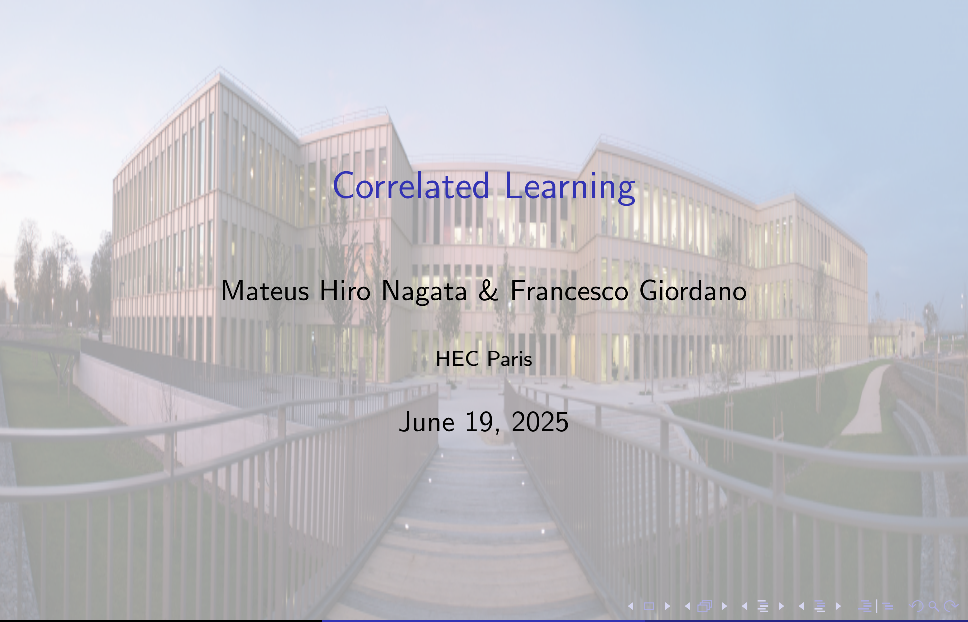


Correlated Learning

Mateus Hiro Nagata & Francesco Giordano

HEC Paris

June 19, 2025

Before Proceeding

Q

Can a learning algorithm converge to a welfare-improving Correlated Equilibrium (WICE)?

- Introduction of messages to learning (information design)
- Analytical results: stability, fixed-points, relation to quantal equilibrium
- Simulation results: positive convergence to WICE

Interpretations of Learning

1. Boundedly rational model - Epistemological conditions for NE (Aumann and Brandenburger, 1995)
2. Algorithm self-play
3. How to play against an algorithm?

Learning Generates Equilibrium?

Can bounded rational agents/AI/algorithms achieve equilibrium?

- Mixed results: no uncoupled learning algorithm that guarantee NE in all games (Hart and Mas-Colell, 2003) but empirical distributions of play converge to the set of CCE 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" (Shoham et al., 2007)
- Strategic decision-making with possibly non-human subjects can have unexpected outcomes (Calvano et al., 2020)

Motivation

The Gaps

P.1 Which equilibrium does it converges to? (Canyakmaz et al., 2024; Borowski et al., 2019)

P.2 How to explain failure of CE? (Cason and Sharma, 2007; Friedman et al., 2022)

P.3 Convergence guarantees are on distribution of play, not on actual play (Borowski et al., 2019)

Similar Papers

The Assumption vs Result Gap

- Correlated 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; Zhang et al., 2024)

Correlated Equilibrium

- Game with Messages $\Gamma = (N, (A_i)_i, (M_i)_i, (u_i)_i,)$
- Message distribution $\eta \sim \Delta(M)$, set of messages $M = \times_i M_i$
- Mixed action $x^t_{a_i} : M_i \to \Delta(A_i)$, utility $u_i : (A_i)_i \to \mathbb{R}$
- Expected utility of action $a_i \in A_i$ given message $m_i \in M_i$ as:

$$u^t_i(a_i|m_i) = \sum_{a_{-i}} \sum_{m_{-i}} u_i(a_i, a_{-i}) x^t_{a_{-i}}(m_{-i}) \eta(m_{-i}|m_i). \qquad (1)$$

- Correlated equilibrium: $(x^t_{a_i}(m_i))_{a_i,m_i,i}$ and message distribution $,\eta$ such that $\forall i \in N, \forall a_i, a'_i \in A_i, \forall m_i \in M_i$:

$$\eta(m_i) x^t_{a_i}(m_i)[u^t_i(a_i \mid m_i) - u^t_i(a'_i \mid m_i)] \geq 0$$

Hedge

Attraction player i has towards action a_i at time t: $Q^t_{a_i}$

$$x^t_{a_i} = \frac{\exp(\beta Q^t_{a_i})}{\sum_{a'_i \in A_i} \exp(\beta Q^t_{a'_i})} \qquad \text{(softmax)}$$

$$Q^{t+1}_{a_i} = (1-\alpha) Q^t_{a_i} + u_i(a_i, a^t_{-i}) \qquad \text{(attraction update)}$$

- α $\in [0, 1]$ Memory-loss
- β $\in [0, \infty)$ Choice intensity
- Full feedback

Algorithm with Messages

$$x_{a_i}^t(m_i) = \begin{cases} \frac{exp(\beta Q_{a_i}^t(m_i)}{\sum_{a_i' \in A_i} exp(\beta Q_{a_i'}^t(m_i))}, \text{ if } m_i^t = m_i \\ 0, \text{ otherwise} \end{cases} \qquad \text{(softmax)}$$

$$Q_{a_i}^{t+1}(m_i) = \begin{cases} (1-\alpha) Q_{a_i}^t(m_i) + u_i(a_i, a_{-i}^t), \text{ if } m_i^t = m_i \\ Q_{a_i}^t(m_i), \text{ otherwise} \end{cases} \qquad \text{(attraction update)}$$

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

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

Results - Preview

- Lemma 1: Tractable continuous-time approximation
- Lemma 2: Alternative formula
- Proposition 1: Deterministic actions given message are fixed points. If $\alpha = 0$, those are stable if and only if they are CE.
- Proposition 2: All fixed points are Quantal Correlated Equilibrium (Černỳ et al., 2022)

Continuous-time equivalent

The **continuous-time evolution** is defined via the expected infinitesimal increment:

$$\dot{Q}^t_{a_i}(m_i) = \frac{d}{dt}\mathbb{E}[Q^t_{a_i}(m_i) \mid \mathcal{F}^t] =$$

$$\lim_{\delta \to 0^+} \frac{\mathbb{E}_{m \sim \eta, a_{-i} \sim x^{t+\delta}_{a_{-i}}(m_{-i})}[Q^{t+\delta}_{a_i}(m_i) - Q^t_{a_i}(m_i) \mid \mathcal{F}^t]}{\delta}$$

Continuous-time

Lemma 1

The correlated Hedge algorithm has the following continuous-time equivalent: Proof

$$\dot{Q}^t_{a_i}(m_i) = \eta(m_i)[u^t_i(a_i|m_i) - \alpha Q^t_{a_i}(m_i)] \tag{2}$$

$$\dot{x}^t_{a_i}(m_i) = \beta x^t_{a_i}(m_i)\left[\dot{Q}^t_{a_i}(m_i) - \sum_{a' \in A_i} \dot{Q}^t_{a'_i}(m_i) x^t_{a'_i}(m_i)\right] \tag{3}$$

Alternative

Lemma 2

The continuous-time correlated hedge can be alternatively described by the formula: Proof

$$\dot{x}_{a_i}^t(m_i) = \beta x_{a_i}^t(m_i)\eta(m_i)\left[u_i^t(a_i|m_i) - \sum_{a_i' \in A_i} x_{a_i'}^t(m_i)u_i^t(a_i'|m_i)\right] -$$

$$\alpha x_{a_i}^t(m_i)\eta(m_i)\left[\ln(x_{a_i}^t(m_i)) - \sum_{a_i' \in A_i} x_{a_i'}^t(m_i)\ln(x_{a_i'}^t(m_i))\right]$$

Interpretation

Observation 1 (Reinforcement Condition)

The first expression in brackets is positive $\Leftrightarrow$ *Action* a_i*'s average utility is higher than the utility of playing the mixed action* $x_{a_i}^t$*.*

$$\beta x_{a_i}^t(m_i)\eta(m_i)\left[u_i^t(a_i|m_i)-\sum_{a_i'\in A_i}x_{a_i'}^t(m_i)u_i^t(a_i'|m_i)\right].$$

Furthermore, β *only influences the system through the reinforcement condition.*

Fixed-point Definition

Definition

A fixed point $x^* = (x^*_{a_i}(m_i))_{a_i \in A_i, m_i \in M_i, i \in N}$ *is* **stable** *if for every* $\varepsilon > 0$*, there is a* $\delta > 0$ *such that* $\forall t \geq 0$

$$||x^0 - x^*|| \leq \delta \Rightarrow ||x^t - x^*|| \leq \varepsilon$$

where $x^t = (x^t_{a_i}(m_i))_{a_i \in A_i, m_i \in M_i, i \in N}$.

Definition

A stable fixed point $x^* = (x^*_{a_i}(m_i))_{a_i \in A_i, m_i \in M_i, i \in N}$ *is* **asymptotically stable** *if there is a* $\delta > 0$ *such that*

$$||x^0 - x^*|| \leq \delta \Rightarrow \lim_{t \to \infty} x^t = x^*.$$

Fixed-points and Stability

Proposition 1

Correlated hedge on 2×2 game with 2 messages.

- *Pure strategy $|$message profile is a fixed point* [1]
- *If full-memory, $\alpha = 0$: pure strategy$|$message profile is stable $\Leftrightarrow$ correlated equilibrium.* Proof

[1] $(x_{a_i}(m_i))_{m_i \in M_i, a_i \in A_i, i \in N} \in \{0,1\}^{|M| \times |A|}$

QCE

Proposition 2

Correlated hedge on a 2×2 game with 2 messages. Then, all fixed points are equivalent to a Per-signal Quantal Correlated Equilibrium (S-QCE). Proof

Per-signal Quantal Correlated Equilibrium (S-QCE): Signaling scheme $\eta \in \Delta(M)$ and mixed action $(x_{a_i}(m_i))_{a_i,m_i,i}$ if there is a positive and increasing function $q_i(\cdot)$:

$$x_{a_i}(m_i) = \frac{q_i(u_i(a_i|m_i))}{\sum_{a_i' \in A_i} q_i(u_i(a_i'|m_i))}$$

We prove for $q_i(z) = \exp(\frac{\beta}{\alpha} z)$.

Game

$N = \{1, 2\}$

	a_2	b_2
a_1	$6, 6$	$2, 7$
b_1	$7, 2$	$0, 0$

η :

	m_{a_2}	m_{b_2}
m_{a_1}	$\frac{1}{3}$	$\frac{1}{3}$
m_{b_1}	$\frac{1}{3}$	0

η' :

	m_{a_2}	m_{b_2}
m_{a_1}	$\frac{1}{2}$	$\frac{1}{4}$
m_{b_1}	$\frac{1}{4}$	0

Simulation

For each combination (α, β) :

- Hawk-dove game
- Set initial values: $Q^t_{a_i}(m_i) = 0$, so $x^t_{a_i}(m_i) = \frac{1}{|A_i|}$; fixed η
- Obedient CE: $(x_{a_i}(m_{a_i}), x_{b_i}(m_{b_i})) = (1, 1), i = 1, 2$[2]
- Episode length T $= 500$, last-iterate check if $(x^T_{a_i}(m_i))_{a_i, m_i, i \in A_i \times M_i \times N}$ is the obedient CE, NE or else (99.5% thershold)
- 100 simulations for each parameter combination

[2]Pure NE: 1) $(x_{a_1}(m_{a_1}), x_{b_1}(m_{b_1})) = (1, 0), (x_{a_2}(m_{a_2}), x_{b_2}(m_{b_2})) = (0, 1)$, 2) $(x_{a_1}(m_{a_1}), x_{b_1}(m_{b_1})) = (0, 1), (x_{a_2}(m_{a_2}), x_{b_2}(m_{b_2})) = (1, 0)$

Sim 1 (η)

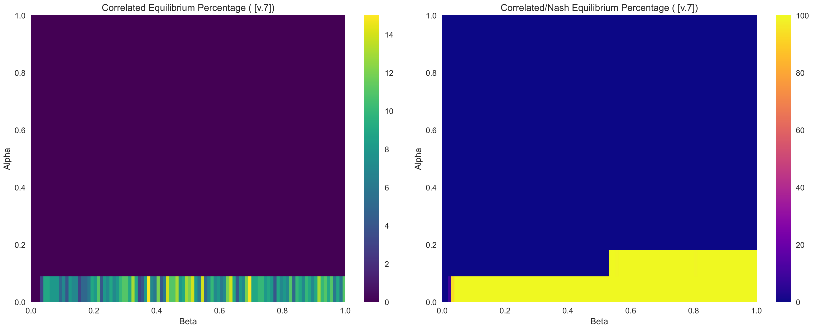


Figure: Convergence to Obedience (left) to Obedience or Nash (right). Grid Heatmap: $\alpha \in [0,1], \beta \in [0,1]$

Sim 1' (η')

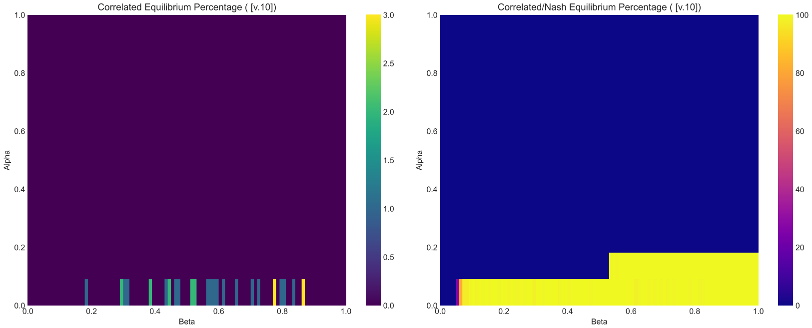


Figure: Convergence to Obedience (left) to Obedience or Nash (right). Grid Heatmap: $\alpha \in [0,1], \beta \in [0,1]$

Robust Information Design

Price of Learning

$$PoL_\eta = \frac{\text{Theoretically Optimal Welfare - Welfare Induced by } \eta}{\text{Theoretically Optimal Welfare}}. \tag{4}$$

- $SW = \sum_i \sum_m \sum_a \eta(m) x^t_{a_i}(m_i) x^t_{a_{-i}}(m_{-i}) u_i(a_i, a_{-i})$
- $SW_\eta = 9.3, SW_{\eta'} = 9, SW^* = 10.5$
- Everything is learnt in this framework: Payoffs, opponent's behavior, correlation. Cason and Sharma (2007): CE may not be achieved because of epistemological reasons.
- Robust Information Design

Sim 2

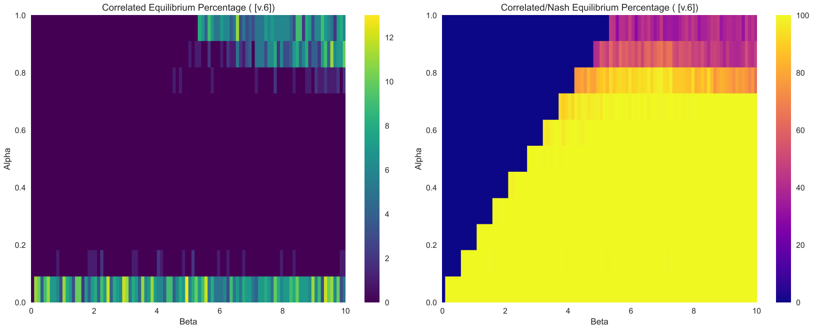


Figure: Convergence to Obedience (left) to Obedience or Nash (right). Grid Heatmap: $\alpha \in [0, 1], \beta \in [0, 10]$

Sim 3

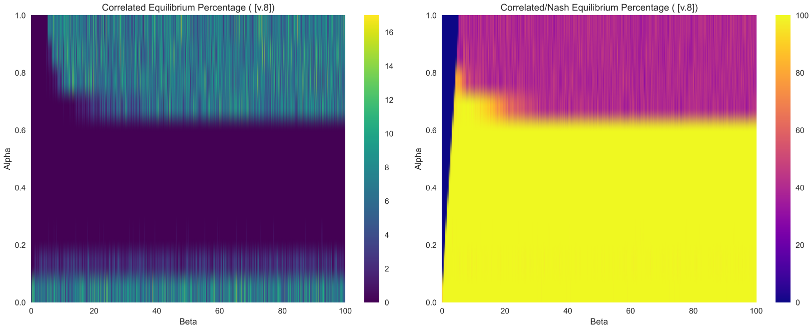


Figure: Convergence to Obedience (left) to Obedience or Nash (right). Grid Heatmap: $\alpha \in [0, 1], \beta \in [0, 100]$

Conclusion

- Algorithm with Messages
- Fixed-point analysis, simulations
- Robust Information Design
- α/β ratio

References I

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

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.

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). Learning and steering game dynamics towards desirable outcomes. *arXiv preprint arXiv:2404.01066*.

References II

Cason, T. N. and Sharma, T. (2007). Recommended play and correlated equilibria: an experimental study. *Economic Theory*, 33(1):11–27.

Černỳ, J., An, B., and Zhang, A. N. (2022). Quantal correlated equilibrium in normal form games. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 210–239.

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

Friedman, D., Rabanal, J. P., Rud, O. A., and Zhao, S. (2022). On the empirical relevance of correlated equilibrium. *Journal of Economic Theory*, 205:105531.

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.

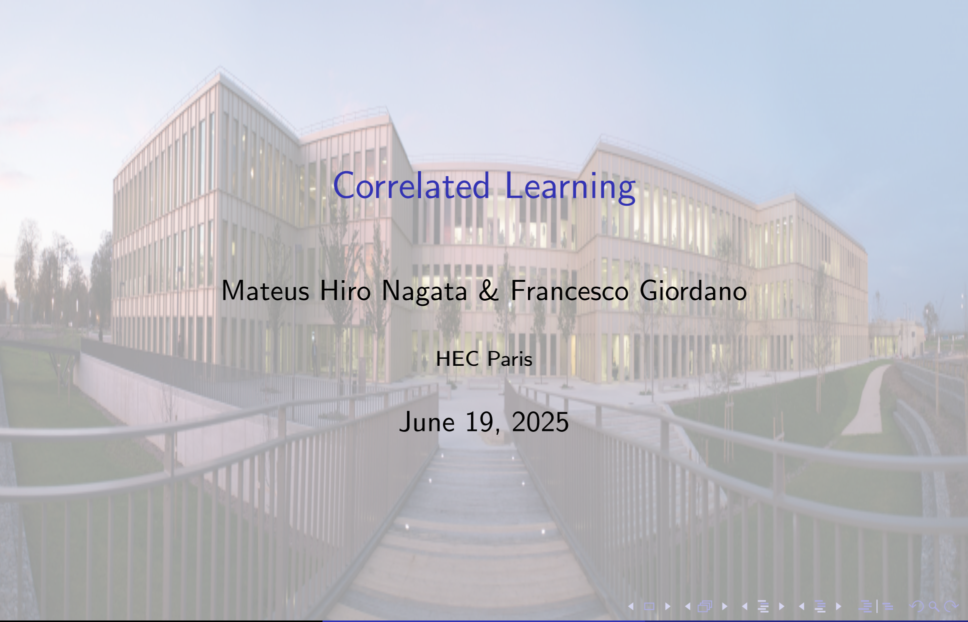
Hirsch, M. W., Smale, S., and Devaney, R. L. (2013). *Differential equations, dynamical systems, and an introduction to chaos*. Academic press.

References IV

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

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.

Zhang, R., Shamma, J., and Li, N. (2024). Equilibrium selection for multi-agent reinforcement learning: A unified framework. *arXiv preprint arXiv:2406.08844*.



Correlated Learning

Mateus Hiro Nagata & Francesco Giordano

HEC Paris

June 19, 2025

Lemma 1

Now, evaluating the expected value:

$$\begin{aligned}
&\mathbb{E}[\mathbb{1}(m_i^{t+\delta} = m_i) u_i(a_i, a_{-i}^{t+\delta})] \\
&= \eta(m_i)\mathbb{E}[u_i(a_i, a_{-i}^{t+\delta}) \mid m_i^{t+\delta} = m_i] \\
&= \eta(m_i) \sum_{m_{-i}} \eta(m_{-i} \mid m_i) \sum_{a_{-i}} u_i(a_i, a_{-i}) x_{a_{-i}}^{t+\delta}(m_{-i}) \\
&= \eta(m_i) u_i^{t+\delta}(a_i \mid m_i)
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\lim_{\delta \to 0^+} \frac{\mathbb{E}[Q_{a_i}^{t+\delta}(m_i) - Q_{a_i}^t(m_i) \mid \mathcal{F}^t]}{\delta} \\
&= \lim_{\delta \to 0^+} \eta(m_i) \left[\mathbb{E}[u_i(a_i, a_{-i}^{t+\delta}) \mid m_i^{t+\delta} = m_i, \mathcal{F}^t] - \alpha Q_{a_i}^t(m_i)\right] \\
&= \eta(m_i) \left[u_i^t(a_i \mid m_i) - \alpha Q_{a_i}^t(m_i)\right]
\end{aligned}$$

Q-value Dynamics

The average Q-value update for a small step size $\delta > 0$ is:

$$\frac{\mathbb{E}[Q_{a_i}^{t+\delta}(m_i) - Q_{a_i}^{t}(m_i) \mid \mathcal{F}^t]}{\delta} = \mathbb{E}[\mathbb{1}(m_i^{t+\delta} = m_i)(u_i(a_i, a_{-i}^{t+\delta}) - \alpha Q_{a_i}^t(m_i))]$$

Time derivative of the softmax policy:

$$\begin{aligned}
\dot{x}_{a_i}^t(m_i) &= \frac{d}{dt}\frac{\exp(\beta Q_{a_i}^t(m_i))}{\sum_{a_i'} \exp(\beta Q_{a_i'}^t(m_i))} \\
&= \frac{\beta \dot{Q}_{a_i}^t(m_i)\exp(\beta Q_{a_i}^t(m_i))\sum_{a_i'} \exp(\beta Q_{a_i'}^t(m_i))}{\left(\sum_{a_i'} \exp(\beta Q_{a_i'}^t(m_i))\right)^2} - \\
&\quad \frac{\exp(\beta Q_{a_i}^t(m_i))\sum_{a_i'} \beta \dot{Q}_{a_i'}^t(m_i)\exp(\beta Q_{a_i'}^t(m_i))}{\left(\sum_{a_i'} \exp(\beta Q_{a_i'}^t(m_i))\right)^2}
\end{aligned}$$

Which results in

$$\dot{x}^t_{a_i}(m_i) = \beta x^t_{a_i}(m_i)\left[\dot{Q}^t_{a_i}(m_i) - \sum_{a'\in A_i} \dot{Q}^t_{a'_i}(m_i)x^t_{a'_i}(m_i)\right].$$

Lemma 1

Lemma 2

We can rewrite the mixed action as Lemma 2

$$\ln(x^t_{a_i}(m_i)) = \beta Q^t_{a_i}(m_i) - \ln\left(\sum_{a'_i \in A_i} e^{\beta Q^t_{a'_i}(m)}\right)$$

Rearranging

$$Q^t_{a_i}(m_i) = \frac{1}{\beta}\ln\left(x^t_{a_i}(m_i)\right) + \frac{1}{\beta}\ln\left(\sum_{a'_i \in A_i} e^{\beta Q^t_{a'_i}(m)}\right).$$

Thus, arriving at:

$$\dot{x}^t_{a_i}(m_i) = \beta x^t_{a_i}(m_i)\eta(m_i)\left[u^t_i(a_i|m_i) - \sum_{a_i \in A_i} x^t_{a_i}(m_i)u^t_i(a_i|m_i)\right] -$$

$$\alpha x^t_{a_i}(m_i)\eta(m_i)\left[\ln(x^t_{a_i}(m_i)) - \sum_{a'_i \in A_i} x^t_{a'_i}(m_i)\ln(x^t_{a'_i}(m_i))\right]$$

Proposition 1

Suppose $x^t_{a_i}(m_i) = 0$, then

$$\dot{x}^t_{a_i}(m_i) = 0 \ln(0).$$

Suppose $x^t_{a_i}(m_i) = 1$, then

$$\dot{x}^t_{a_i}(m_i) = \beta\eta(m_i)\left[u^t_i(a_i \mid m_i) - u^t_i(a_i \mid m_i)\right] - \alpha\eta(m_i)\left[\ln(1) - \ln(1)\right] = 0.$$

Proposition 1

Lyapunov Linearization Theorem which states that if all eigenvalues of the Jacobian have strictly negative parts, then it is asymptotically stable (Hirsch et al., 2013).

Proposition 1

And since we are in 2×2 game, we could write the continuous-time equivalent as

$$\dot{x}^t_{a_i}(m_i) = \beta x^t_{a_i}(m_i)\eta(m_i)(1 - x^t_{a_i}(m_i))\left[u^t_i(a_i|m_i) - u^t_i(a'_i|m_i)\right] - \alpha x^t_{a_i}(m_i)\eta(m_i)(1 - x^t_{a_i}(m_i))\left[\ln\left(\frac{x^t_{a_i}(m_i)}{1 - x^t_{a_i}(m_i)}\right)\right]$$

and the average utility of playing a_i given m_i as

$$u^t_i(a_i|m_i) = \eta(m_{-i}|m_i)[u_i(a_i, a_{-i})x^t_{a_{-i}}(m_{-i}) + u_i(a_i, a'_{-i})(1 - x^t_{a_{-i}}(m_{-i}))] + \eta(m'_{-i}|m_i)[u_i(a_i, a_{-i})(1 - x^t_{a'_{-i}}(m'_{-i})) + u_i(a_i, a'_{-i})x_{a'_{-i}}(m'_{-i}))].$$

Proposition 1

The Jacobian is defined as follows:

$$J(x) = \begin{bmatrix} \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a_i}(m_i)} & \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a'_i}(m'_i)} & \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a_{-i}}(m_{-i})} & \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a'_{-i}}(m'_{-i})} \\ \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a_i}(m_i)} & \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a'_i}(m'_i)} & \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a_{-i}}(m_{-i})} & \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a'_{-i}}(m'_{-i})} \\ \frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x_{a_i}(m_i)} & \frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x_{a'_i}(m'_i)} & \frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x_{a_{-i}}(m_{-i})} & \frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x_{a'_{-i}}(m'_{-i})} \\ \frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x_{a_i}(m_i)} & \frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x_{a'_i}(m'_i)} & \frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x_{a_{-i}}(m_{-i})} & \frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x_{a'_{-i}}(m'_{-i})} \end{bmatrix}.$$

Proposition 1

$$J(x) = \begin{bmatrix} \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a_i}(m_i)} & 0 & \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a_{-i}}(m_{-i})} & \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a'_{-i}}(m'_{-i})} \\ 0 & \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a'_i}(m'_i)} & \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a_{-i}}(m_{-i})} & \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a'_{-i}}(m'_{-i})} \\ \frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x_{a_i}(m_i)} & \frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x_{a'_i}(m'_i)} & \frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x_{a_{-i}}(m_{-i})} & 0 \\ \frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x_{a_i}(m_i)} & \frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x_{a'_i}(m'_i)} & 0 & \frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x_{a'_{-i}}(m'_{-i})} \end{bmatrix}.$$

Proposition 1

$$J(x) = \begin{bmatrix} \frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x_{a_i}(m_i)} & 0 & 0 & 0 \\ 0 & \frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a'_i}(m'_i)} & 0 & 0 \\ 0 & 0 & \frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x_{a_{-i}}(m_{-i})} & 0 \\ 0 & 0 & 0 & \frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x_{a'_{-i}}(m'_{-i})} \end{bmatrix}$$

Proposition 1

We are supposing $\alpha = 0$. Then the self-interaction terms are:

- $\frac{\partial \dot{x}^t_{a_i}(m_i)}{\partial x^t_{a_i}(m_i)} = \beta\eta(m_i)(1 - 2x^t_{a_i}(m_i))[u^t_i(a_i|m_i) - u^t_i(a'_i|m_i)]$
- $\frac{\partial \dot{x}^t_{a'_i}(m'_i)}{\partial x_{a'_i}(m'_i)} = \beta\eta(m'_i)(1 - 2x^t_{a'_i}(m'_i))[u^t_i(a'_i|m'_i) - u^t_i(a_i|m'_i)]$
- $\frac{\partial \dot{x}^t_{a_{-i}}(m_{-i})}{\partial x_{a_{-i}}(m_{-i})} = \beta\eta(m_{-i})(1-2x^t_{a_{-i}}(m_{-i}))[u^t_{-i}(a_{-i}|m_{-i})-u^t_{-i}(a'_{-i}|m_{-i})]$
- $\frac{\partial \dot{x}^t_{a'_{-i}}(m'_{-i})}{\partial x_{a'_{-i}}(m'_{-i})} = \beta\eta(m'_{-i})(1-2x^t_{a'_{-i}}(m'_{-i}))[u^t_{-i}(a'_{-i}|m'_{-i})-u^t_{-i}(a_{-i}|m'_{-i})]$

Proposition 1

Correlated equilibrium: $\forall i \in N, \forall a_i, a_i' \in A_i, \forall m_i \in M_i$:

$$\eta(m_i) x^t_{a_i}(m_i)[u^t_i(a_i \mid m_i) - u^t_i(a_i' \mid m_i)] \geq 0.$$

Proposition 1

Proposition 2

The fixed-point condition in Lemma 2 can be expressed as:

$$\beta \left[u_i^t(a_i|m_i) - \sum_{a_i' \in A_i} x_{a_i'}^t(m_i) u_i^t(a_i'|m_i) \right] = \alpha \left[\ln(x_{a_i}^t(m_i)) - \sum_{a_i' \in A_i} x_{a_i'}^t(m_i) \ln \right.$$

Since we restrict our attention to 2×2 games, we have that

$$\beta \left[(1 - x_{a_i}^t(m_i)) u_i^t(a_i|m_i) - (1 - x_{a_i}^t(m_i)) u_i^t(a_i'|m_i) \right] = \alpha \left[(1 - x_{a_i}^t(m_i)) \ln \right.$$

$$\beta \left[u_i^t(a_i|m_i) - u_i^t(a_i'|m_i) \right] = \alpha \left[\ln(x_{a_i}^t(m_i)) - \ln(x_{a_i'}^t(m_i)) \right]$$

by the properties of logarithm, we have

$$\frac{\beta}{\alpha}\left[u_i^t(a_i|m_i) - u_i^t(a_i'|m_i)\right] = \left[\ln\left(\frac{x_{a_i}^t(m_i)}{x_{a_i'}^t(m_i)}\right)\right].$$

Since $1 - x_{a_i}^t(m_i) = x_{a_i'}^t(m_i)$, it follows that

$$\frac{\beta}{\alpha}\left[u_i^t(a_i|m_i) - u_i^t(a_i'|m_i)\right] = \left[\ln\left(\frac{x_{a_i}^t(m_i)}{1 - x_{a_i}^t(m_i)}\right)\right]$$

$$\left(\frac{x_{a_i}^t(m_i)}{1 - x_{a_i}^t(m_i)}\right) = e^{\frac{\beta}{\alpha}\left[u_i^t(a_i|m_i) - u_i^t(a_i'|m_i)\right]}$$

$$\left(x_{a_i}^t(m_i)\right) = e^{\frac{\beta}{\alpha}\left[u_i^t(a_i|m_i) - u_i^t(a_i'|m_i)\right]}\left(1 - x_{a_i}^t(m_i)\right)$$

$$\left(x^t_{a_i}(m_i)\right)\left(1 + e^{\frac{\beta}{\alpha}\left[u^t_i(a_i|m_i) - u^t_i(a'_i|m_i)\right]}\right) = e^{\frac{\beta}{\alpha}\left[u^t_i(a_i|m_i) - u^t_i(a'_i|m_i)\right]}$$

$$x^t_{a_i}(m_i) = \frac{e^{\frac{\beta}{\alpha}\left[u^t_i(a_i|m_i) - u^t_i(a'_i|m_i)\right]}}{1 + e^{\frac{\beta}{\alpha}\left[u^t_i(a_i|m_i) - u^t_i(a'_i|m_i)\right]}}$$

Multiplying the right-hand-side numerator and denominator by the same factor, we have

$$x^t_{a_i}(m_i) = \frac{e^{\frac{\beta}{\alpha}\left[u^t_i(a_i|m_i)\right]}}{e^{\frac{\beta}{\alpha}[u^t_i(a'_i|m_i)]} + e^{\frac{\beta}{\alpha}\left[u^t_i(a_i|m_i)\right]}},$$

which is the formula for S-QCE when $q_i(z) = \exp(\frac{\beta}{\alpha} z)$.