A Procedure to Nudge Outcome Selection with Algorithmic Learners

Mateus Hiro Nagata & Francesco Giordano

HEC Paris

May 14, 2025

Before Proceeding

Applications/Interpretations of Learning

1. Boundedly rational model
2. Algorithm self-play
3. How to play against an algorithm?

Learning Generates Equilibrium?

Can bounded rational agents/AI/algorithms achieve equilibrium?

- Epistemological conditions for NE (Aumann and Brandenburger, 1995)
- Mixed results: There is no uncoupled learning algorithm that achieve NE in all games (Hart and Mas-Colell, 2003) but empirical distributions of play converge to the set of CE of the game (Hart and Mas-Colell, 2000; Foster and Vohra, 1997)
- "Game theory is somewhat unusual in having the notion of an equilibrium without associated dynamics that give rise to the equilibrium" Arrow (1986); Shoham et al. (2007)
- Strategic decision-making with possibly non-human subjects can have surprising outcomes (Calvano et al., 2020)

The Gap

- P.1 Which equilibrium does it converges to? (Canyakmaz et al., 2024; Borowski et al., 2019)
- P.2 Complexity of finding payoff improving CE (Barman and Ligett, 2015)
- P.3 Convergence guarantees are on distribution of play, not on actual play (Borowski et al., 2019)

Q

Can we devise a welfare-improving learning algorithm?

- New algorithm and information design

This Paper

1. Introduce message-mediated augmentation of a learning algorithm
2. Apply to a variant of fictitious play
3. Simulate the discrete-time version
4. Analyze convergence in continuous time
5. Compare welfare with and without messages

Results

1. Achieve welfare correlated version of the equilibrium (NE or QRE)
2. Can be payoff increasing
3. Show which message distribution it works (Robust Information Design)

Similar Papers

The Assumption vs Result Gap

- Q-learning (forward-looking) can converge to correlated equilibria but requires state-dependent payoffs (Greenwald et al., 2003)
- Requires a complicated algorithm that alternates between exploration and exploitation (Borowski et al., 2019; Marden, 2017)
- Induce learning by varying the utilities in each stage game (Canyakmaz et al., 2024)

Game

$$\Gamma = (N, (A_i)_i, (u_i)_i)$$

- $N = \{1, 2\}$, Repeated
- Equilibrium analysis of the stage-game

	a	b
a	6, 6	2, 7
b	7, 2	0, 0

Correlated equilibrium

- Message distribution $\mu \sim \Delta(M), M = \times_i M_i$
- μ_i marginal distribution of μ
- Information Design perspective vs ML perspective

	m_1	m_2
m_1	$\frac{1}{3}$	$\frac{1}{3}$
m_2	$\frac{1}{3}$	0

	m_1	m_2
m_1	$\frac{1}{2}$	$\frac{1}{4}$
m_2	$\frac{1}{4}$	0

Weighted Stochastic Fictitious Play

$$Q_i^a(t) = (1-\alpha)Q_i^a(t-1) + u_i(a, a_{-i}(t)) \quad \text{(attraction update)}$$

$$\mathcal{X}_i^a(t) = \frac{exp(\beta Q_i^a(t))}{\sum_{a \in A_i} exp(\beta Q_i^a(t))} \quad \text{(softmax)}$$

$\alpha \in [0,1]$ Memory-loss

$\beta \in [0,\infty)$ Choice intensity

Approach

- Learning information: payoffs and non payoff-relevant messages
- Extension of EWA, extend Fictitious Play
- Intuitive algorithm
- Probability of action is monotonic in its past performance
- Encompass exploration, error, uncertainty
- Does not require knowledge of opponent's payoffs - Uncoupled

Weighted Stochastic Fictitious Play

$$Q_i^a(t) = (1-\alpha)Q_i^a(t) + u_i(a, a_{-i}(t)) \qquad \text{(attraction update)}$$

$$\mathcal{X}_i^a(t) = \frac{exp(\beta Q_i^a(t))}{\sum_{a \in A_i} exp(\beta Q_i^a(t))} \qquad \text{(softmax)}$$

$\alpha \in [0, 1]$ Memory-loss

$\beta \in [0, \infty)$ Choice intensity

Algorithm with Messages

Let $\mu \in \Delta(M)$ be a distribution over messages, $M = M_1 \times \cdots \times M_n$

$$Q_i^{a|m}(t) = \begin{cases} (1-\alpha)Q_i^{a|m}(t-1) + u_i(a, a_{-i}(t)), \text{ if } m_t = m \\ Q_i^{a|m}(t-1), \text{ otherwise} \end{cases} \quad \text{(attraction update)}$$

$$\mathcal{X}_i^{a|m}(t) = \begin{cases} \frac{exp(\beta Q_i^{a|m}(t))}{\sum_{a \in A_i} exp(\beta Q_i^{a|m}(t))}, \text{ if } m_t = m \\ 0, \text{ otherwise} \end{cases} \quad \text{(softmax)}$$

$\alpha \in [0,1]$ Memory-loss

$\beta \in [0,\infty)$ Choice intensity

Model

Primitives

Continuous-time limit

Simulations

Results

Conclusion

Simulation

For each combination (α, β) :

- 2 players play the Hawk-dove game
- Set initial values: $Q_i^{a|m}(0) = 0$, so $\mathcal{X}_i^{a|m}(t) = \frac{1}{|A_i|}$; fixed $\mu \in \Delta(M)$
- Episode length T $= 500$, last-iterate check if $\mathcal{X}_i^{a|m}(T)$ convergence to CE, NE or else (99.5% thershold)
- Repeat 100 times

Sim 1

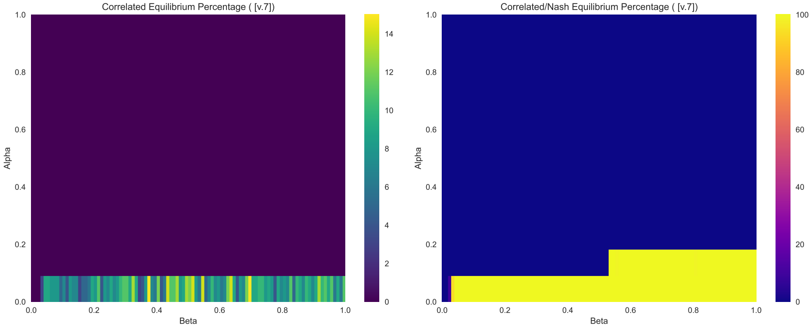


Figure: Convergence Percentage to the Pure Correlated Equilibrium (left) and to Correlated or Nash Equilibrium (right). Grid Heatmap: $\alpha \in [0, 1], \beta \in [0, 1]$

Sim 1'

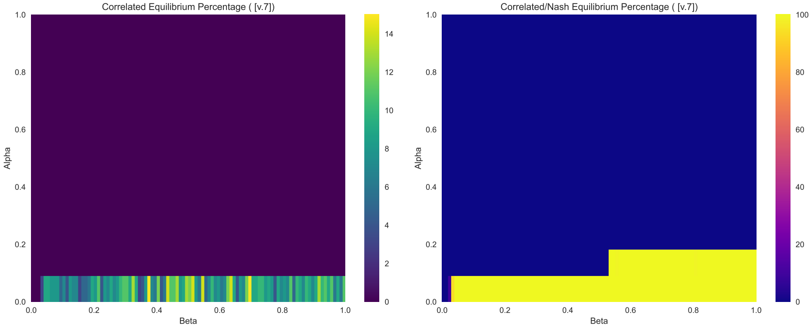


Figure: Convergence Percentage to the Pure Correlated Equilibrium (left) and to Correlated or Nash Equilibrium (right). Grid Heatmap: $\alpha \in [0, 1], \beta \in [0, 1]$

Robust Information Design

Price of Learning

$$PoL = \frac{\text{Welfare with messages}}{\text{Welfare without messages}}$$

- Everything is learnt in this framework: Payoffs, opponent's behavior, message correlation
- Bauch and Hartmann (2025)
- In the search of Optimal algorithm to calculate the optimal Price of Learning
- Robust Information Design (Feng et al., 2024)

Sim 2

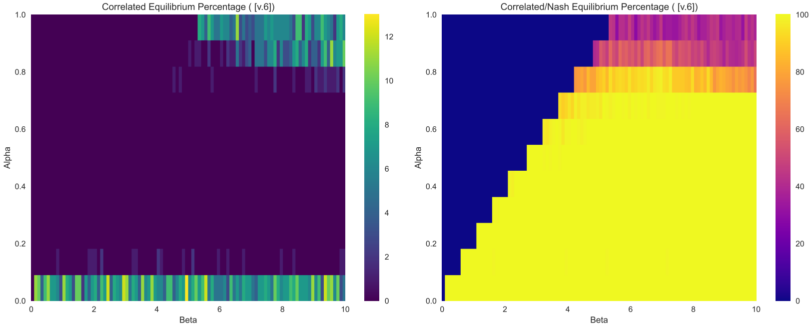


Figure: Convergence Percentage to the Pure Correlated Equilibrium (left) and to Correlated or Nash Equilibrium (right). Grid Heatmap: $\alpha \in [0, 1], \beta \in [0, 10]$

Sim 3

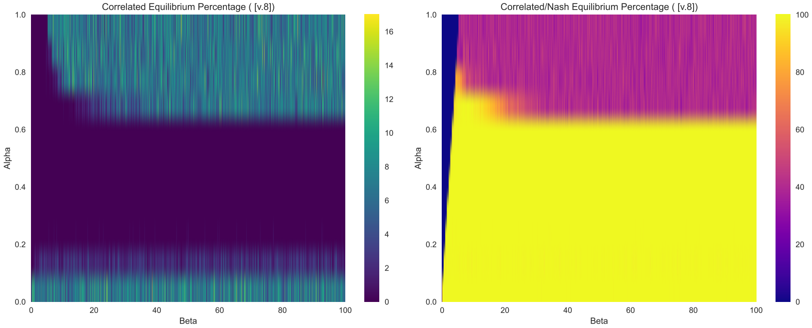


Figure: Convergence Percentage to the Pure Correlated Equilibrium (left) and to Correlated or Nash Equilibrium (right). Grid Heatmap: $\alpha \in [0, 1], \beta \in [0, 100]$

Time Difference Graph

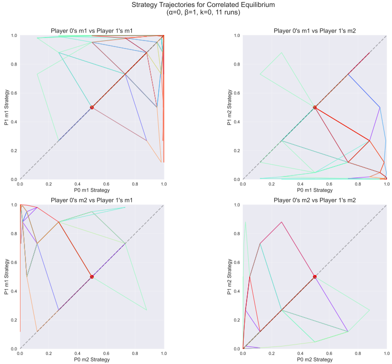


Figure: Strategy Trajectories for CE

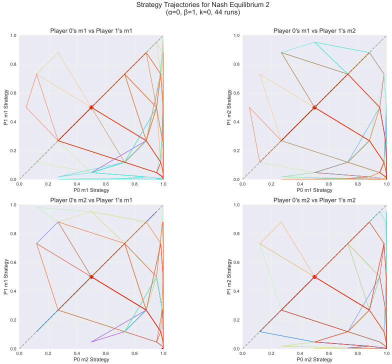


Figure: Strategy Trajectories for NE

Weighted Stochastic Fictitious Play

$$\mathcal{X}_i^a(t+1) = \frac{\mathcal{X}_i^a(t) e^{\beta[Q^a(t+1) - Q^a(t)]}}{\sum_{a' \in A_i} \mathcal{X}_i^{a'}(t) e^{\beta\left[Q^{a'}(t+1) - Q^{a'}(t)\right]}}$$

$$Q_i^a(t+1) - Q_i^a(t) = -\alpha Q_i^a(t) + u_i^t(a, a_{-i}(t))$$

Weighted Stochastic Fictitious Play

$$\dot{Q}_i^a(t) = u_i^a(t) - \alpha Q_i^a(t)$$

$$\dot{\mathcal{X}}_i^a(t) = \frac{d}{dt}\mathcal{X}_i^a(t) = \frac{d}{dt}\frac{\exp(\beta Q_i^a(t))}{\sum_{a' \in A_i} \exp(\beta Q_i^{a'}(t))}$$

$$= \beta \mathcal{X}_i^a(t) \left[\dot{Q}_i^a(t) - \sum_{a' \in A_i} \dot{Q}_i^{a'}(t) \mathcal{X}_i^{a'}(t) \right]$$

Weighted Stochastic Fictitious Play

$$\frac{\dot{\mathcal{X}}_i^a(t)}{\mathcal{X}_i^a(t)} = \beta \left[u_i^a(t) - \sum_{a' \in A_i} u_i^{a'}(t) \mathcal{X}_i^{a'}(t) \right] - \alpha \left[\ln \mathcal{X}_i^a(t) - \sum_{a' \in A_i} \mathcal{X}_i^{a'}(t) \ln \mathcal{X}_i^{a'}(t) \right]$$

Algorithm with Messages

$$Q_i^{a|m}(t) - Q_i^{a|m}(t-1) = \mathbb{1}(m_t = m)[u_i(a, a_{-i}(t)) - \alpha Q_i^{a|m}(t-1)] \tag{1}$$

$$\dot{Q}_i^{a|m}(t) = \mathbb{P}(m_t = m)[\mathbb{E}_{a_{-i} \sim \mathcal{X}_{-i}^{a_{-i}}(t)}[u_i(a, a_{-i}(t)] - \alpha Q_i^{a|m}(t-1)] \tag{2}$$

Messages

$$\mathcal{X}_i^{a|m}(t) = P(m_t = m)\frac{exp(\beta Q_i^{a|m}(t))}{\sum_{a \in A_i} exp(\beta Q_i^{a|m}(t))} \tag{3}$$

$$\dot{\mathcal{X}}_i^{a|m}(t) =$$

$$= \mathbb{P}(m_t = m)\beta \mathcal{X}_i^a(t)\left[\dot{Q}_i^a(t) - \sum_{a' \in A_i} \dot{Q}_i^{a'}(t)\mathcal{X}_i^{a'}(t)\right]$$

Final Message

$$\frac{\dot{\mathcal{X}}_i^{a|m}(t)}{\mathcal{X}_i^{a|m}(t)} = \beta\left[u_i(a, \mathcal{X}_{-i}^{a_{-i}}(t)) - \sum_{a' \in A_i} u_i(a, \mathcal{X}_i^{a_{-i}}(t))\mathcal{X}_i^{a'|m}(t)\right] - \alpha\left[\ln \mathcal{X}_i^{a|m}(t) - \sum_{a' \in A_i} \mathcal{X}_i^{a'|m}(t) \ln \mathcal{X}_i^{a'|m}(t)\right]$$

Proposition

All pure strategy profiles given message $\mathcal{X}_i^{a|m}(t) \in \{0, 1\}$ *are a fixed point. If memory-loss is positive,* $\alpha > 0$, *then the fixed points are unstable. If* $\alpha = 0$, *the pure-strategy fixed points are stable iff correlated equilibrium.*

Conclusion

- EWA with messages
- Fixed-point analysis, simulations
- Robust Information Design
- Memory and choice intensity have a proportional relation on reaching pure strategy outcomes

Work in Progress

- Time Convergence and Convergence Guarantees
- Can it be manipulated?
- More messages = better?
- Expansion to the dynamic case: can it be manipulated? Dynamic Information Design
- Study chaotic cycles
- Basin of attractions
- Sanders et al. (2018) shows that the basin of attractions get smaller as $n \to \infty$

References I

Arrow, K. J. (1986). Rationality of self and others in an economic system. *Journal of business*, pages S385–S399.

Aumann, R. and Brandenburger, A. (1995). Epistemic conditions for nash equilibrium. *Econometrica: Journal of the Econometric Society*, pages 1161–1180.

Barman, S. and Ligett, K. (2015). Finding any nontrivial coarse correlated equilibrium is hard. *ACM SIGecom Exchanges*, 14(1):76–79.

Bauch, G. and Hartmann, L. (2025). Correlation uncertainty: a decision-theoretic approach. *arXiv preprint arXiv:2503.13416*.

Borowski, H. P., Marden, J. R., and Shamma, J. S. (2019). Learning to play efficient coarse correlated equilibria. *Dynamic Games and Applications*, 9:24–46.

References II

Calvano, E., Calzolari, G., Denicolo, V., and Pastorello, S. (2020). Artificial intelligence, algorithmic pricing, and collusion. *American Economic Review*, 110(10):3267–3297.

Canyakmaz, I., Sakos, I., Lin, W., Varvitsiotis, A., and Piliouras, G. (2024). Steering game dynamics towards desired outcomes. *arXiv e-prints*, pages arXiv–2404.

Feng, Y., Ho, C.-J., and Tang, W. (2024). Rationality-robust information design: Bayesian persuasion under quantal response. In *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 501–546. SIAM.

Foster, D. P. and Vohra, R. V. (1997). Calibrated learning and correlated equilibrium. *Games and Economic Behavior*, 21(1-2):40–55.

References III

Greenwald, A., Hall, K., Serrano, R., et al. (2003). Correlated q-learning. In *ICML*, volume 3, pages 242–249.

Hart, S. and Mas-Colell, A. (2000). A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68(5):1127–1150.

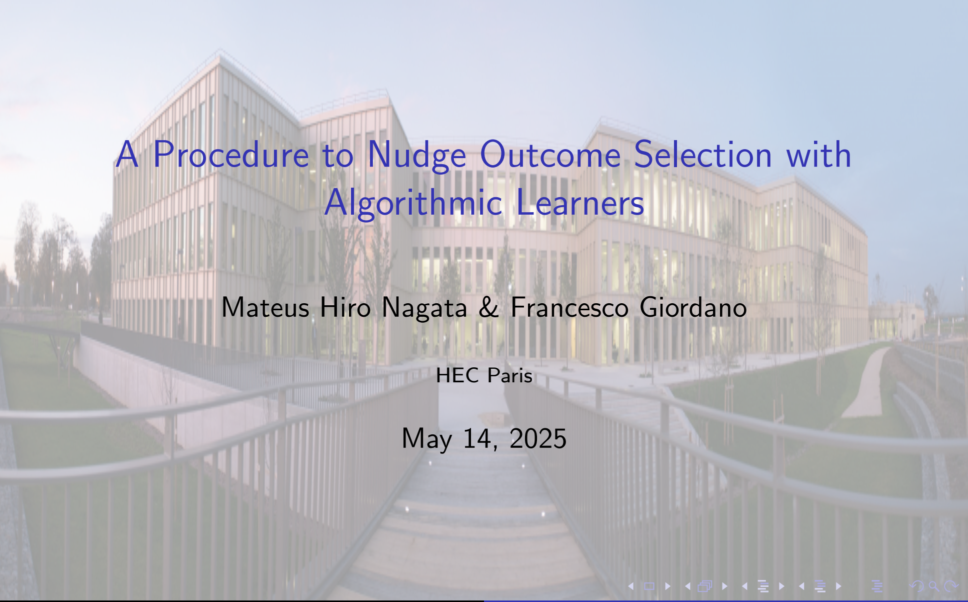
Hart, S. and Mas-Colell, A. (2003). Uncoupled dynamics do not lead to nash equilibrium. *American Economic Review*, 93(5):1830–1836.

Marden, J. R. (2017). Selecting efficient correlated equilibria through distributed learning. *Games and Economic Behavior*, 106:114–133.

References IV

Sanders, J. B., Farmer, J. D., and Galla, T. (2018). The prevalence of chaotic dynamics in games with many players. *Scientific reports*, 8(1):4902.

Shoham, Y., Powers, R., and Grenager, T. (2007). If multi-agent learning is the answer, what is the question? *Artificial intelligence*, 171(7):365–377.



A Procedure to Nudge Outcome Selection with Algorithmic Learners

Mateus Hiro Nagata & Francesco Giordano

HEC Paris

May 14, 2025