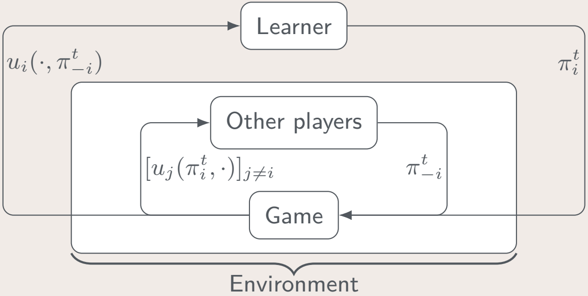
Efficient Deviation Types and Learning in Extensive-Form Games

Dustin Morrill, Ryan D'Orazio, Marc Lanctot,
James R. Wright, Michael Bowling, Amy R. Greenwald

July 16, 2021 – Berkeley MARL Group

Introduction

- We seek more effective algorithms for playing multi-player, general-sum extensive-form games (EFGs).



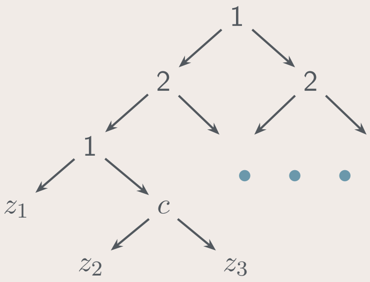
Learner : $\pi_i^1 \quad \pi_i^2 \quad \cdots \quad \pi_i^T \quad \to \frac{1}{T}\sum_{t=1}^{T} u_i(\pi_i^t, \pi_{-i}^t)$

Deviation : $\phi(\pi_i^1) \quad \phi(\pi_i^2) \quad \cdots \quad \phi(\pi_i^T) \quad \to \frac{1}{T}\sum_{t=1}^{T} u_i\left(\phi(\pi_i^t), \pi_{-i}^t\right)$

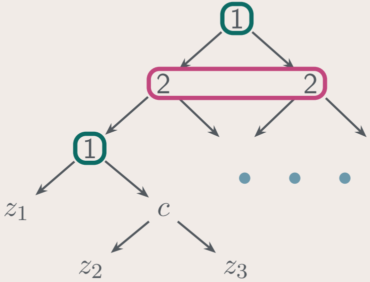
Objective:

$$\underbrace{\frac{1}{T}\sum_{t=1}^{T} u_i(\pi_i^t, \pi_{-i}^t)}_{\text{The learner's average reward.}} \geq \max_{\phi \in \Phi} \underbrace{\frac{1}{T}\sum_{t=1}^{T} u_i(\phi(\pi_i^t), \pi_{-i}^t)}_{\text{Deviation } \phi\text{'s average reward.}} - \underbrace{o(1)}_{\text{Leeway.}} .$$

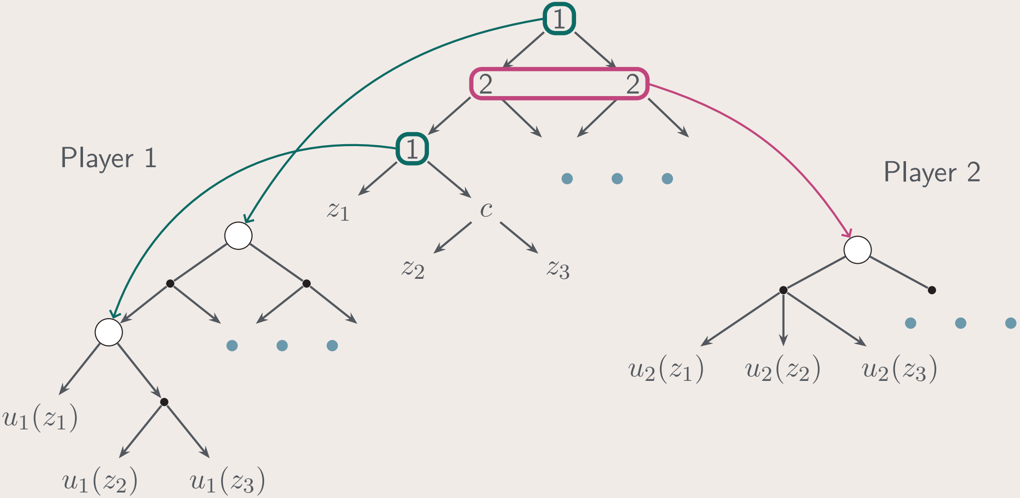
Extensive-Form Game Trees



Extensive-Form Game Trees

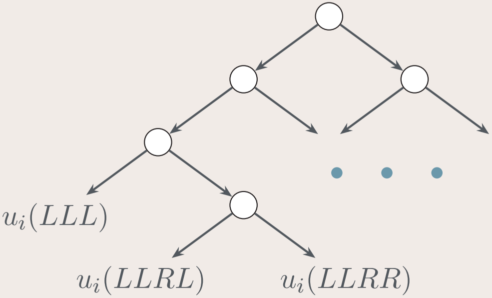


Information Set Trees



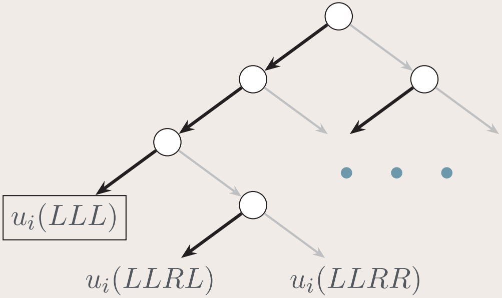
Strategies

Player i's information set tree:



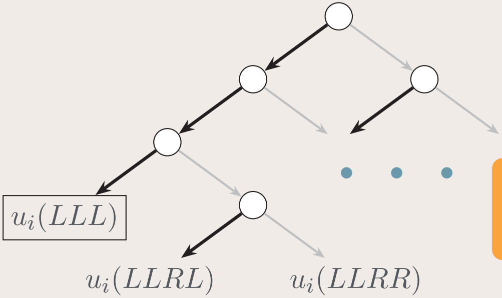
Strategies

The "always left" strategy:



Strategies

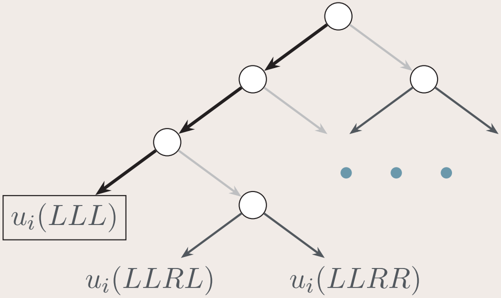
The "always left" strategy:



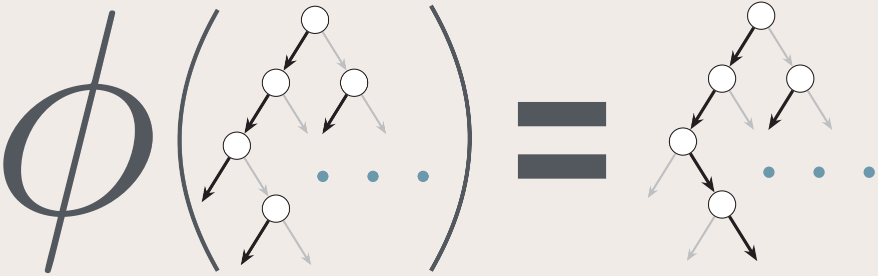
5 information sets with 2 actions
$\implies 2^5 = 32$ strategies.

Reduced Strategies

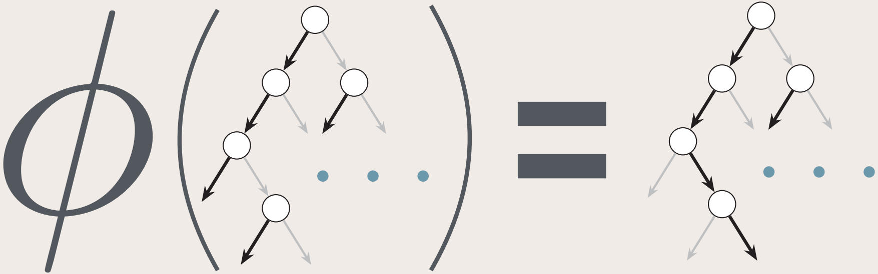
The reduced "always left" strategy:



Deviations

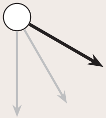


Deviations



At least 32 ways to transform any given strategy and 32^{32} possible deviation functions!

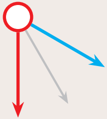
von Stengel and Forges's Deviations [1]

$$\phi(\pi_i) =$$ 

input action, $\pi_i(\cdot)$	deviation action, $[\phi\pi_i](\cdot)$
3	?

[1] von Stengel and Forges, "Extensive-form correlated equilibrium: Definition and computational complexity".

von Stengel and Forges's Deviations [1]

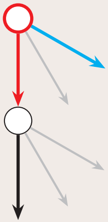
$$\phi(\pi_i) =$$ 

input action, $\pi_i(\cdot)$	**deviation action,** $[\phi\pi_i](\cdot)$
3	1

[1] von Stengel and Forges, "Extensive-form correlated equilibrium: Definition and computational complexity".

von Stengel and Forges's Deviations [1]

$$\phi(\pi_i) =$$

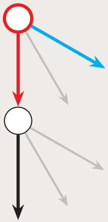


memory	input action, $\pi_i(\cdot)$	deviation action, $[\phi\pi_i](\cdot)$
$\varnothing$	3	1
3	1	?

[1] von Stengel and Forges, "Extensive-form correlated equilibrium: Definition and computational complexity".

von Stengel and Forges's Deviations [1]

$$\phi(\pi_i) =$$

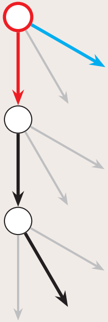


memory	input action, $\pi_i(\cdot)$	deviation action, $[\phi\pi_i](\cdot)$
$\varnothing$	3	1
3	1	1

[1] von Stengel and Forges, "Extensive-form correlated equilibrium: Definition and computational complexity".

von Stengel and Forges's Deviations [1]

$$\phi(\pi_i) =$$

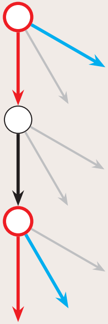


memory	input action, $\pi_i(\cdot)$	deviation action, $[\phi\pi_i](\cdot)$
$\varnothing$	3	1
3	1	1
3 1	2	?

[1] von Stengel and Forges, "Extensive-form correlated equilibrium: Definition and computational complexity".

von Stengel and Forges's Deviations [1]

$$\phi(\pi_i) =$$

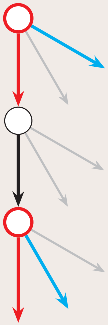


memory	input action, $\pi_i(\cdot)$	deviation action, $[\phi\pi_i](\cdot)$
$\varnothing$	3	1
3	1	1
3 1	2	1

[1] von Stengel and Forges, "Extensive-form correlated equilibrium: Definition and computational complexity".

von Stengel and Forges's Deviations [1]

$$\phi(\pi_i) =$$



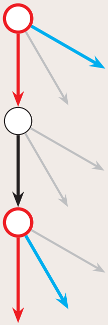
memory	input action, $\pi_i(\cdot)$	deviation action, $[\phi\pi_i](\cdot)$
$\varnothing$	3	1
3	1	1
3 1	2	1

Deviation player behavior can only depend on their observations
$\Longrightarrow$ many deviation functions are ruled out.

[1] von Stengel and Forges, "Extensive-form correlated equilibrium: Definition and computational complexity".

von Stengel and Forges's Deviations [1]

$$\phi(\pi_i) =$$

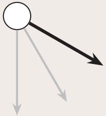


memory	input action, $\pi_i(\cdot)$	deviation action, $[\phi\pi_i](\cdot)$
$\varnothing$	3	1
3	1	1
3 1	2	1

But, the memory string grows linearly with depth
$\implies$ the # of memory states grows exponentially
$\implies$ the # of deviations is exponential in depth.

[1] von Stengel and Forges, "Extensive-form correlated equilibrium: Definition and computational complexity".

von Stengel and Forges's Deviations [2] With a Reduced Strategy

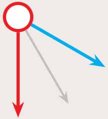
$$\phi(\bar{\pi}_i) =$$ 

memory	input action, $\bar{\pi}_i(\cdot)$	deviation action, $[\phi\bar{\pi}_i](\cdot)$
$\varnothing$	3	?

[2] von Stengel and Forges, "Extensive-form correlated equilibrium: Definition and computational complexity".

von Stengel and Forges's Deviations [2] With a Reduced Strategy

$$\phi(\bar{\pi}_i) =$$

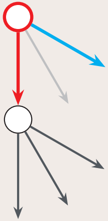


memory	input action, $\bar{\pi}_i(\cdot)$	deviation action, $[\phi\bar{\pi}_i](\cdot)$
$\varnothing$	3	1

[2] von Stengel and Forges, "Extensive-form correlated equilibrium: Definition and computational complexity".

von Stengel and Forges's Deviations [2] With a Reduced Strategy

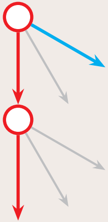
$$\phi(\bar{\pi}_i) =$$



memory	input action, $\bar{\pi}_i(\cdot)$	deviation action, $[\phi\bar{\pi}_i](\cdot)$
$\varnothing$	3	1
3	*	?

[2] von Stengel and Forges, "Extensive-form correlated equilibrium: Definition and computational complexity".

von Stengel and Forges's Deviations [2] With a Reduced Strategy

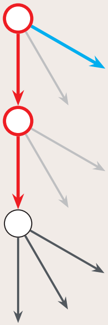
$\phi(\bar{\pi}_i) =$ 

memory	input action, $\bar{\pi}_i(\cdot)$	deviation action, $[\phi\bar{\pi}_i](\cdot)$
$\varnothing$	3	1
3	*	1

[2] von Stengel and Forges, "Extensive-form correlated equilibrium: Definition and computational complexity".

von Stengel and Forges's Deviations [2] With a Reduced Strategy

$\phi(\bar{\pi}_i) =$

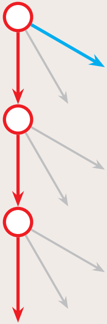


memory	input action, $\bar{\pi}_i(\cdot)$	deviation action, $[\phi\bar{\pi}_i](\cdot)$
$\varnothing$	3	1
3	*	1
3 *	*	?

[2] von Stengel and Forges, "Extensive-form correlated equilibrium: Definition and computational complexity".

von Stengel and Forges's Deviations [2] With a Reduced Strategy

$$\phi(\bar{\pi}_i) =$$

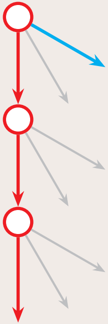


memory	input action, $\bar{\pi}_i(\cdot)$	deviation action, $[\phi\bar{\pi}_i](\cdot)$
∅	3	1
3	*	1
3 *	*	1

[2] von Stengel and Forges, "Extensive-form correlated equilibrium: Definition and computational complexity".

von Stengel and Forges's Deviations [2] With a Reduced Strategy

$$\phi(\bar{\pi}_i) =$$



memory	input action, $\bar{\pi}_i(\cdot)$	deviation action, $[\phi\bar{\pi}_i](\cdot)$
∅	3	1
3	*	1
3 *	*	1

Now the # of possible memory states (and thus deviations) grows linearly.

[2] von Stengel and Forges, "Extensive-form correlated equilibrium: Definition and computational complexity".

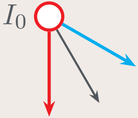
Behavioral Deviations: A Flexible Formalization of von Stengel and Forges's Deviations

Behavioral Deviations

$$\phi = \{\phi_{I,g} : \mathcal{A}(I) \to \mathcal{A}(I)\}_{\substack{\text{information set } I, \\ \text{memory state } g}}$$

Behavioral Deviations

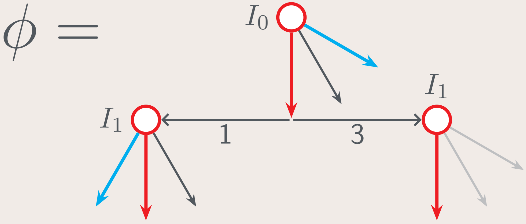
$\phi =$



memory state: transformation

$\varnothing$: $\phi^{3\to 1}$

Behavioral Deviations



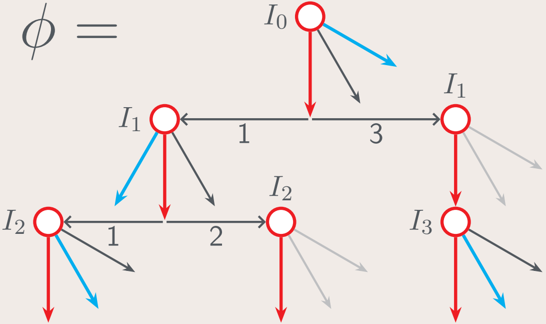
memory state: transformation

$\varnothing$: $\phi^{3\to 1}$

$1 : \phi^{1\to 2}, \quad 3 : \phi^{\to 1}$

Arriving in $I_1 \implies [\phi^{3\to 1}\pi_i](I_0) = 1$
$\implies \pi_i(I_0) = 1$ or $\pi_i(I_0) = 3$.

Behavioral Deviations



memory state: transformation

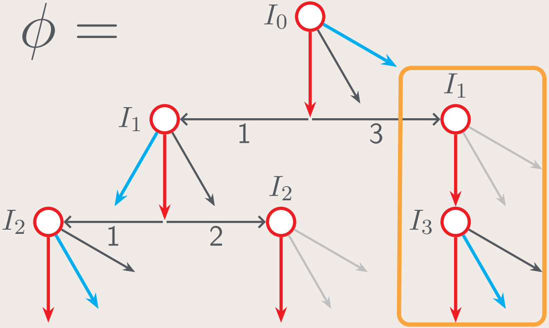
$\varnothing$: $\phi^{3\to 1}$

$1 : \phi^{1\to 2}, \quad 3 : \phi^{\to 1}$

$1\ 1 : \phi^{2\to 1}, \quad 1\ 2 : \phi^{\to 1}, \quad 3 * : \phi^{2\to 1}$

Arriving in I_3 neither requires nor reveals $\pi_i(I_1)$ since $\phi^{\to 1}$ is external/constant.

$\phi =$



memory state: transformation

$\varnothing$: $\phi^{3\to 1}$

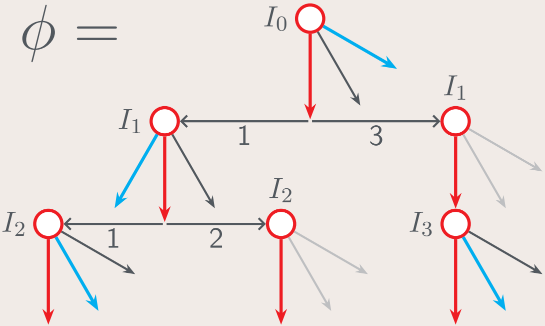
$1 : \phi^{1\to 2}, \quad 3 : \phi^{\to 1}$

$1\ 1 : \phi^{2\to 1}, \quad 1\ 2 : \phi^{\to 1}, \quad 3 * : \phi^{2\to 1}$

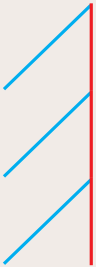
The action at I_1 in memory state "3" can be hidden from the deviation player, but the action at I_3 can be revealed.

Behavioral Deviations

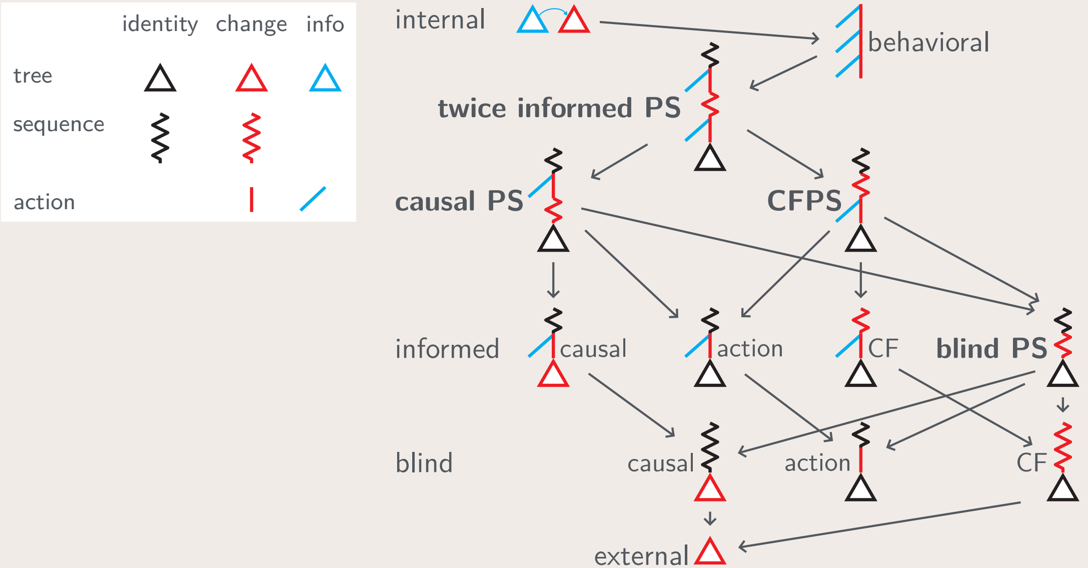
$\phi =$



$\in \Phi^{\mathrm{IN}}_{\mathcal{I}_i} =$

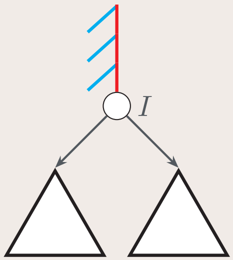


The EFG Deviation Landscape



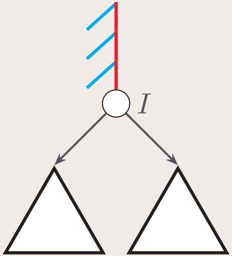
Extensive-Form Regret Minimization (EFR)

Extensive-Form Regret Minimization (EFR)



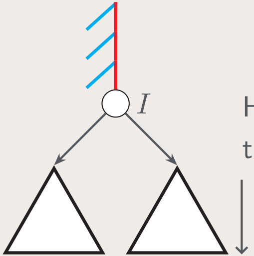
EFR works by learning $\pi_i^t(I) \in \Delta^{|\mathcal{A}(I)|}$.

Extensive-Form Regret Minimization (EFR)



How much incentive does the deviation player have to employ transformations at I?

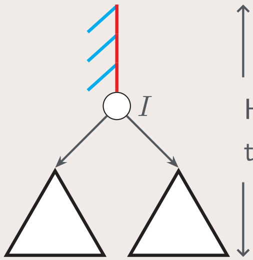
Extensive-Form Regret Minimization (EFR)



How much incentive does the deviation player have to employ transformations at I?

What is the value of each action a under i's current strategy, π_i^t?

Extensive-Form Regret Minimization (EFR)

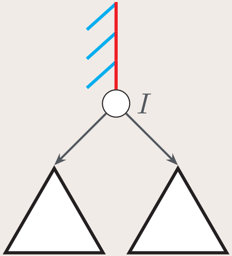


How often does each $\phi \in \Phi^{\text{IN}}_{\mathcal{I}_i}$ reach I and with what memory state?

How much incentive does the deviation player have to employ transformations at I?

What is the value of each action a under i's current strategy, π^t_i?

Extensive-Form Regret Minimization (EFR)



How often does each $\phi \in \Phi^{\mathrm{IN}}_{\mathcal{I}_i}$ reach I and with what memory state?

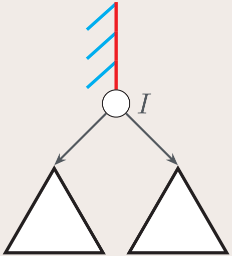
How much incentive does the deviation player have to employ transformations at I?

$\forall a, \underbrace{v_I(a; \pi^t)}$

Counterfactual value,

i.e., expected payoff for a assuming i plays to I.

Extensive-Form Regret Minimization (EFR)



Memory probability,

i.e., the chance that $\phi(\pi_i^t)$ reaches I in memory state g.

$$\forall \phi \in \Phi_{\mathcal{I}_i}^{\text{IN}}, g, \overbrace{w_\phi(I, g; \pi_i^t)} \in [0, 1]$$

How much incentive does the deviation player have to employ transformations at I?

$$\forall a, \underbrace{v_I(a; \pi^t)}$$

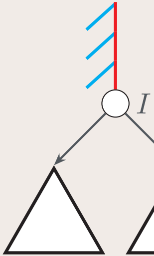
Counterfactual value,

i.e., expected payoff for a assuming i plays to I.

Extensive-Form Regret Minimization (EFR)

Memory probability,

i.e., the chance that $\phi(\pi_i^t)$ reaches I in memory state g.



$$\forall \phi \in \Phi_{\mathcal{I}_i}^{\mathrm{IN}}, g, \; \overbrace{w_\phi(I, g; \pi_i^t)} \in [0, 1]$$

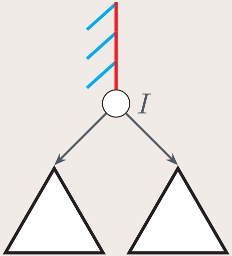
$$w_\phi(I, g; \pi_i^t) \underbrace{\left(v_I([\phi_{I,g}\pi_i^t](I); \pi^t) - v_I(\pi_i^t(I); \pi^t)\right)}_{\text{Counterfactual regret, } \rho_I^{\mathrm{CF}}(\phi_{I,g}; \pi^t).}$$

$$\forall a, \; \underbrace{v_I(a; \pi^t)}$$

Counterfactual value,

i.e., expected payoff for a assuming i plays to I.

Extensive-Form Regret Minimization (EFR)



Memory probability,

i.e., the chance that $\phi(\pi_i^t)$ reaches I in memory state g.

$$\downarrow \; \forall \phi \in \Phi^{\mathrm{IN}}_{\mathcal{I}_i}, g, \; \overbrace{w_\phi(I, g; \pi_i^t)} \in [0, 1]$$

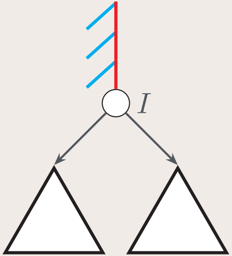
$$w_\phi(I, g; \pi_i^t)\rho_I^{\mathrm{CF}}(\phi_{I,g}; \pi^t)$$

$$\uparrow \; \forall a, \; \underbrace{v_I(a; \pi^t)}$$

Counterfactual value,

i.e., expected payoff for a assuming i plays to I.

Extensive-Form Regret Minimization (EFR)



Memory probability,

i.e., the chance that $\phi(\pi_i^t)$ reaches I in memory state g.

$$\forall \phi \in \Phi_{\mathcal{I}_i}^{\mathrm{IN}}, g, \overbrace{w_\phi(I, g; \pi_i^t)} \in [0, 1]$$

$$w_\phi(I, g; \pi_i^t)\rho_I^{\mathrm{CF}}(\phi_{I,g}; \pi^t)$$

This is a time selection problem! [4]

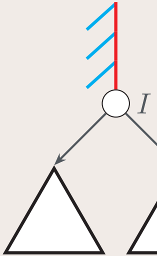
$$\forall a, \underbrace{v_I(a; \pi^t)}$$

Counterfactual value,

i.e., expected payoff for a assuming i plays to I.

[4] Blum and Mansour, "From external to internal regret".

Extensive-Form Regret Minimization (EFR)



Memory probability,

i.e., the chance that $\phi(\pi_i^t)$ reaches I in memory state g.

$$\forall \phi \in \Phi^{\mathrm{IN}}_{\mathcal{I}_i}, g, \overbrace{w_\phi(I, g; \pi_i^t)} \in [0, 1]$$

$$w_\phi(I, g; \pi_i^t)\rho_I^{\mathrm{CF}}(\phi_{I,g}; \pi^t)$$

Our solution:
time selection regret matching.

$$\forall a, \underbrace{v_I(a; \pi^t)}$$

Counterfactual value,

i.e., expected payoff for a assuming i plays to I.

Time Selection Regret Matching

- Sets $\pi_i^t(I) \leftarrow$ to be a fixed point of a linear operator, L^t:

Time Selection Regret Matching

- Sets $\pi_i^t(I) \leftarrow$ to be a fixed point of a linear operator, L^t:

$$x^t_{\phi_{I,g}} \leftarrow \sum_{k}^{t-1} w_\phi(I, g; \pi_i^k)\rho_I^{\text{CF}}(\phi_{I,g}; \pi^k) \qquad \triangleright \text{ Cumulative immediate regret.}$$

Time Selection Regret Matching

- Sets $\pi_i^t(I) \leftarrow$ to be a fixed point of a linear operator, L^t:

$$x^t_{\phi_{I,g}} \leftarrow \sum_{k}^{t-1} w_\phi(I, g; \pi_i^k) \rho_I^{\mathrm{CF}}(\phi_{I,g}; \pi^k) \qquad \triangleright \text{ Cumulative immediate regret.}$$

$$y^t_{\phi_I} \leftarrow \sum_{g} w_\phi(I, g; \pi_i^t) f\left(x^t_{\phi_{I,g}}\right) \qquad \triangleright \text{ Link output/preference for } \phi_I.$$

Time Selection Regret Matching

- Sets $\pi_i^t(I) \leftarrow$ to be a fixed point of a linear operator, L^t:

$$x^t_{\phi_{I,g}} \leftarrow \sum_k^{t-1} w_\phi(I, g; \pi_i^k) \rho_I^{\mathrm{CF}}(\phi_{I,g}; \pi^k) \qquad \triangleright \text{ Cumulative immediate regret.}$$

$$y^t_{\phi_I} \leftarrow \sum_g w_\phi(I, g; \pi_i^t) f\left(x^t_{\phi_{I,g}}\right) \qquad \triangleright \text{ Link output/preference for } \phi_I.$$

$$L^t : \Delta^{|\mathcal{A}(I)|} \ni \sigma \mapsto \frac{1}{z^t} \sum_{\phi_I} y^t_{\phi_I} \phi_I(\sigma). \qquad \triangleright \text{ Convex combination.}$$

Time Selection Regret Matching

- Sets $\pi_i^t(I) \leftarrow$ to be a fixed point of a linear operator, L^t:

$$x^t_{\phi_{I,g}} \leftarrow \sum_k^{t-1} w_\phi(I, g; \pi_i^k)\rho_I^{\mathrm{CF}}(\phi_{I,g}; \pi^k) \qquad \triangleright \text{ Cumulative immediate regret.}$$

$$y^t_{\phi_I} \leftarrow \sum_g w_\phi(I, g; \pi_i^t) f\left(x^t_{\phi_{I,g}}\right) \qquad \triangleright \text{ Link output/preference for } \phi_I.$$

$$L^t : \Delta^{|\mathcal{A}(I)|} \ni \sigma \mapsto \frac{1}{z^t} \sum_{\phi_I} y^t_{\phi_I} \phi_I(\sigma). \qquad \triangleright \text{ Convex combination.}$$

- Permits:

Time Selection Regret Matching

- Sets $\pi_i^t(I) \leftarrow$ to be a fixed point of a linear operator, L^t:

$$x^t_{\phi_{I,g}} \leftarrow \sum_k^{t-1} w_\phi(I, g; \pi_i^k) \rho_I^{\mathrm{CF}}(\phi_{I,g}; \pi^k) \qquad \triangleright \text{ Cumulative immediate regret.}$$

$$y^t_{\phi_I} \leftarrow \sum_g w_\phi(I, g; \pi_i^t) f\left(x^t_{\phi_{I,g}}\right) \qquad \triangleright \text{ Link output/preference for } \phi_I.$$

$$L^t : \Delta^{|\mathcal{A}(I)|} \ni \sigma \mapsto \frac{1}{z^t} \sum_{\phi_I} y^t_{\phi_I} \phi_I(\sigma). \qquad \triangleright \text{ Convex combination.}$$

- Permits:
  - choice of link function f, *e.g.*, $f(\cdot) = \max\{0, \cdot\}$ or $f(\cdot) = \mathrm{e}^{\eta \cdot}$.

Time Selection Regret Matching

- Sets $\pi_i^t(I) \leftarrow$ to be a fixed point of a linear operator, L^t:

$$x^t_{\phi_{I,g}} \leftarrow \sum_k^{t-1} w_\phi(I, g; \pi_i^k)\rho_I^{\mathrm{CF}}(\phi_{I,g}; \pi^k) \qquad \triangleright \text{ Cumulative immediate regret.}$$

$$y^t_{\phi_I} \leftarrow \sum_g w_\phi(I, g; \pi_i^t) f\left(x^t_{\phi_{I,g}}\right) \qquad \triangleright \text{ Link output/preference for } \phi_I.$$

$$L^t : \Delta^{|\mathcal{A}(I)|} \ni \sigma \mapsto \frac{1}{z^t} \sum_{\phi_I} y^t_{\phi_I} \phi_I(\sigma). \qquad \triangleright \text{ Convex combination.}$$

- Permits:
  - choice of link function f, *e.g.*, $f(\cdot) = \max\{0, \cdot\}$ or $f(\cdot) = \mathrm{e}^{\eta \cdot}$.
  - approximating $x^t_{\phi_{I,g}}$ instead of storing it in a table, and

Time Selection Regret Matching

- Sets $\pi_i^t(I) \leftarrow$ to be a fixed point of a linear operator, L^t:

$$x^t_{\phi_{I,g}} \leftarrow \sum_k^{t-1} w_\phi(I, g; \pi_i^k)\rho_I^{\mathrm{CF}}(\phi_{I,g}; \pi^k) \qquad \triangleright \text{ Cumulative immediate regret.}$$

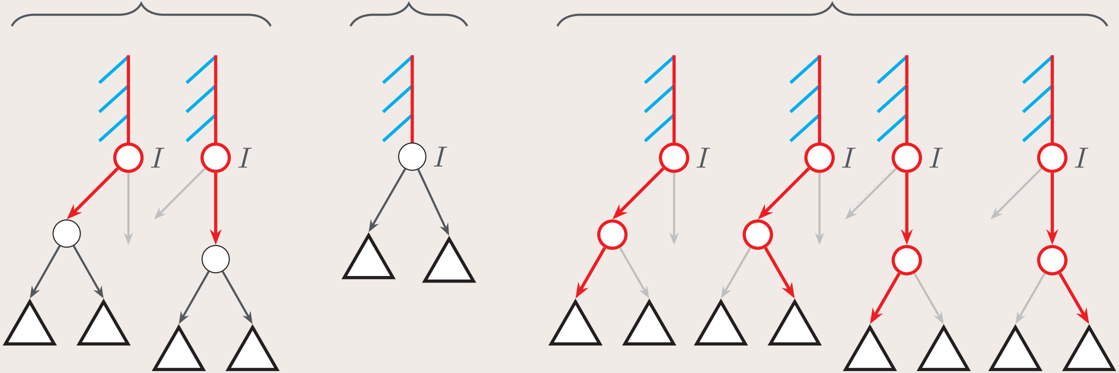
$$y^t_{\phi_I} \leftarrow \sum_g w_\phi(I, g; \pi_i^t) f\left(x^t_{\phi_{I,g}}\right) \qquad \triangleright \text{ Link output/preference for } \phi_I.$$

$$L^t : \Delta^{|\mathcal{A}(I)|} \ni \sigma \mapsto \frac{1}{z^t} \sum_{\phi_I} y^t_{\phi_I} \phi_I(\sigma). \qquad \triangleright \text{ Convex combination.}$$

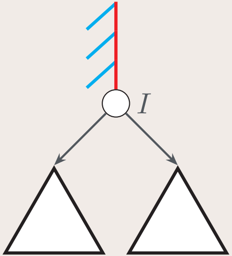
- Permits:
  - choice of link function f, *e.g.*, $f(\cdot) = \max\{0, \cdot\}$ or $f(\cdot) = \mathrm{e}^{\eta \cdot}$.
  - approximating $x^t_{\phi_{I,g}}$ instead of storing it in a table, and
  - predicting the next instantaneous regret for each $\phi_{I,g}$.

EFR: Regret Decomposition

Minimize immediate regret at children. Minimize immediate regret at parent. $\Longrightarrow$ Two-step regret is minimized.



Restricting EFR's Deviations to Improve Efficiency



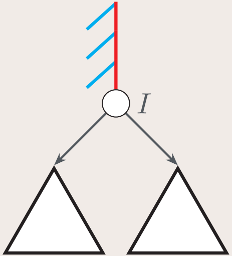
$\forall \phi \in \Phi^{\mathrm{IN}}_{\mathcal{I}_i}, g$, $w_\phi(I, g; \pi_i^t) \in [0, 1]$

$w_\phi(I, g; \pi_i^t)\rho_I^{\mathrm{CF}}(\phi_{I,g}; \pi^t)$

$\forall a,\ v_I(a; \pi^t)$

The # of regrets depends on the # of deviations and their realizable ($\exists \pi_i, w_\phi(I, g; \pi_i) > 0$) memory states.

Restricting EFR's Deviations to Improve Efficiency



$\forall \phi \in \Phi, g$, $w_\phi(I, g; \pi_i^t) \in [0, 1]$

$w_\phi(I, g; \pi_i^t)\rho_I^{\mathrm{CF}}(\phi_{I,g}; \pi^t)$

$\forall a,\ v_I(a; \pi^t)$

The # of regrets depends on the # of deviations and their realizable $(\exists \pi_i, w_\phi(I, g; \pi_i) > 0)$ memory states.

We can restrict EFR's deviation set to $\Phi \subseteq \Phi_{\mathcal{I}_i}^{\mathrm{IN}}$ to ensure efficiency and re-construct previous algorithms!

Reductions to Previous Algorithms

- EFR() = CFR [3]

[3] Zinkevich et al., “Regret Minimization in Games with Incomplete Information”.

Reductions to Previous Algorithms

- EFR() = CFR [3]
- EFR() ≈ ICFR [4]

[3] Zinkevich et al., "Regret Minimization in Games with Incomplete Information".
[4] Celli et al., "No-regret learning dynamics for extensive-form correlated equilibrium".

Reductions to Previous Algorithms

- EFR() = CFR [3]
- EFR() ≈ ICFR [4]
- EFR() ≈ PGPI [5]

[3] Zinkevich et al., “Regret Minimization in Games with Incomplete Information”.
[4] Celli et al., “No-regret learning dynamics for extensive-form correlated equilibrium”.
[5] Srinivasan et al., “Actor-Critic Policy Optimization in Partially Observable Multiagent Environments”; Morrill et al., “Hindsight and Sequential Rationality of Correlated Play”.

New Efficient Variants

- TIPS: EFR(), # deviations: $\mathcal{O}(d_* n_{\mathcal{A}}^3)$.

- TIPS: EFR(), # deviations: $\mathcal{O}(d_* n_{\mathcal{A}}^3)$.
- CSPS: EFR(), # deviations: $\mathcal{O}(d_* n_{\mathcal{A}}^2)$.

New Efficient Variants

- TIPS: EFR(), # deviations: $\mathcal{O}(d_* n_{\mathcal{A}}^3)$.
- CSPS: EFR(), # deviations: $\mathcal{O}(d_* n_{\mathcal{A}}^2)$.
- CFPS: EFR(), # deviations: $\mathcal{O}(d_* n_{\mathcal{A}}^2)$.

New Efficient Variants

- TIPS: EFR(), # deviations: $\mathcal{O}(d_* n_{\mathcal{A}}^3)$.
- CSPS: EFR(), # deviations: $\mathcal{O}(d_* n_{\mathcal{A}}^2)$.
- CFPS: EFR(), # deviations: $\mathcal{O}(d_* n_{\mathcal{A}}^2)$.
- BPS: EFR(), # deviations: $\mathcal{O}(d_* n_{\mathcal{A}})$.

A Sample of Representative Experimental Results

Learning Curves

— ACT_{IN} — CFR — TIPS — BHV

Learning Curves

— ACT_{IN} — CFR — TIPS — BHV

goofspiel$(5, \uparrow, N = 2)$

goofspiel$(4, \uparrow, N = 3)$

Learning Curves

— ACT_{IN} — CFR — TIPS — BHV

goofspiel$(5, \uparrow, N = 2)$

goofspiel$(4, \uparrow, N = 3)$

Sheriff$(N = 2)$

Learning Curves

— $\mathrm{ACT}_{\mathrm{IN}}$ — CFR — TIPS — BHV

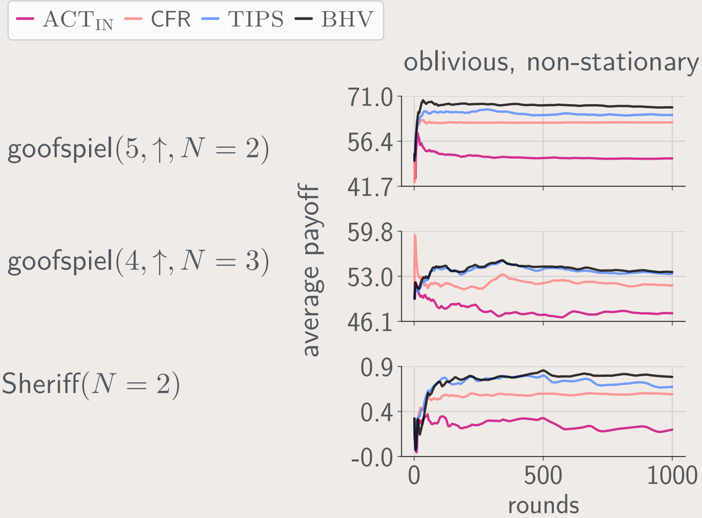
oblivious, non-stationary

goofspiel$(5, \uparrow, N = 2)$

goofspiel$(4, \uparrow, N = 3)$

Sheriff$(N = 2)$

Learning Curves

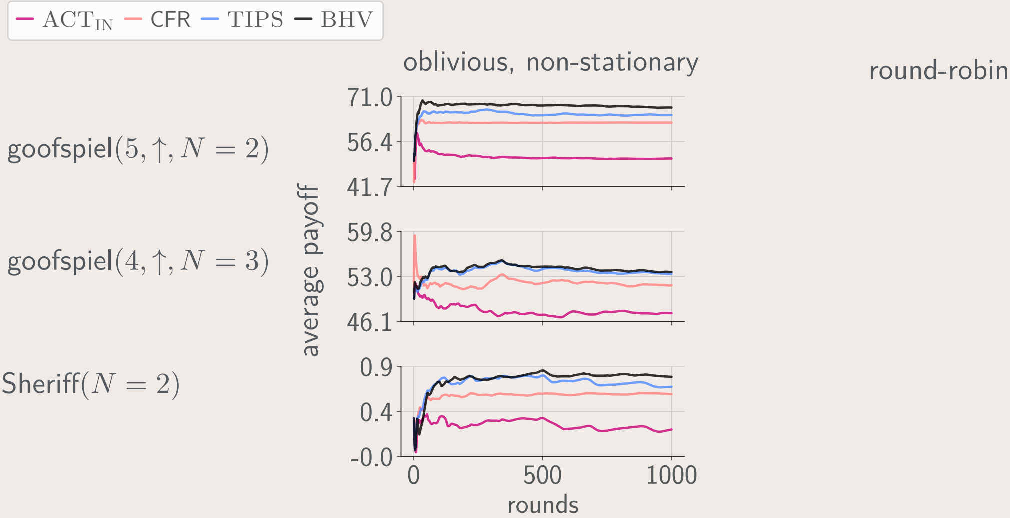


goofspiel$(5, \uparrow, N = 2)$

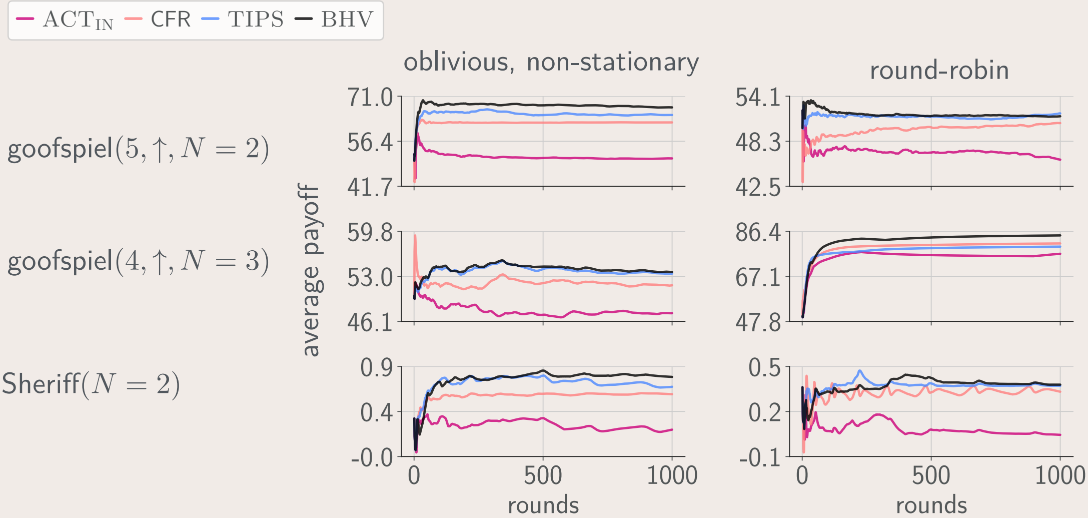
goofspiel$(4, \uparrow, N = 3)$

Sheriff$(N = 2)$

Learning Curves



Learning Curves



Conclusions

Conclusions

- Behavioral deviations are natural and expressive.

Conclusions

- Behavioral deviations are natural and expressive.
- EFR is hindsight rational for any given behavioral deviation subset with computation that scales closely with that set's complexity.

Conclusions

- Behavioral deviations are natural and expressive.
- EFR is hindsight rational for any given behavioral deviation subset with computation that scales closely with that set's complexity.
- The partial sequence deviations are efficient and powerful.

Conclusions

- Behavioral deviations are natural and expressive.
- EFR is hindsight rational for any given behavioral deviation subset with computation that scales closely with that set's complexity.
- The partial sequence deviations are efficient and powerful.
- EFR with a stronger deviation type tends to perform better.

Conclusions

Some remaining challenges:

Conclusions

Some remaining challenges:

- Stronger deviation types require more computation.

Conclusions

Some remaining challenges:

- Stronger deviation types require more computation.
- Stronger deviation types lead to worse bounds with respect to weaker types.

Conclusions

Some remaining challenges:

- Stronger deviation types require more computation.
- Stronger deviation types lead to worse bounds with respect to weaker types.

Possible solutions:

Conclusions

Some remaining challenges:

- Stronger deviation types require more computation.
- Stronger deviation types lead to worse bounds with respect to weaker types.

Possible solutions:

- Characterize the potential benefit of a stronger deviation type in a given game.

Conclusions

Some remaining challenges:

- Stronger deviation types require more computation.
- Stronger deviation types lead to worse bounds with respect to weaker types.

Possible solutions:

- Characterize the potential benefit of a stronger deviation type in a given game.
- Navigate tradeoffs by using ideas from the fixed-share forecaster [6] and context tree weighting [7].

[6] Herbster and Warmuth, "Tracking the best expert".

[7] Willems, Shtarkov, and Tjalkens, "Context tree weighting: a sequential universal source coding procedure for FSMX sources".

Conclusions

Some remaining challenges:

- Stronger deviation types require more computation.
- Stronger deviation types lead to worse bounds with respect to weaker types.

Possible solutions:

- Characterize the potential benefit of a stronger deviation type in a given game.
- ~~Navigate tradeoffs by using ideas from the fixed-share forecaster [6] and context tree weighting [7].~~
- Weighting deviation regrets to improve performance with respect to weaker deviation types.

[6] Herbster and Warmuth, "Tracking the best expert".

[7] Willems, Shtarkov, and Tjalkens, "Context tree weighting: a sequential universal source coding procedure for FSMX sources".

Efficient Deviation Types and Learning in Extensive-Form Games

Dustin Morrill, Ryan D'Orazio, Marc Lanctot,
James R. Wright, Michael Bowling, Amy R. Greenwald

July 16, 2021 – Berkeley MARL Group