16

3-Player Limit Hold'em - 2011

Agent	Instant Run-off: Round 0
Sartre3p	243 ± 20
Hyperborean-3p-limit-iro	204 ± 20
LittleRock	113 ± 19
AAIMontybot	96 ± 44
dcubot3plr	77 ± 19
OwnBot	-4 ± 30
Bnold3	-91 ± 22
Entropy	-108 ± 36
player.zeta.3p	-530 ± 33

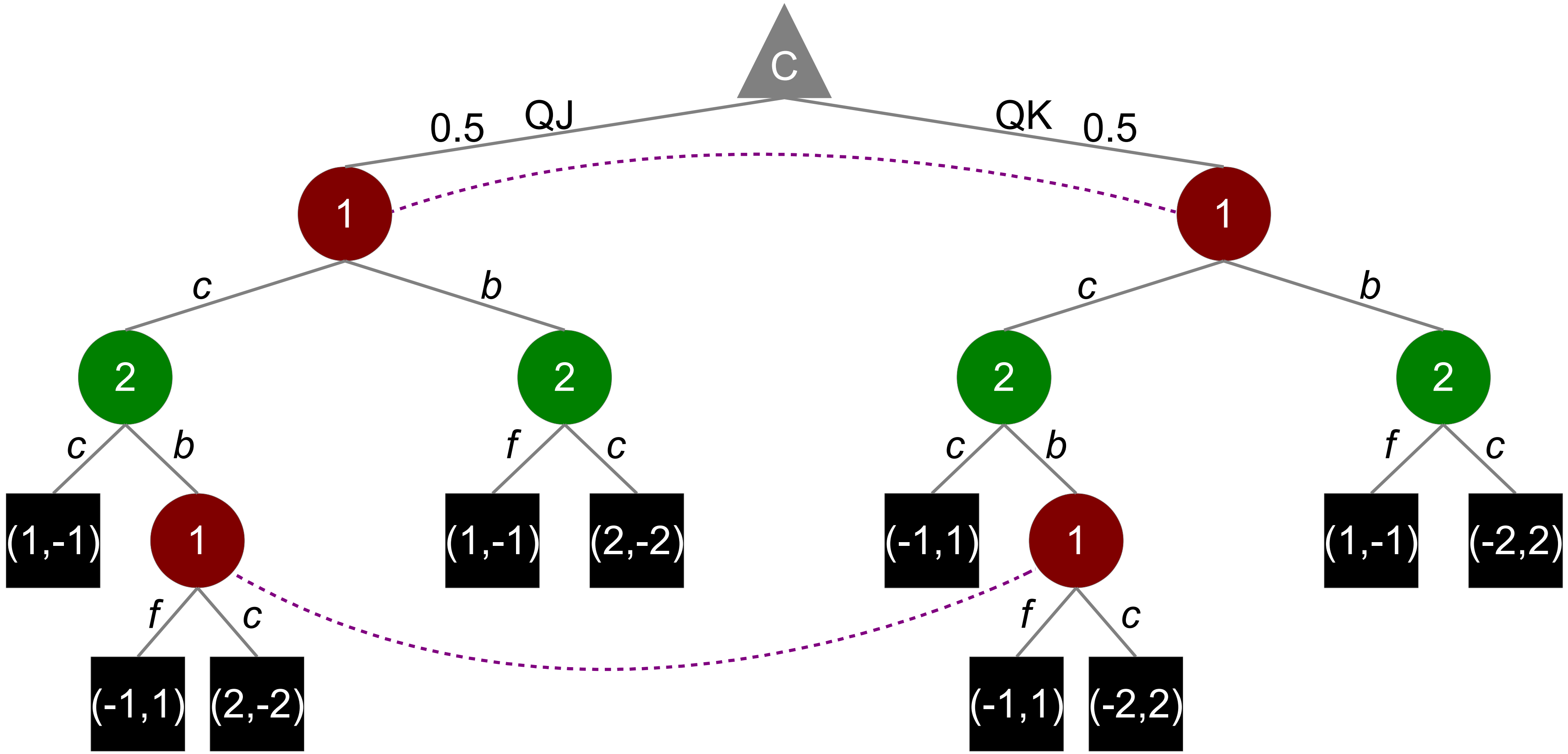
Counterfactual Regret Minimization (CFR)

- In games with more than 2-players, σ^{AVG} is a "good" strategy. **Why?**

- What properties make a strategy good in games with more than 2-players?

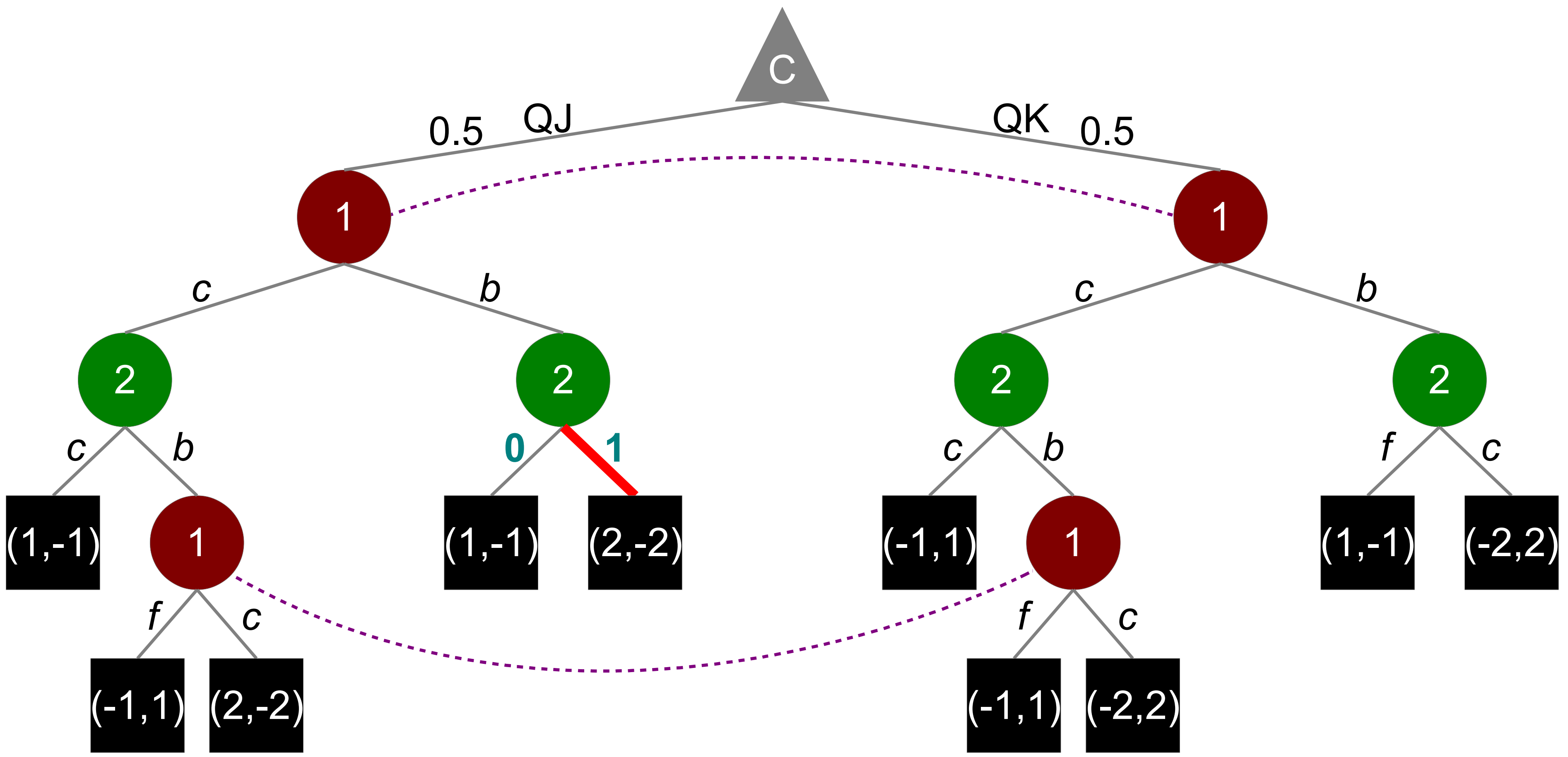
- We know what a bad strategy is...

Domination



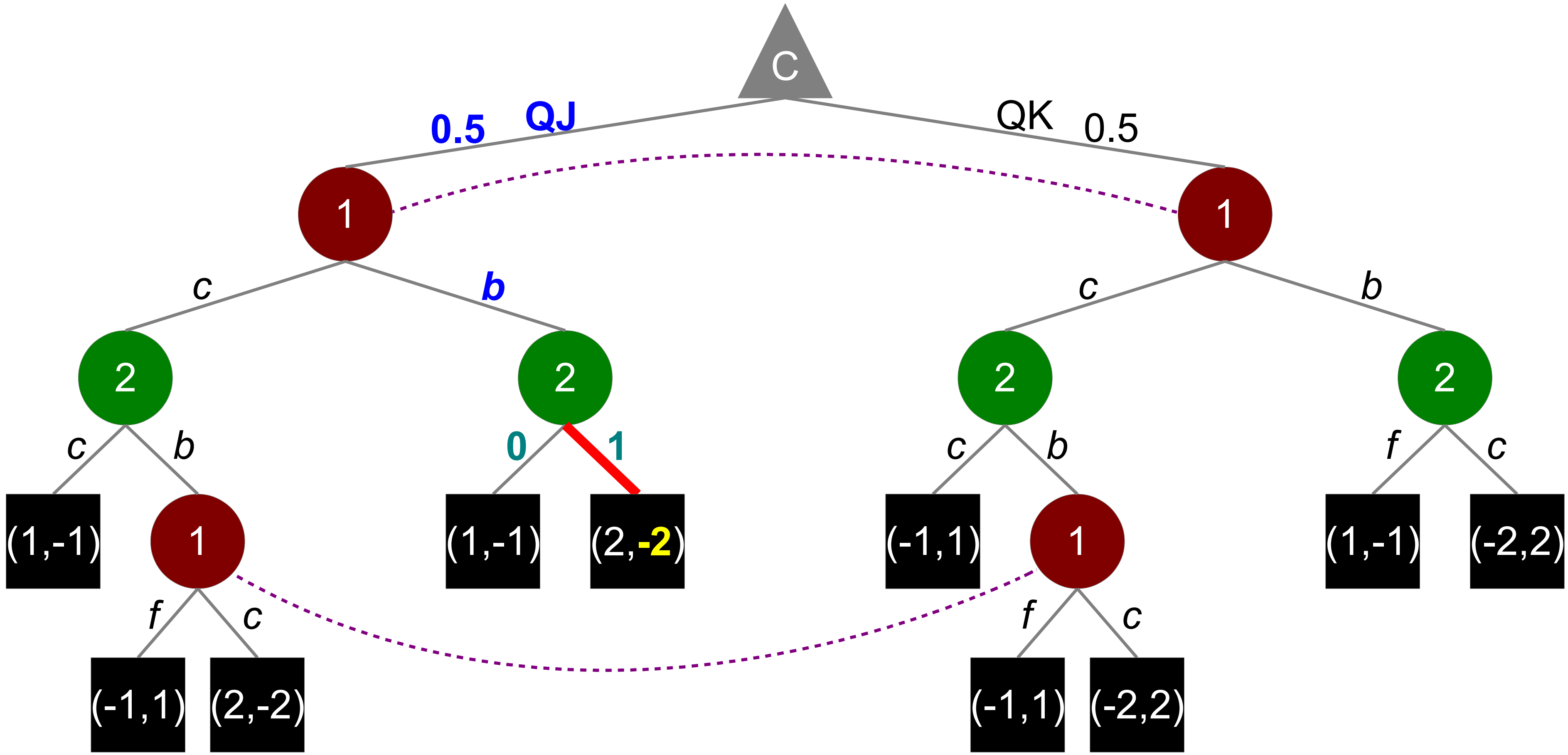
C
QJ
0.5
QK
0.5
1
1
c
b
c
b
2
2
2
2
c
b
f
c
c
b
f
c
(1,-1)
1
(1,-1)
(2,-2)
(-1,1)
1
(1,-1)
(-2,2)
f
c
f
c
(-1,1)
(2,-2)
(-1,1)
(-2,2)

Domination



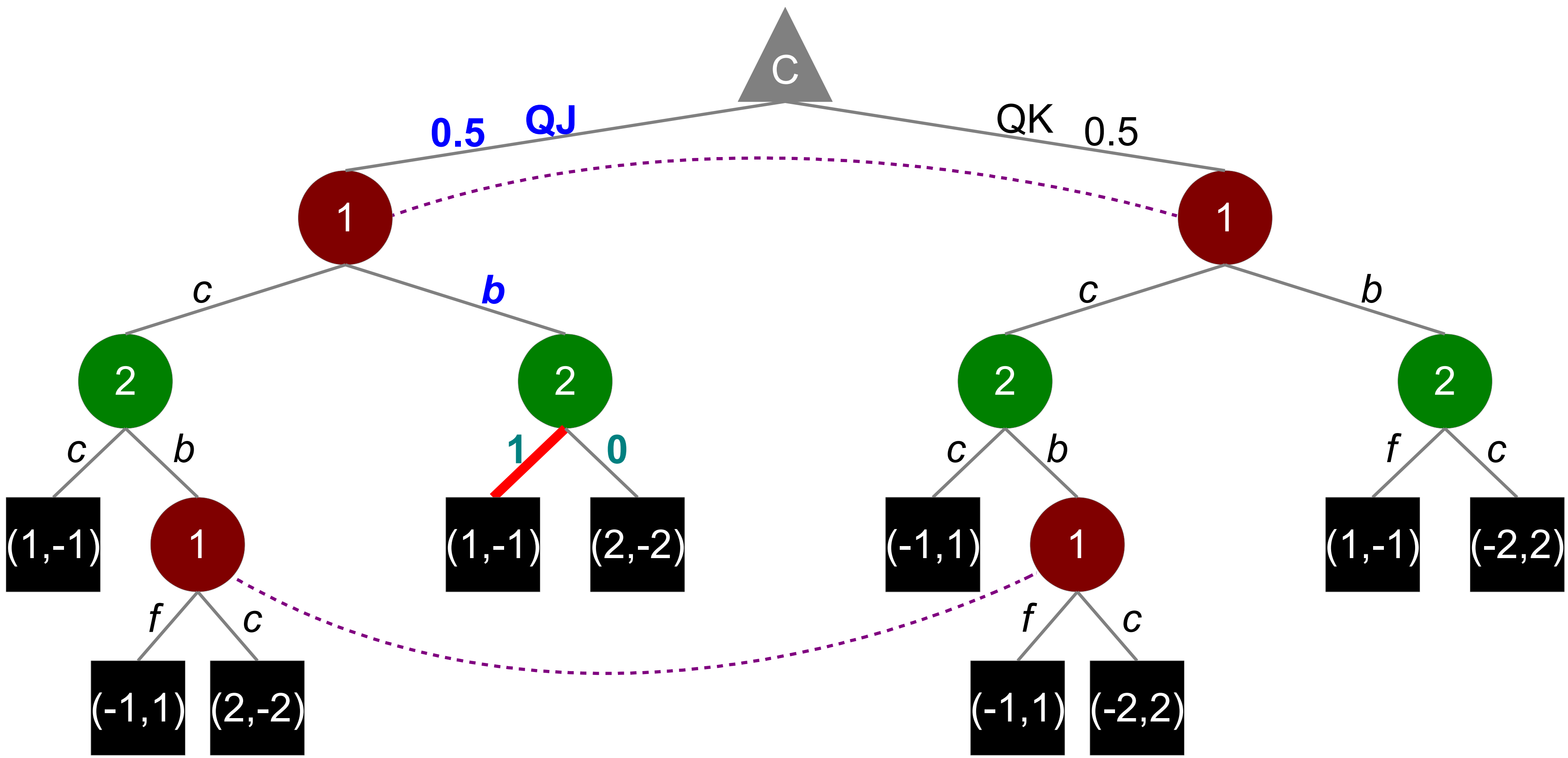
Consider any player 2 strategy $\sigma_2^{J,c}$ that always calls with the Jack when faced with a bet.

Domination



$$u_2(\sigma_1, \sigma_2^{J,c}) = \ldots + 0.5\sigma_1(Q?, b)1(-2) + \ldots$$

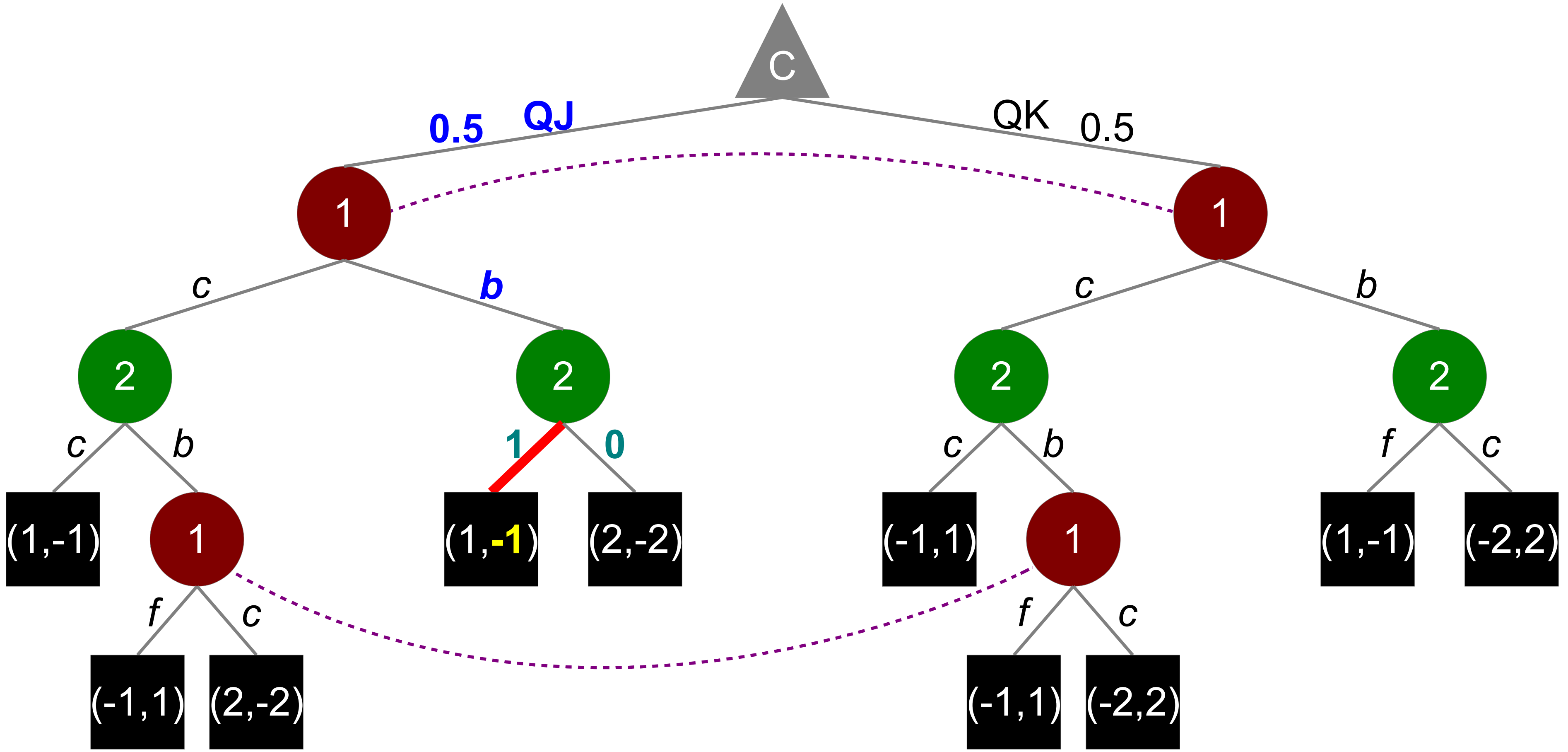
Domination



$$\boldsymbol{u_2}(\ \sigma_1,\ \sigma_2^{J,c}\ ) = \ldots + \mathbf{0.5}\sigma_1(\ \mathbf{Q?},\ \boldsymbol{b}\ )\mathbf{1}(\ -2\ ) + \ldots$$

Now consider the same player 2 strategy, except always folds the J. Call it $\sigma_2^{J,f}$.

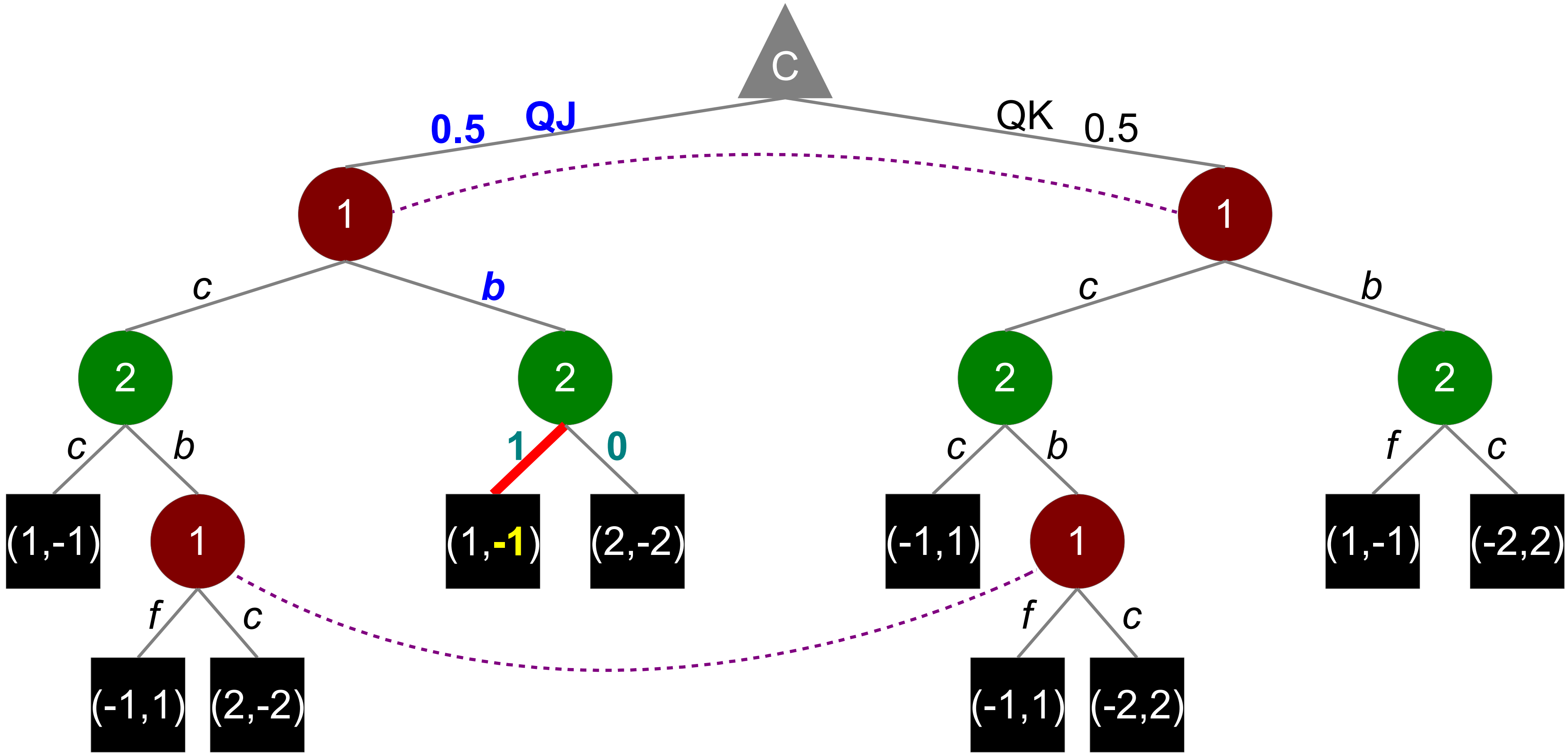
Domination



$$u_2(\ \sigma_1,\ \sigma_2^{J,c}\ ) = \ldots + 0.5\sigma_1(\ Q?,\ b\ )1(\ -2\ ) + \ldots$$

$$u_2(\ \sigma_1,\ \sigma_2^{J,f}\ ) = \text{''} + 0.5\sigma_1(\ Q?,\ b\ )1(\ \mathbf{-1}\ ) + \text{''}$$

Domination



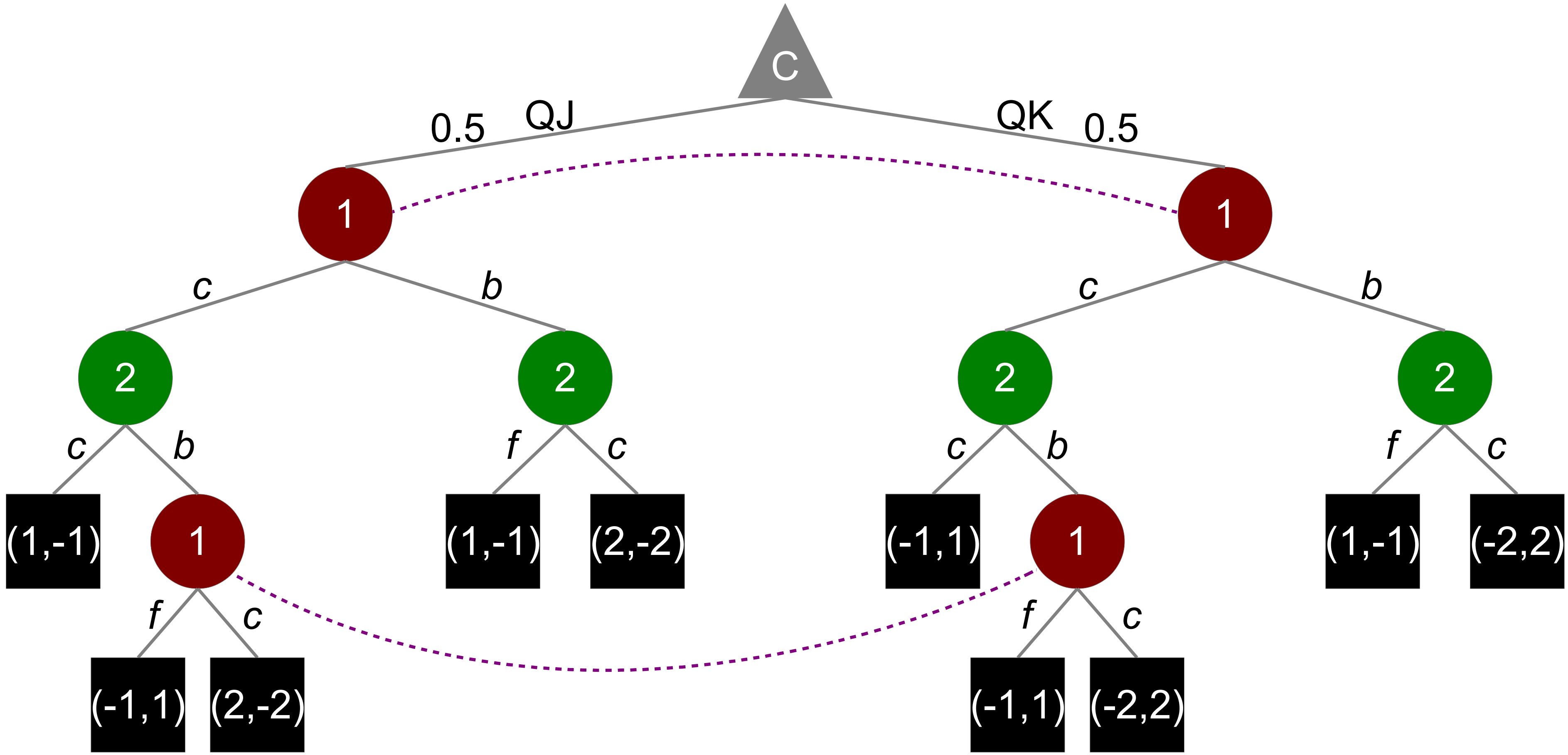
$$u_2(\,\sigma_1,\ \sigma_2^{J,c}\,) = \ldots + 0.5\sigma_1(\,Q?,\ b\,)1(\,-2\,) + \ldots$$

$$u_2(\,\sigma_1,\ \sigma_2^{J,f}\,) = \text{ '' } + 0.5\sigma_1(\,Q?,\ b\,)1(\,\mathbf{-1}\,) + \text{ '' }$$

$$u_2(\,\sigma_1,\ \sigma_2^{J,c}\,) \le u_2(\,\sigma_1,\ \sigma_2^{J,f}\,) \text{ for all } \sigma_1.$$

$$u_2(\,\sigma_1,\ \sigma_2^{J,c}\,) < u_2(\,\sigma_1,\ \sigma_2^{J,f}\,) \text{ if } \sigma_1(\,Q?,\ b\,) > 0$$

Domination

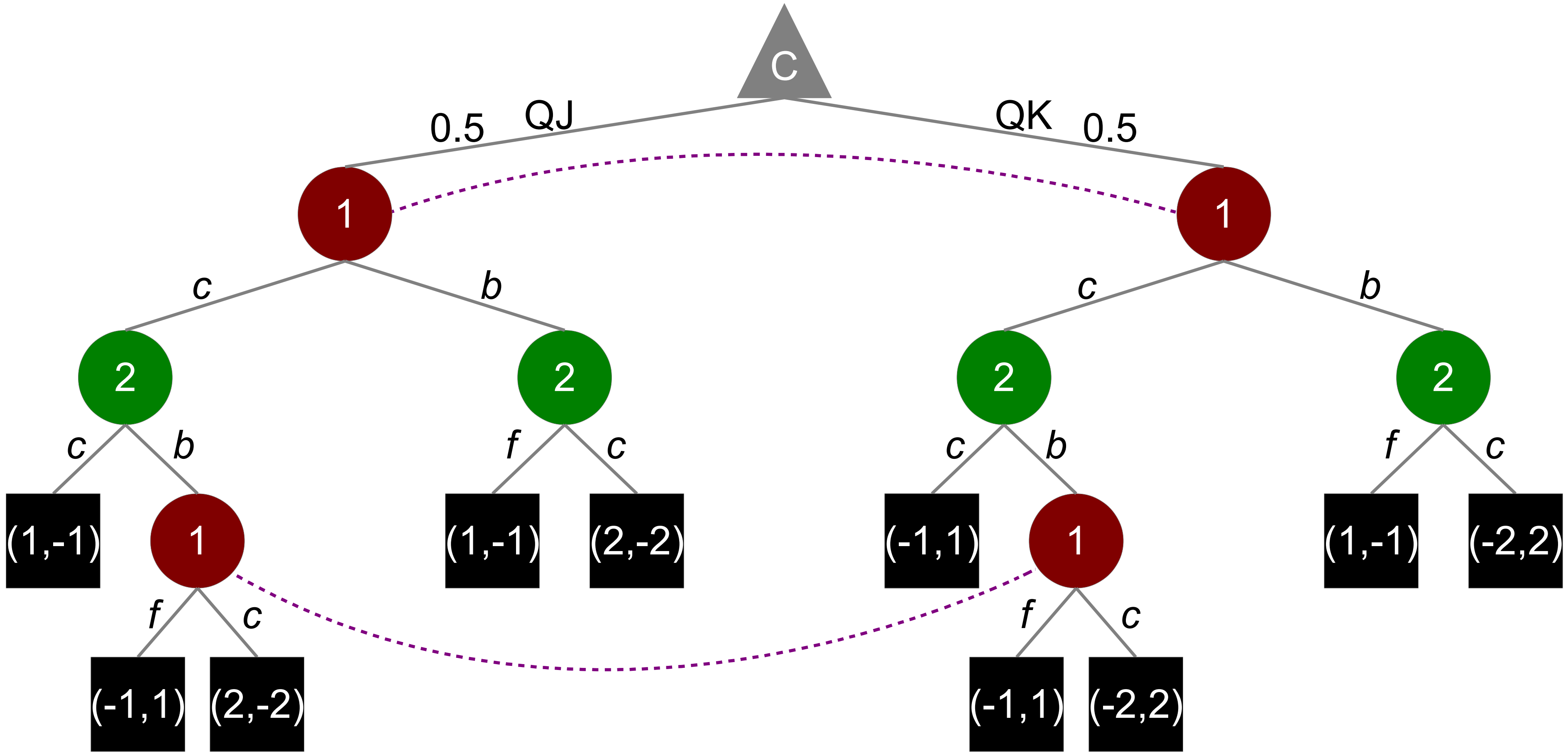


σ_2 is a **dominated strategy** if there exists σ_2' such that

$u_2(\sigma_1, \sigma_2, \sigma_3, \ldots) \leq u_2(\sigma_1, \sigma_2', \sigma_3, \ldots)$ for all $\sigma_1, \sigma_3, \ldots$

$u_2(\sigma_1, \sigma_2, \sigma_3, \ldots) < u_2(\sigma_1, \sigma_2', \sigma_3, \ldots)$ for some $\sigma_1, \sigma_3, \ldots$

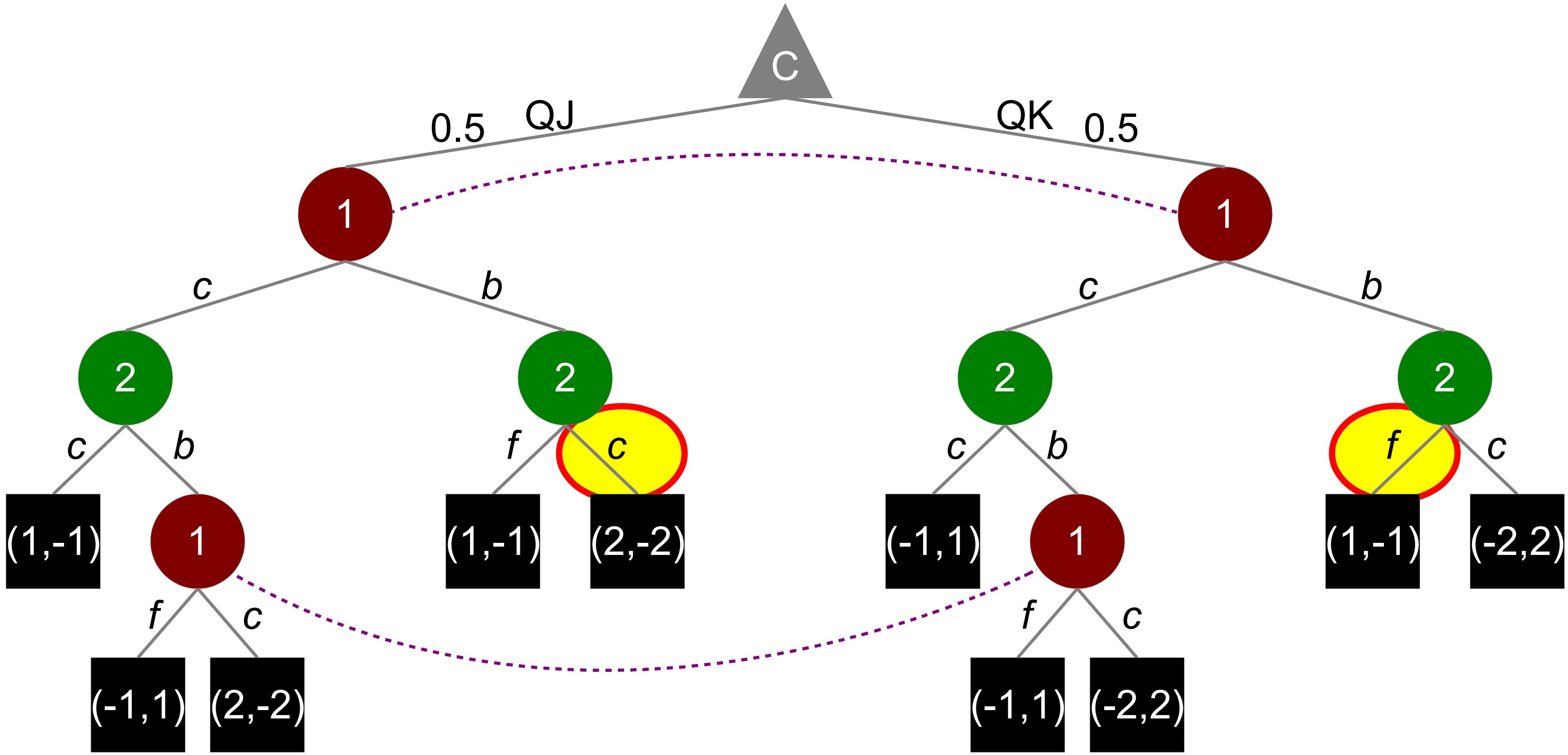
Domination



$\sigma_2^{J,c}$ is **dominated** by $\sigma_2^{J,f}$

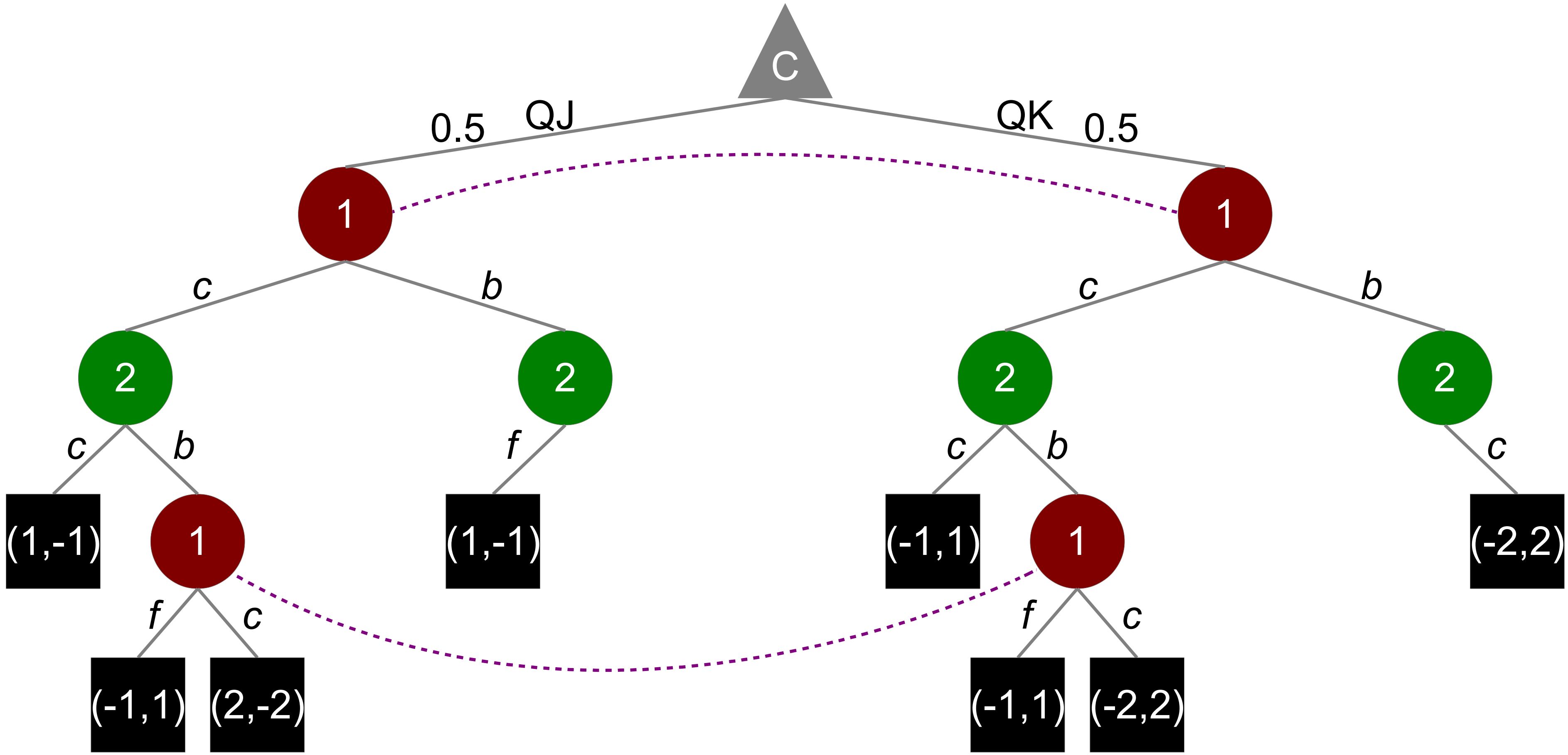
$\sigma_2^{K,f}$ is **dominated** by $\sigma_2^{K,c}$

Domination

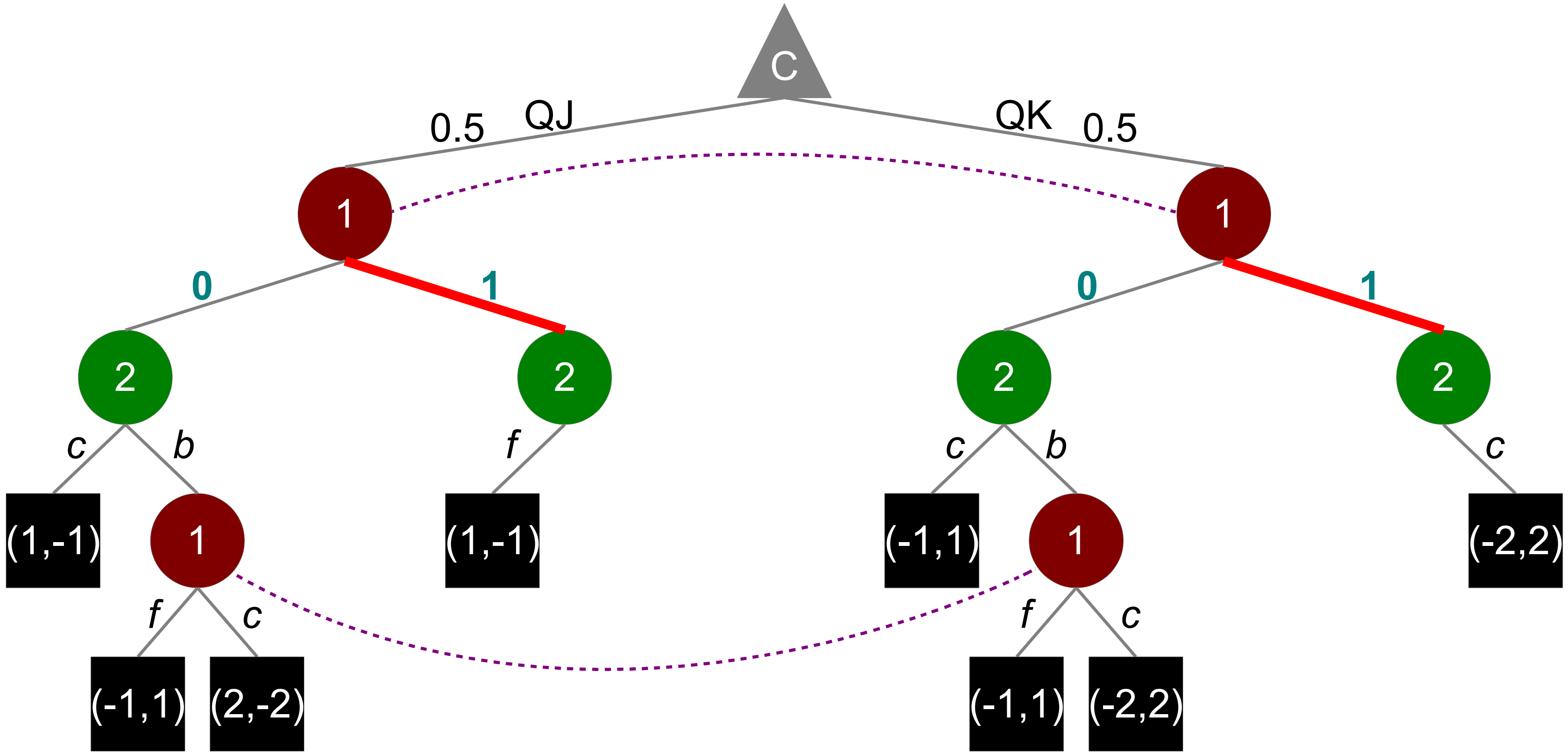


Define a **dominated action** to be an action such that any strategy that always plays that action is dominated (assuming that player plays to reach that action).

Domination

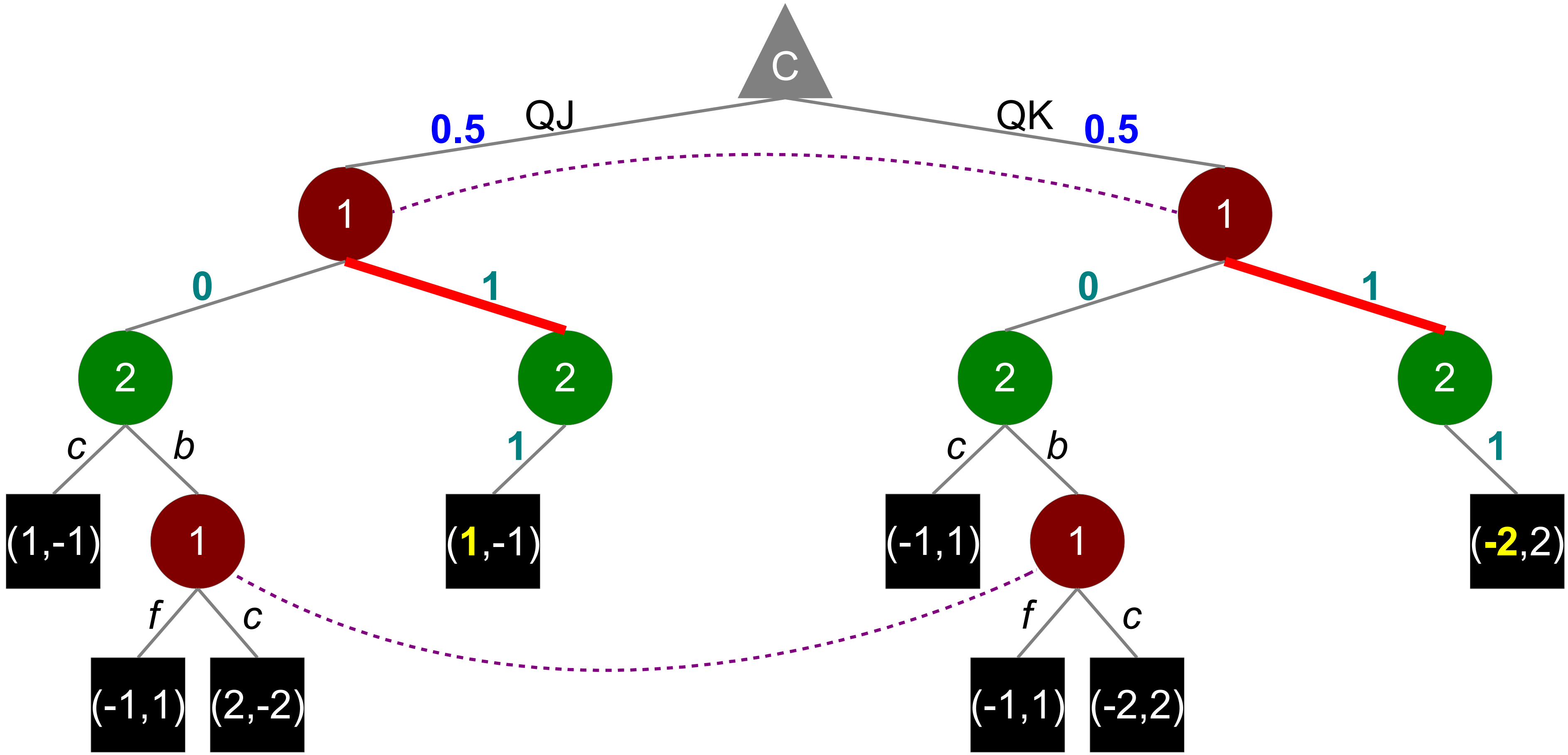


Domination



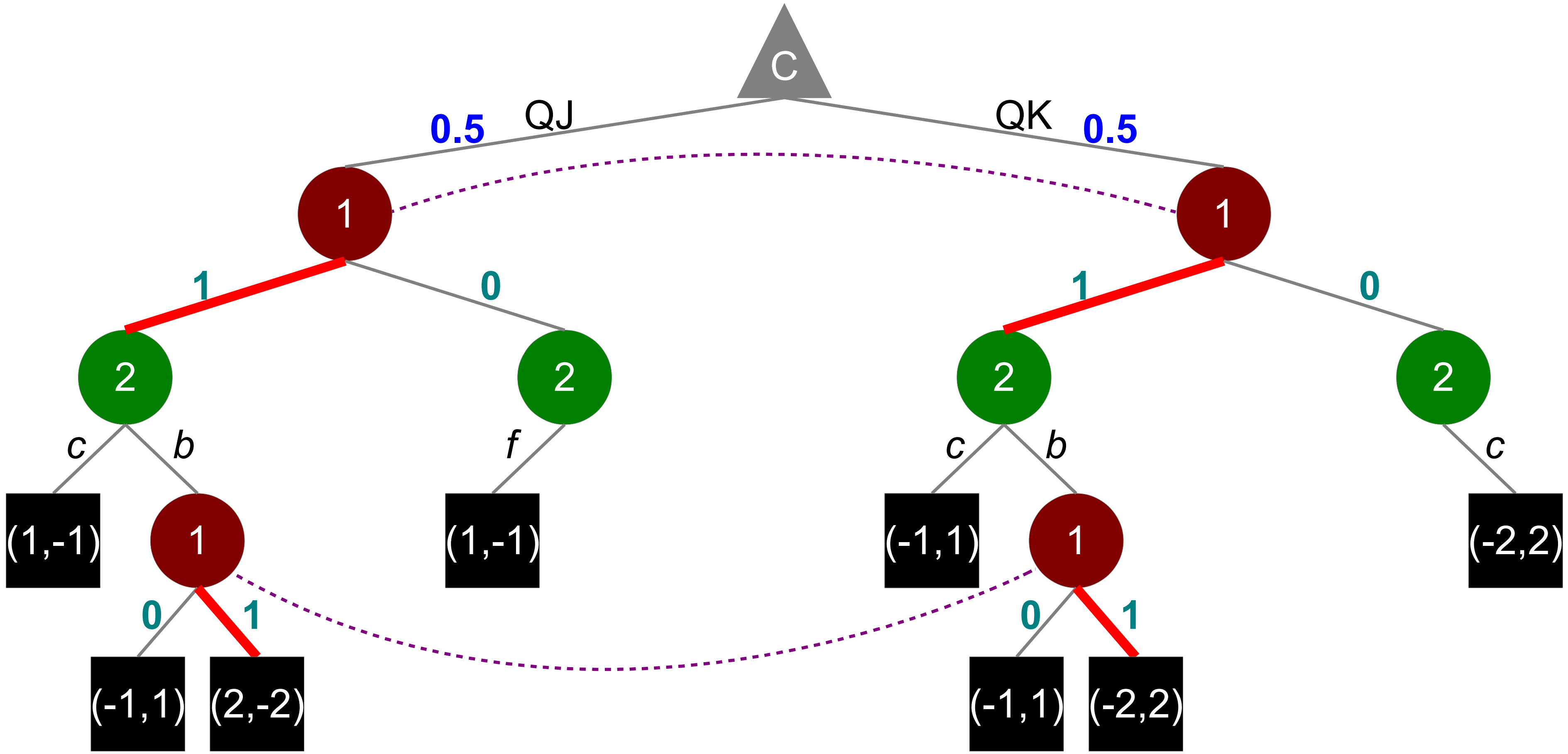
Consider the player 1 strategy σ_1^b that always bets.

Domination



$$u_1(\sigma_1^b, \sigma_2) = 0.5(1)(1)(1) + 0.5(1)(1)(-2) = -0.5$$

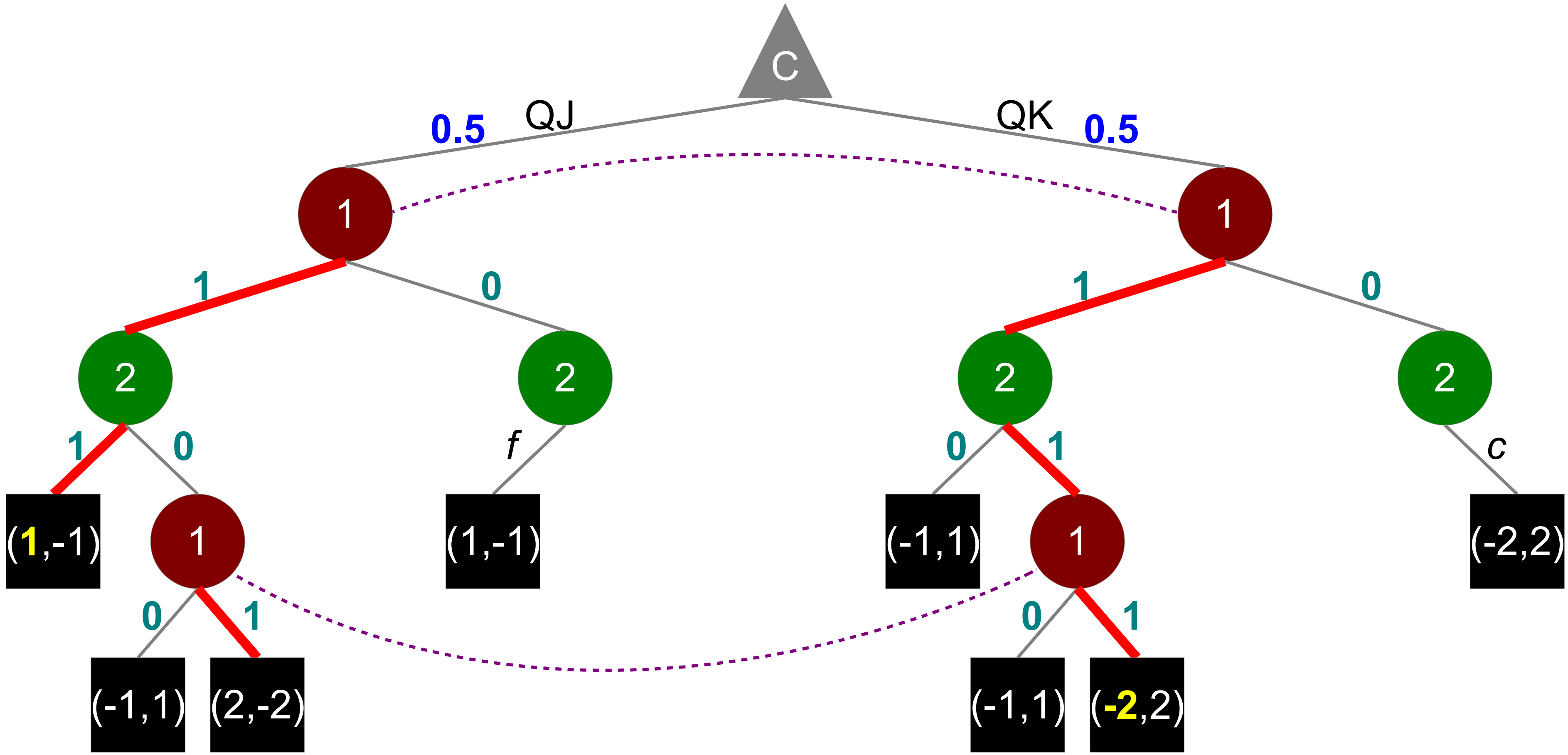
Domination



$u_1(\ \sigma_1^{\ b},\ \sigma_2\ ) = 0.5(\ 1\ )(\ 1\ )(\ 1\ ) + 0.5(\ 1\ )(\ 1\ )(\ -2\ ) = -0.5$

Now consider the player 1 strategy $\sigma_1^{\ cc}$ that checks then calls.

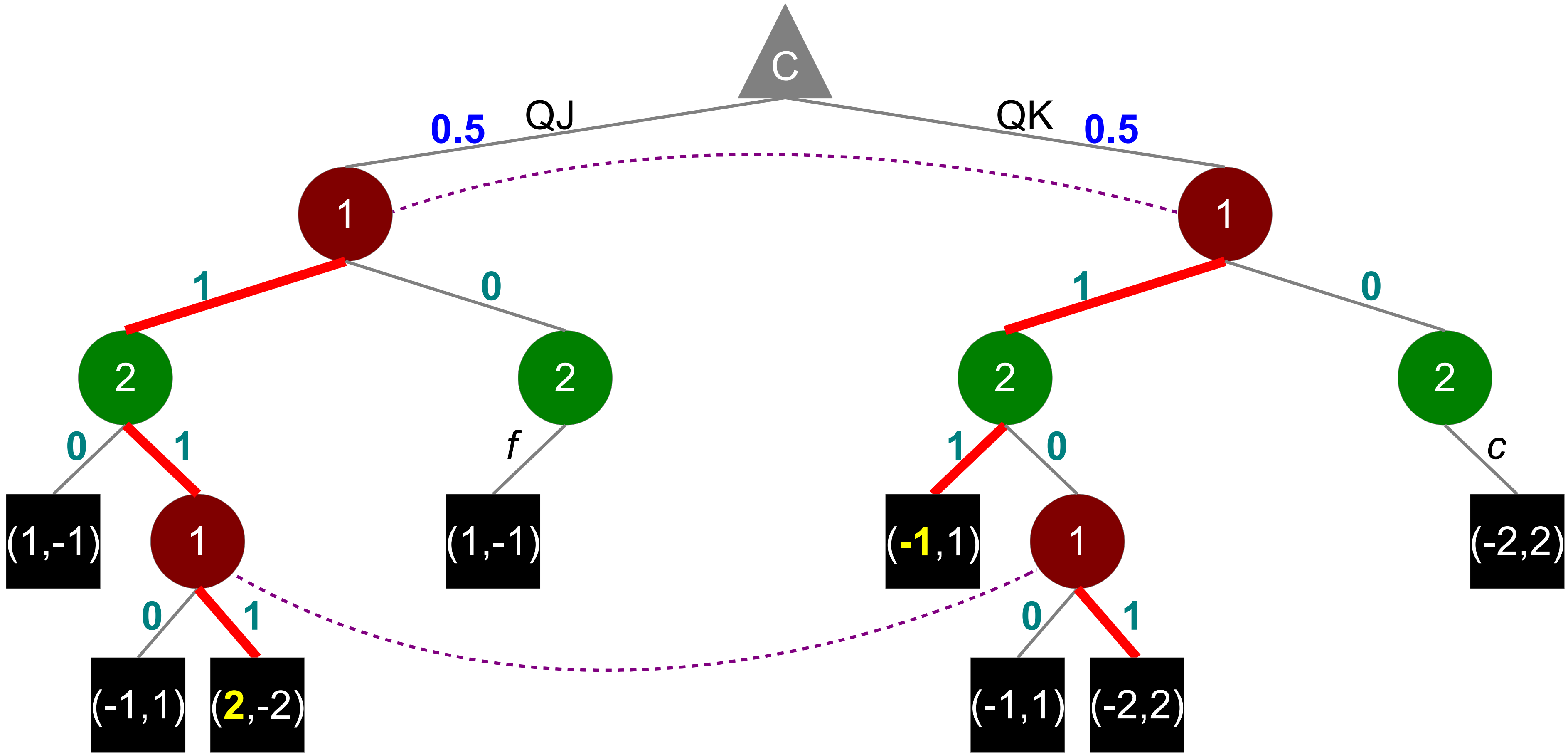
Domination



$$u_1(\ \sigma_1^{\ b},\ \sigma_2\ ) = 0.5(\ 1\ )(\ 1\ )(\ 1\ ) + 0.5(\ 1\ )(\ 1\ )(\ -2\ ) = -0.5$$

$$u_1(\ \sigma_1^{\ cc},\ \sigma_2^{\ Jc,Kb}\ ) = 0.5(\ 1\ )(\ 1\ )(\ 1\ ) + 0.5(\ 1\ )(\ 1\ )(\ 1\ )(\ -2\ ) = -0.5$$

Domination

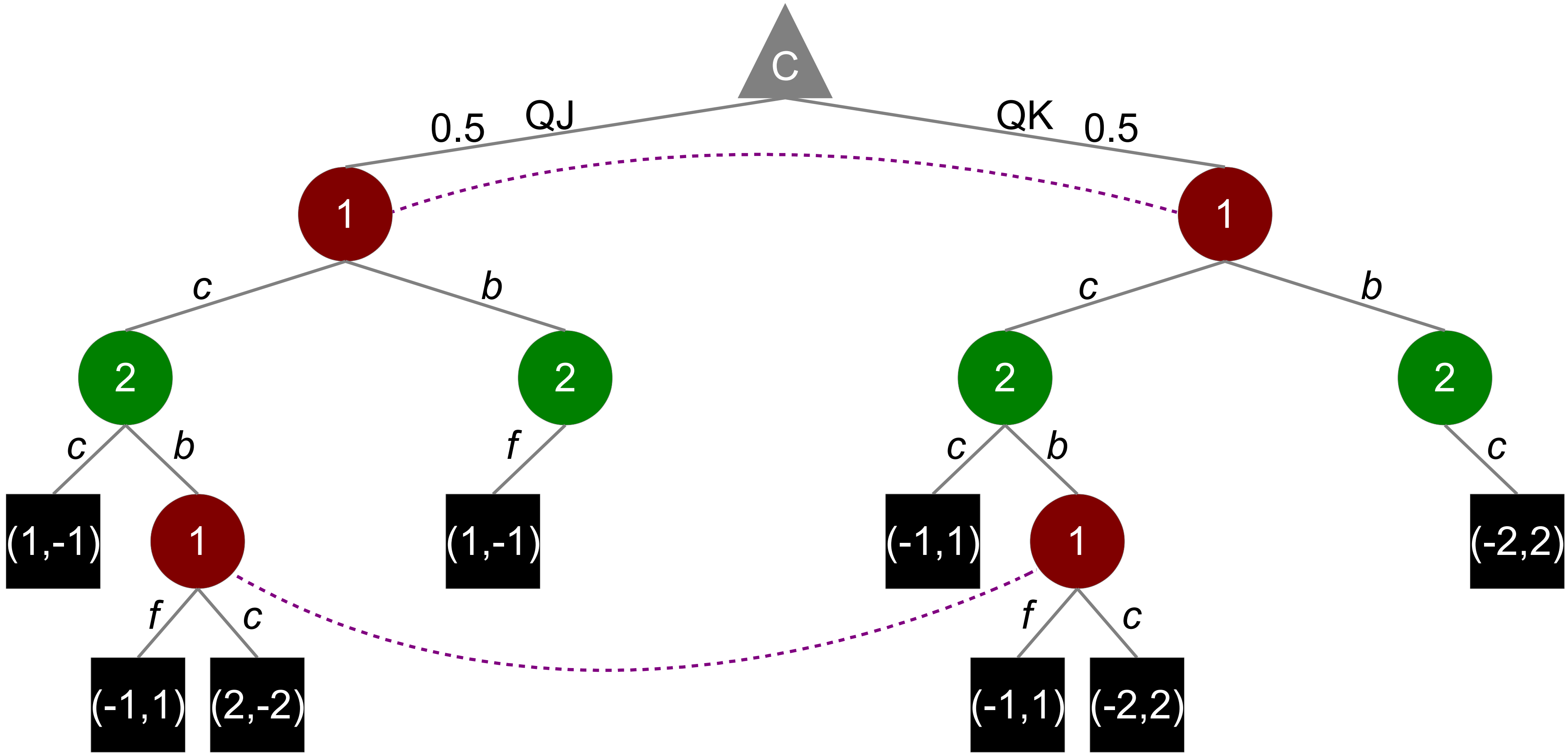


$$u_1(\sigma_1^{b}, \sigma_2) = 0.5(1)(1)(1) + 0.5(1)(1)(-2) = -0.5$$

$$u_1(\sigma_1^{cc}, \sigma_2^{Jc,Kb}) = 0.5(1)(1)(1) + 0.5(1)(1)(1)(-2) = -0.5$$

$$u_1(\sigma_1^{cc}, \sigma_2^{Jb,Kc}) = 0.5(1)(1)(1)(2) + 0.5(1)(1)(-1) = +0.5$$

Domination

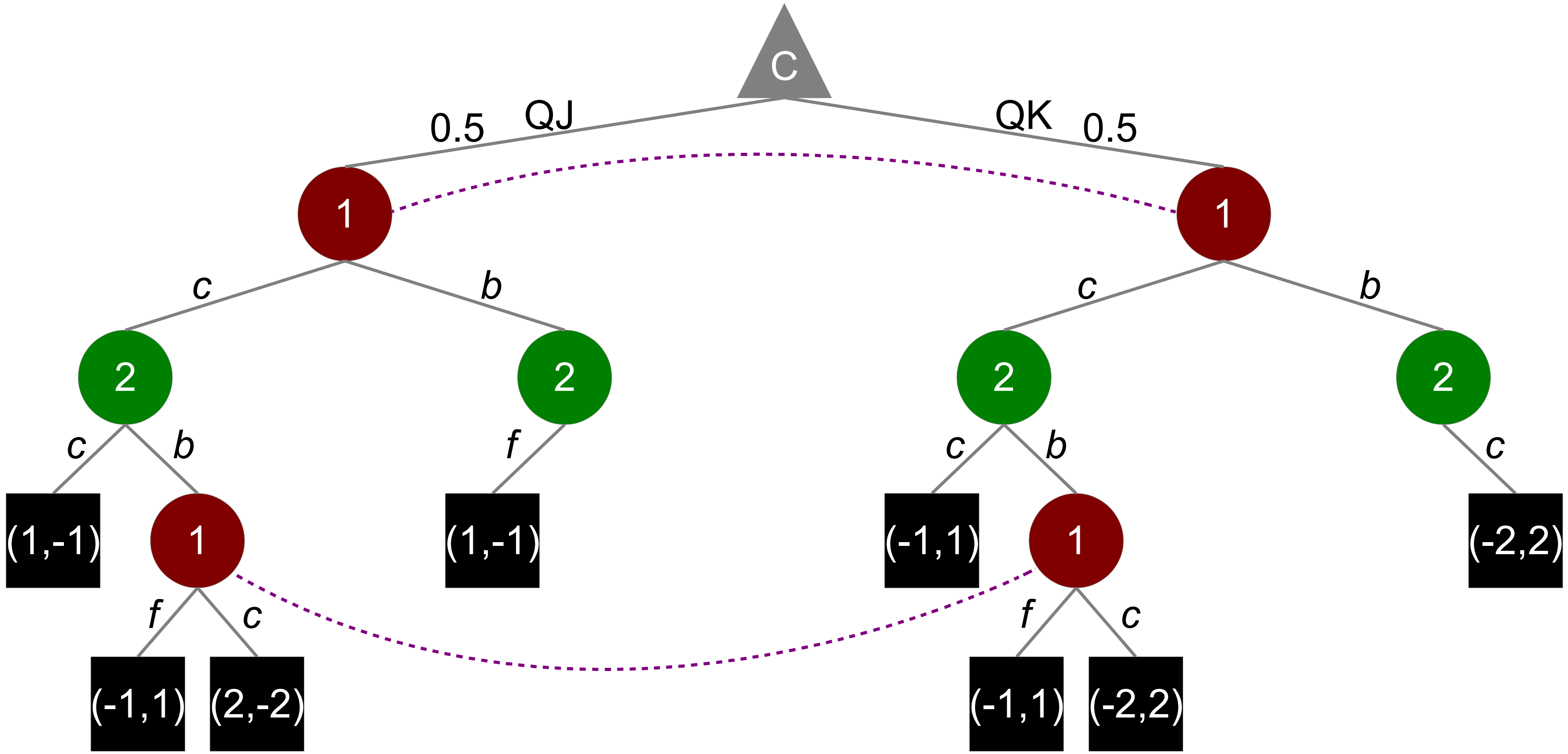


σ_1 is an **iteratively dominated strategy** if there exists σ_1' such that

$u_1(\sigma_1, \sigma_2, \sigma_3, \ldots) \le u_1(\sigma_1', \sigma_2, \sigma_3, \ldots)$ for all non-iteratively dominated $\sigma_2, \sigma_3, \ldots$

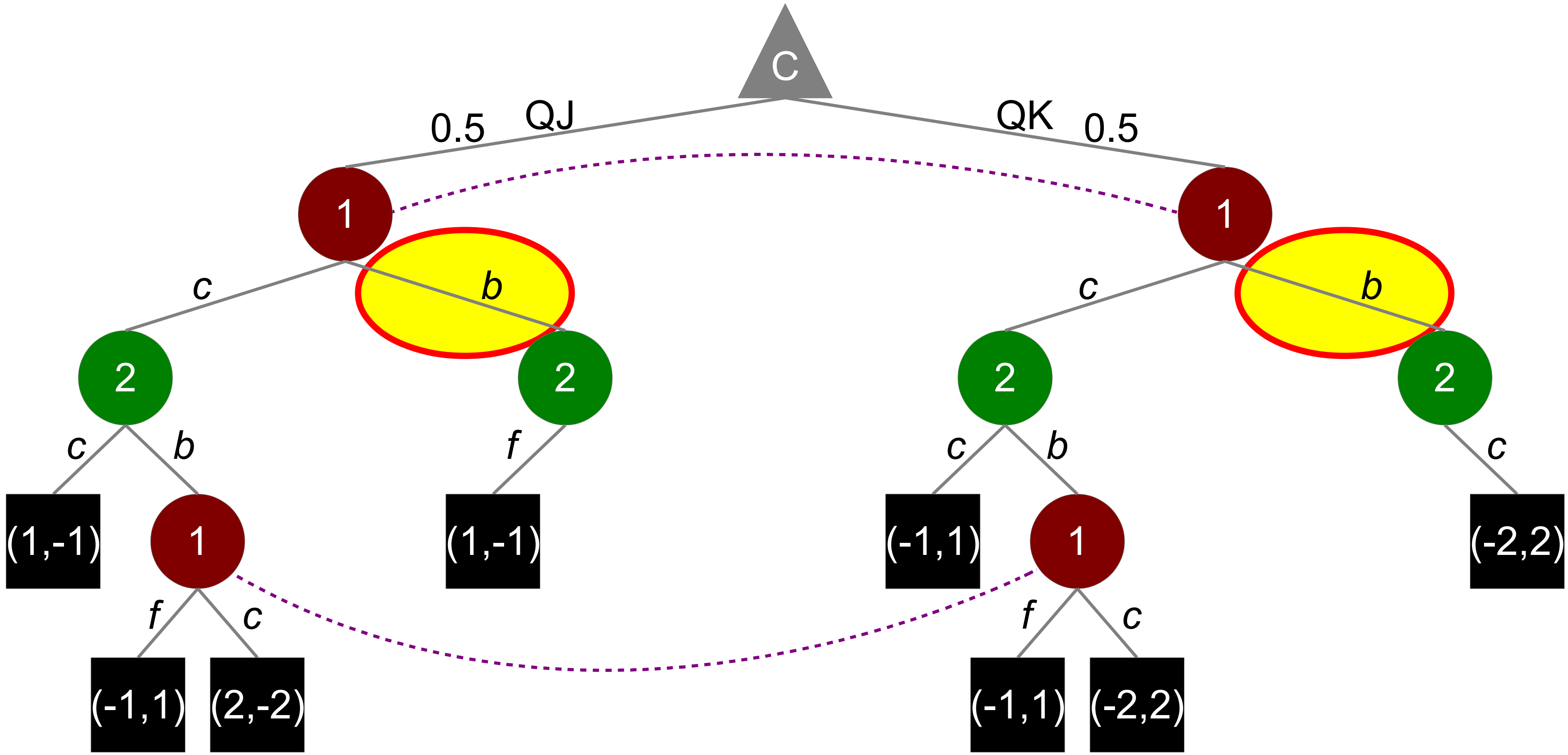
$u_1(\sigma_1, \sigma_2, \sigma_3, \ldots) < u_1(\sigma_1', \sigma_2, \sigma_3, \ldots)$ for some non-iteratively dominated $\sigma_2, \sigma_3, \ldots$

Domination



σ_1^{b} is **iteratively dominated** by σ_1^{cc}

Domination



Define an **iteratively dominated action** to be an action such that any strategy that always plays that action is iteratively dominated (assuming that player plays to reach that action).

Domination and CFR

- Clearly, one should not play a **dominated action**.
- If we assume our opponents are rational, then we should also not play an **iteratively dominated action**.
- **Theorem**: If *a* is an iteratively **strictly** dominated action, and the players play to reach *a* "often enough," then when running CFR,

$$\sigma^{\mathrm{AVG}}(a) \to 0 \text{ as } T \to \infty.$$

- Can also prove a weaker result regarding CFR avoiding strictly dominated strategies.

Discussion

- We can show that CFR avoids dominated actions and strategies, but how important is it to avoid such actions and strategies?
  - Need to measure correlation between playing dominated actions or strategies and performance.
  - Hard to identify all dominated actions in large games, but may be computationally possible in smaller games.

Discussion

- Recall that CFR generates a sequence of strategy profiles, ($\boldsymbol{\sigma}^1$, $\boldsymbol{\sigma}^2$, ... , $\boldsymbol{\sigma}^T$) over many iterations T.
- Can show that for an iteratively strictly dominated action a, after a finite number of iterations T_0, the profiles generated play a with probability 0.
  - If avoiding iteratively dominated actions is enough to perform well, then perhaps there is no need to use the average profile $\boldsymbol{\sigma}^{\mathbf{AVG}}$ as is needed in 2-player zero-sum games.

Conclusion

- CFR can generate strong strategies outside of 2-player zero-sum games, but we do not have a good understanding of why this is so.
- Iteratively dominated actions and strategies should typically be avoided in any game.
- We have shown that the strategies produced by CFR tend to avoid playing iteratively strictly dominated actions.
  - More work is required to conclude that this really does help explain the strong performance of CFR-generated strategies.

Thanks for listening!

Richard Gibson

Twitter: @RichardGGibson

Email: rggibson@cs.ualberta.ca

Website: http://cs.ualberta.ca/~rggibson

CPRG Website: http://cs.ualberta.ca/~poker