# Simplex NeuPL: Any-Mixture Bayes-Optimality in Symmetric Zero-sum Games

read paper

Siqi Liu[1, 2], Marc Lanctot[2], Luke Marris[1, 2], Nicolas Heess[2]

University College London, UK[1]
DeepMind[2]

ICML 2022

# Population Learning: the hammer and the nail.

## If population learning is the solution, what is the problem?

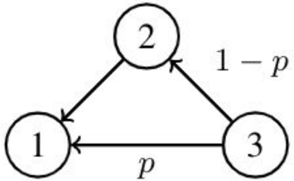
Convergence to an Nash Equilibrium (NE) of the game:

- **Robust to adversarial exploit:** no opponent can profitably deviate from their strategy.
- **Minimax optimal:** optimal when the opponent plays minimax optimally.

Population learning with convergence to NE can be restrictive:

- **Arbitrarily suboptimal:** if the opponent does NOT play minimax optimally.
- **Cannot BR to all but a few mixed-strategies from the population:**
  - E.g. the `(0.0 0.5, 0.5)` mixture policy can be executed at test time, but its best-response requires further training.
- **Cannot incorporate subjective Bayesian opponent priors at test-time.**

$$\begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ p & 1-p & 0 \end{bmatrix}$$




PSRO-NASH

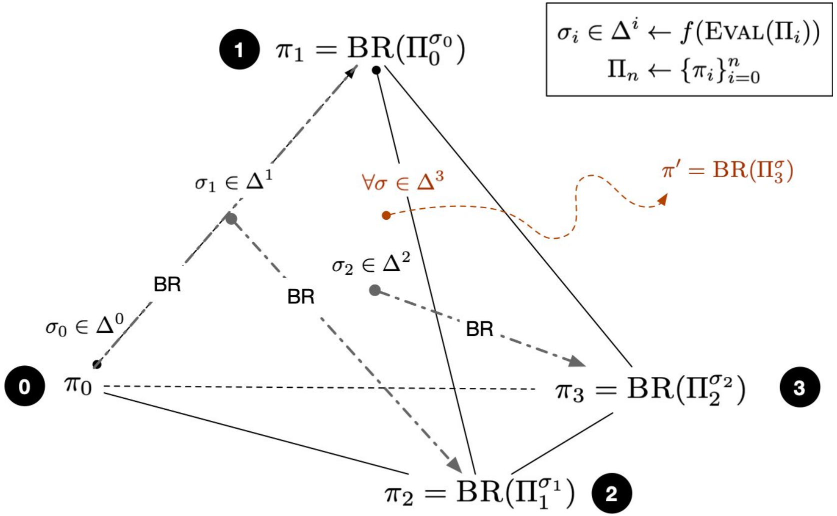
[1] Wellman, Michael P. "Methods for empirical game-theoretic analysis." AAAI. 2006.
[2] Lanctot, Marc, et al. "A unified game-theoretic approach to multiagent reinforcement learning." Advances in neural information processing systems 30 (2017).
[3] Balduzzi, David, et al. "Open-ended learning in symmetric zero-sum games." International Conference on Machine Learning. PMLR, 2019.

# Beyond NE: generalising over the Population Simplex

**Any-mixture Bayes-optimality:**

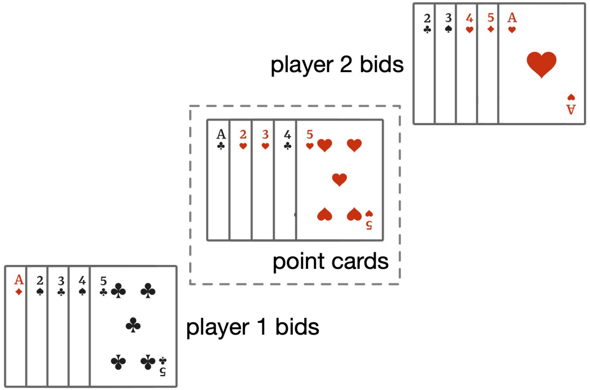
- **Convergence to an Nash Equilibrium (NE):** procedure terminates when we fails to expand the simplex.
- **Bayes-optimality under *any* opponent prior.** Learned policies trades off exploration and exploitation optimally to maximize returns.

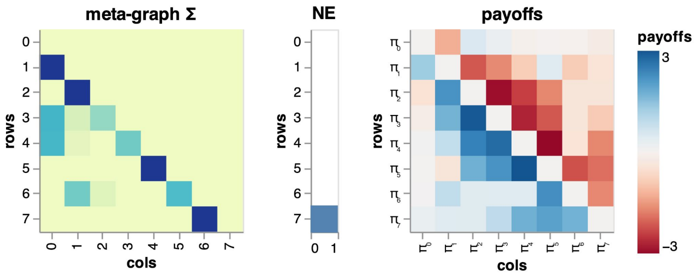
$$\pi^* = \mathrm{BR}(\Pi^{\sigma})$$
$$= \mathrm{argmax}_{\pi}\left[\mathbb{E}_{i\sim\sigma}[\mathbb{E}_{\pi,\pi_i}[\sum_t r_t]]\right]$$

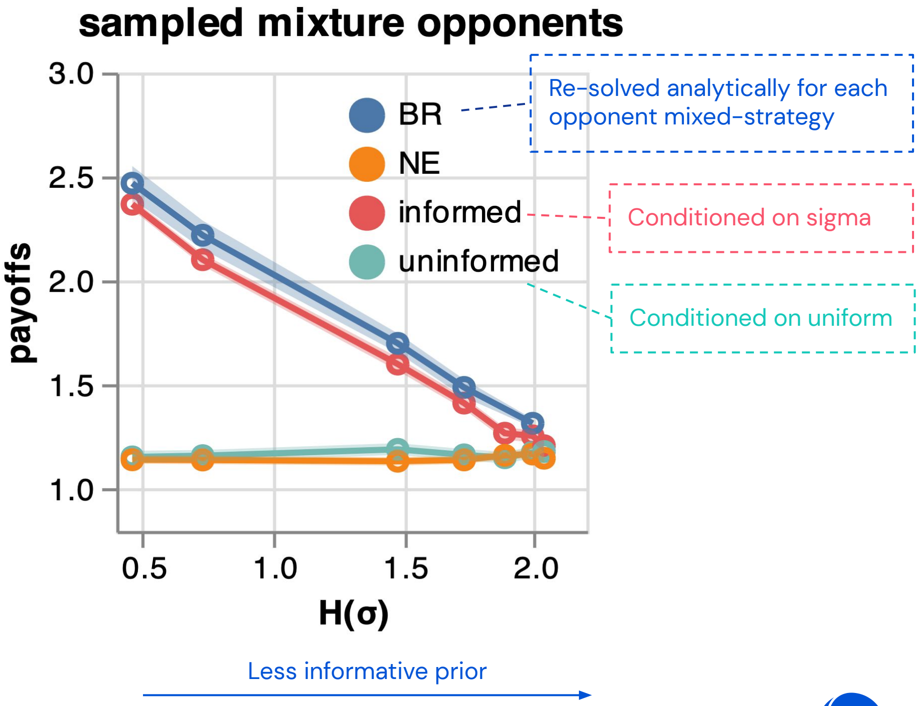



But … we need an efficient BR operator that can generalise across the entire expanding simplex.

# Result 1: Any-mixture Bayes-optimal return

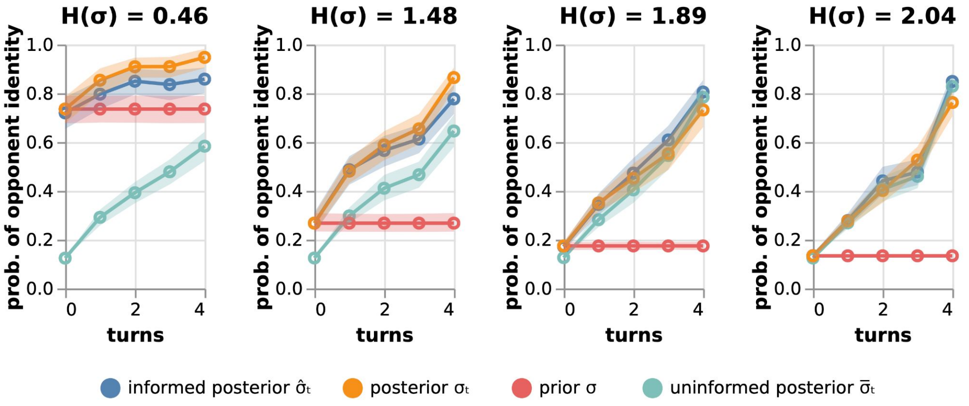


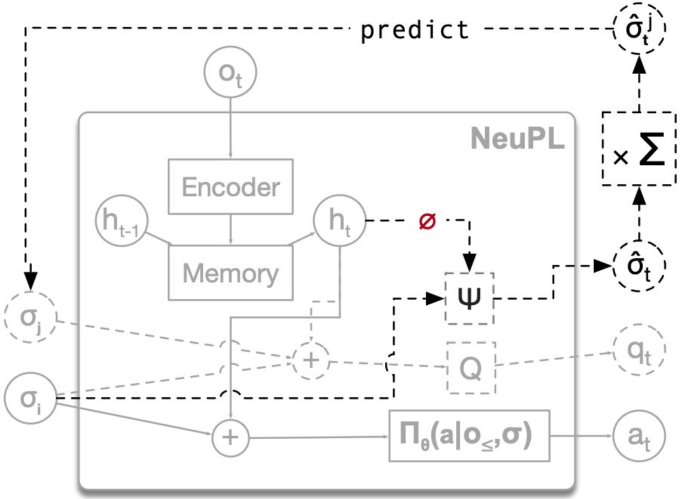



# Result 2: Posterior Inference via Bayesian MTRL

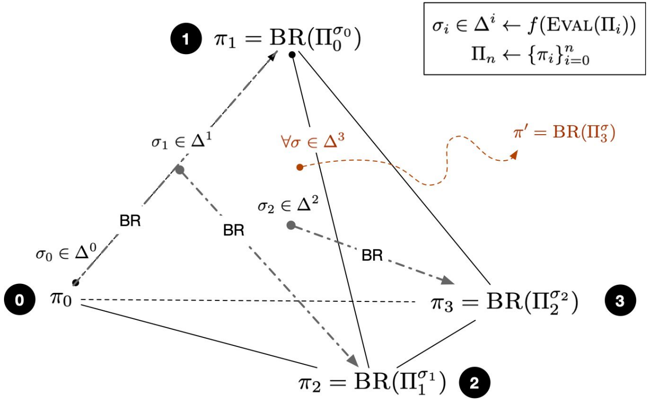
Implicit posterior readout
(with stop-grad)

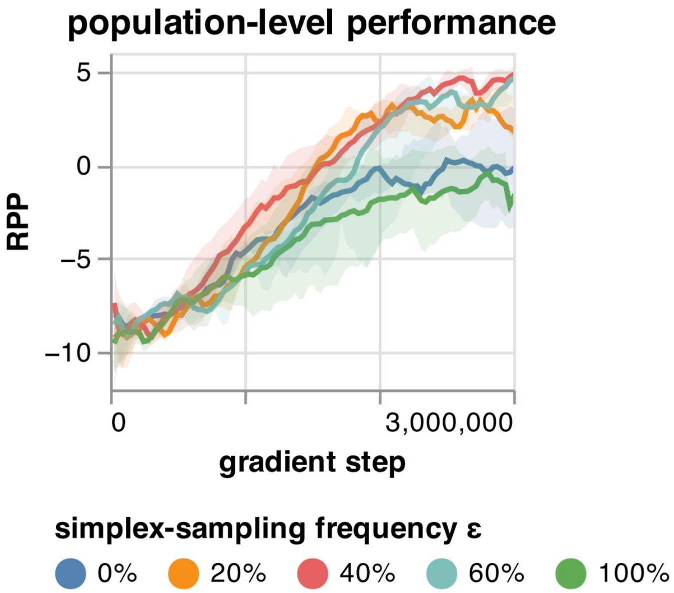



[1] Ortega, Pedro A., et al. "Meta-learning of Sequential Strategies." (2019).
[2] Mikulik, Vladimir, et al. "Meta-trained agents implement bayes-optimal agents." Advances in neural information processing systems 33 (2020): 18691-18703.

# Result 3: Improved Population Learning




- **NeuPL:** transfer learning over vertices of the population simplex.
- **Simplex-NeuPL:** transfer learning across the entire simplex!

# Conclusion & Future Works

- **Game-Theoretic:** preserves convergence guarantees to NE (extending NeuPL);
- **Bayes-optimal adaptive behaviors:** infer and exploit opponents optimally under opponent prior.
- **Transfer of skills** across population simplex;
- **Efficient & Scalable:** represents a population of strategies, **as well as Bayes-optimal responses to all their mixtures**, within a single conditional network.

- **Future Works:**
  - Beyond symmetric zero-sum games.

- **Come visit us at Session 3 Track 8 for more discussions & results!**