Phase Transitions in Quantum Games

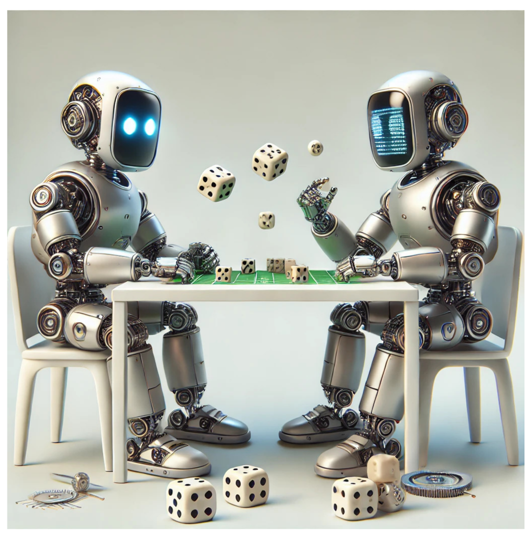
Giovanni Cemin[1], Markus Schmitt[2,3], Marin Bukov[1]

[1] Max Planck Institute for the Physics of Complex Systems, Dresden, Germany
[2] University of Regensburg, Regensburg, Germany
[3] Forschungszentrum Jülich, Institute of Quantum Control, Jülich, Germany

Contact me: gcemin@pks.mpg.de

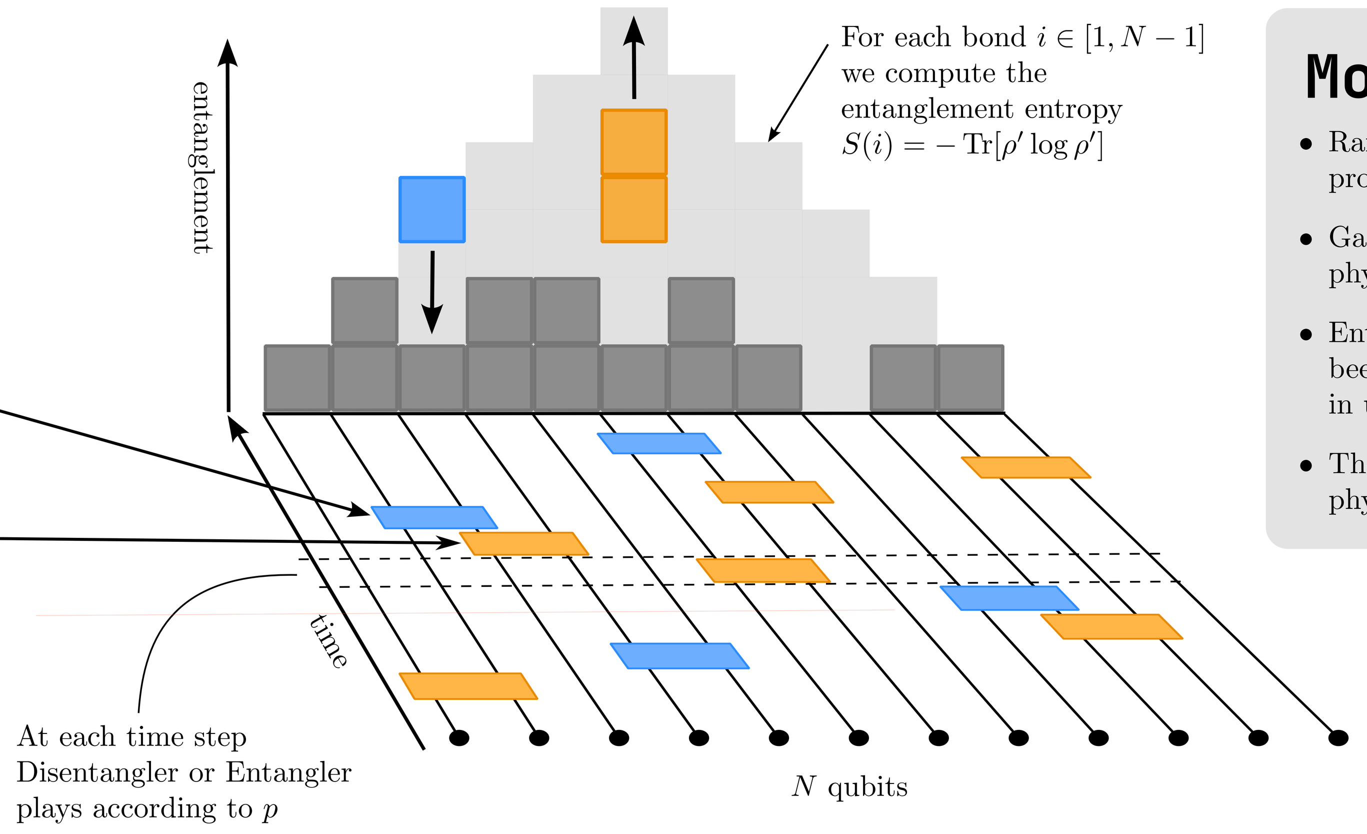
Game set-up

Two agents playing against each other [1]



Entangler
plays with probability $1-p$
1) selects randomly a bond $i \in [1, N-1]$,
2) samples a Clifford gate $g \in \mathcal{C}_2$,
3) acts with g on ith bond.

Disentangler
plays with probability p
1) selects a bond $i \in [1, N-1]$ with *policy*,
2) chooses optimally disentangling Clifford gate g^*,
3) acts with g^* on ith bond.
policy $\in$ { random, greedy, RL }



Motivation

- Random quantum circuits display interesting and universal behaviours, providing a simple tool to probe complex quantum dynamics
- Game-like settings are highly tunable setups that reveal new and intriguing physical phenomena
- Entanglement transitions, from area-law to volume-law steady states, have been observed in both mesurement-doped circuits[6], but most recently also in unitary gates circuits[1]
- The degree and nature of accessible information significantly impact the physics, allowing us to modify entanglement transitions

Clifford circuits in a nutshell

Clifford group $\mathcal{C}_n$: the group of unitaries that normalize the Pauli group (n = # qubits)

$\mathcal{C}_n := \{c \in U(2^n) \mid c\mathcal{P}c^\dagger = \mathcal{P}_n\}$

Pauli group $\mathcal{P}_n := \{\alpha O_1 O_2 \dots O_N\}$, with $\alpha \in \{1, -1, i, -i\}, O_j \in \{I, X, Y, Z\}$

The elements of $\mathcal{C}_n$ are called **Clifford gates**, and are generated by *Hadamard* (H), *Phase* (S) and CNOT gates.

$$\mathrm{H} = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix} \quad \mathrm{S} = \begin{pmatrix}1 & 0\\ 0 & i\end{pmatrix} \quad \mathrm{CNOT} = \begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&0&1\\0&0&1&0\end{pmatrix}$$

Why Clifford circuits?[2,3]
The *Gottesman-Knill theorem* ensures that any Clifford circuit can be simulated on a classical machine in time *poly*(*n*).
→ Can simulate larger system sizes and get closer to the thermodynamic limit

idea: describe state $|\psi\rangle$ with its *stabilizers* instead of the amplitudes.
unitary operator s.t. $s_i|\psi\rangle = |\psi\rangle$
→ stabilizer group Stab($|\psi\rangle$)

$$\text{Tableau } \mathcal{T} = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_N \end{bmatrix} \Rightarrow \left(\begin{array}{ccc|ccc} x_{11} & \dots & x_{1N} & z_{11} & \dots & z_{1N} \\ x_{21} & \dots & x_{2N} & z_{21} & \dots & z_{2N} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_{N1} & \dots & x_{NN} & z_{N1} & \dots & z_{NN} \end{array}\right)$$

since Stab($|\psi\rangle$) $\subset \mathcal{P}$ can decompose each s_i on $\{X_1, \dots, X_N, Z_1, \dots, Z_N\}$

Example
State: $|\psi\rangle = |11\rangle$
stabilized by II, ZI, IZ, ZZ

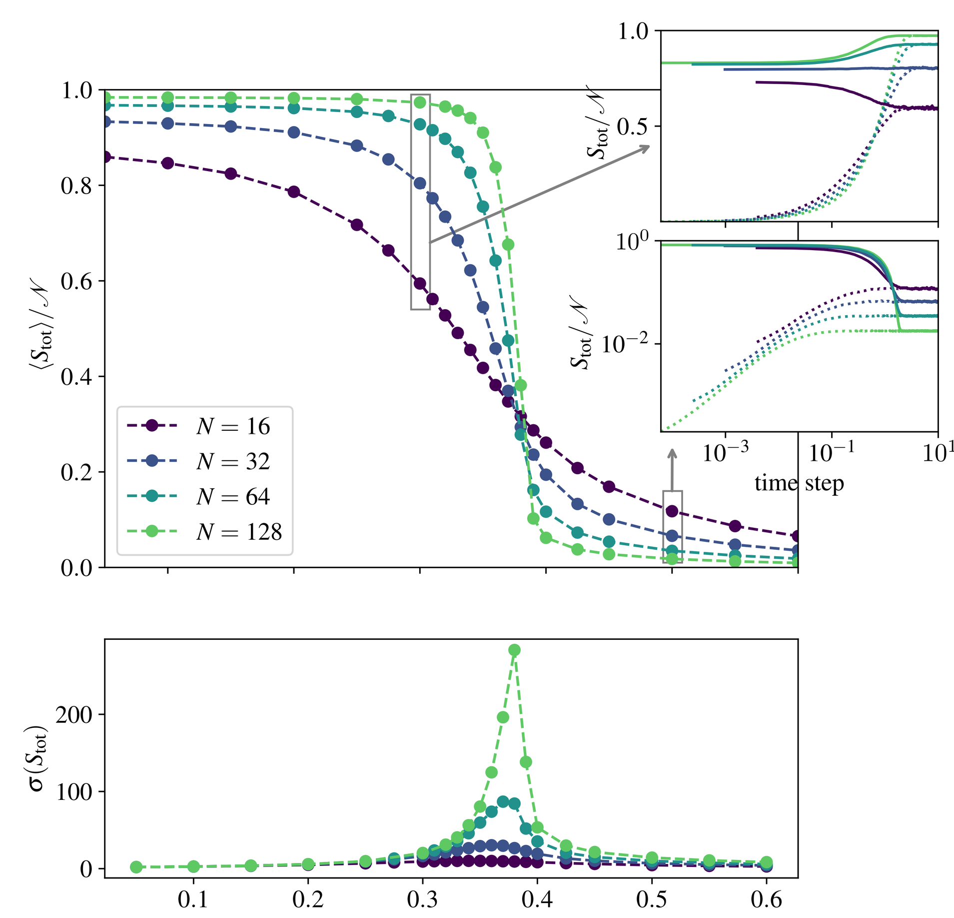
$\Rightarrow$ tableau $\left(\begin{array}{cc|cc} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right)$

Apply H: $X \to Z, Z \to X$

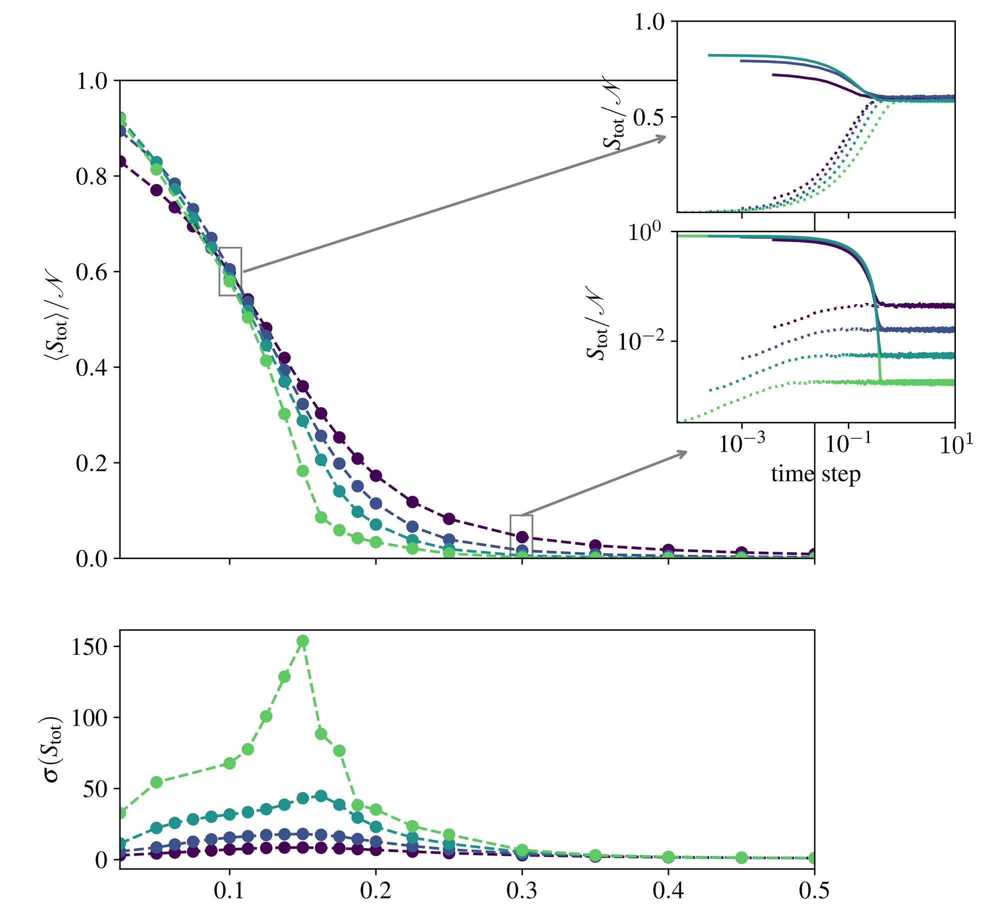
$\Rightarrow$ tableau $\left(\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{array}\right)$

Entanglement: $S(\mathrm{A}) = |\mathrm{A}| - \log_2 |\mathcal{S}_\mathrm{A}|$ for bipartition A, B and $\mathcal{S}_\mathrm{A}$ group of stabilizers acting on A
The subgroup $\mathcal{S}_\mathrm{A}$ has period two, and therefore $\log_2 |\mathcal{S}_\mathrm{A}| \in \mathbb{Z}^{0+}$

zero information vs. complete information



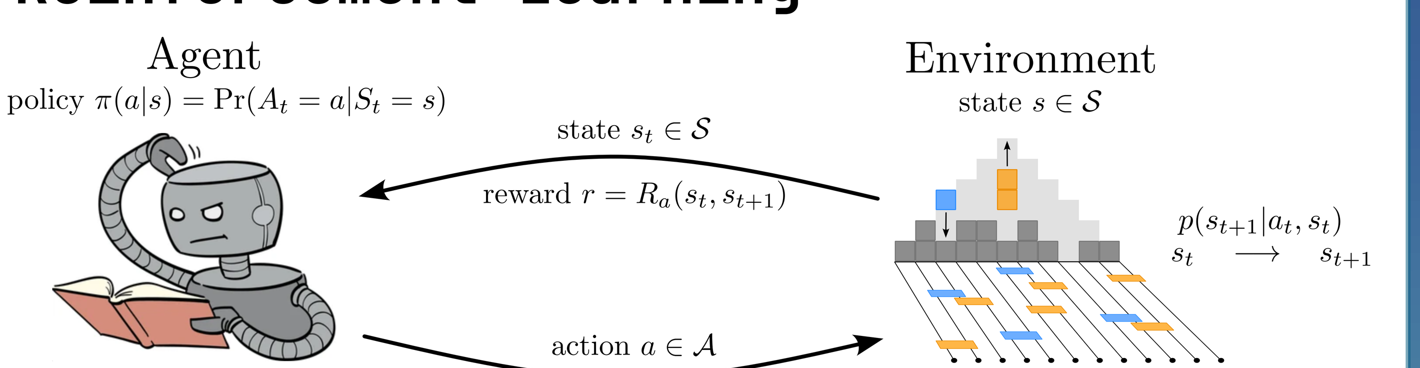
policy: select bond i randomly
→ dicontinuous phase transition between a volume law and an area law phase
→ divergence of fluctuations at the critical point suggests $p_c \approx 0.38$



policy: select bond i greedily – bond that maximally disentangles across all i
→ continuous phase transition
→ divergence of fluctuations at the critical point suggests $p_c \approx 0.15$
→ entanglement has a value < 1 for $p \neq 0$

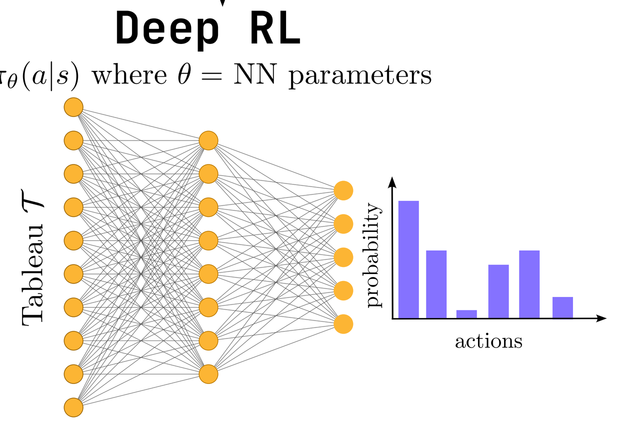
The amount of information available to the Disentanger changes the dynamical properties of the system!

Reinforcement Learning



Deep RL

$\pi_\theta(a|s)$ where θ = NN parameters



state space $\mathcal{S} = \mathcal{T}$
action space $\mathcal{A}$ = bonds
reward $r = -\sum_i^{N-1} S(i)$

Proximal Policy Optimization[4] (PPO)

OpenAI Details

PPO is a **stable, efficient, on-policy, actor-critic algorithm**.

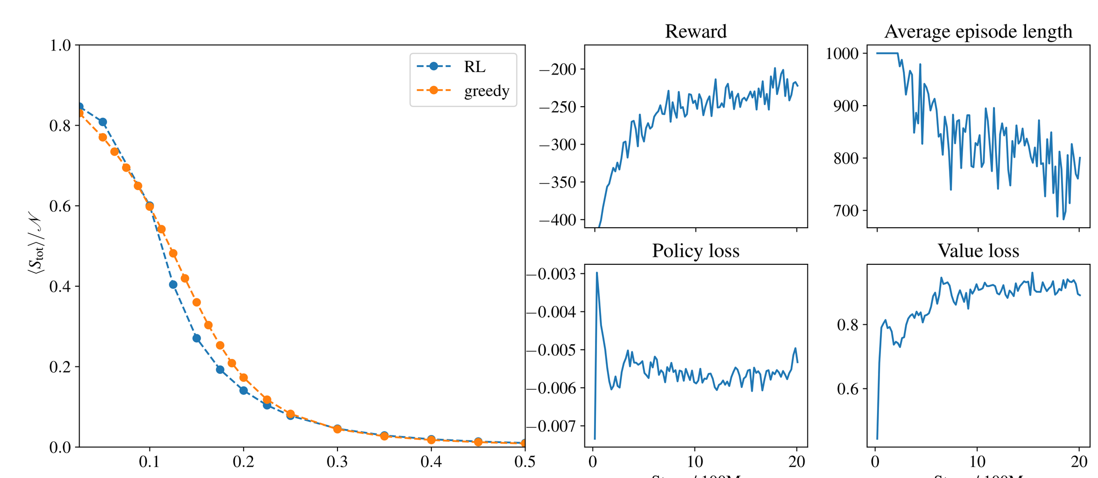
1. **Data collection**
Unroll policy $\pi_\theta(a|s)$ for $t = 0, .., T-1$
collect **trajectories** $\tau = \{(s, a, r)\}$

2. **Compute Rewards and Advantages**
$R_t = \sum_{k=0}^{T} \gamma^k r_{t+k}$
$A_t = \sum_{k=0}^{T} (\gamma\lambda)^k (r_t + \gamma V(s_{t+1}) - V(s_t))$

3. **Policy Update**
$L(\theta) = \mathbb{E}_t[\mathrm{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)A_t]$
where $r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{\mathrm{old}}}(a_t|s_t)}$
Update θ (i.e. policy) minimizing $L(\theta)$

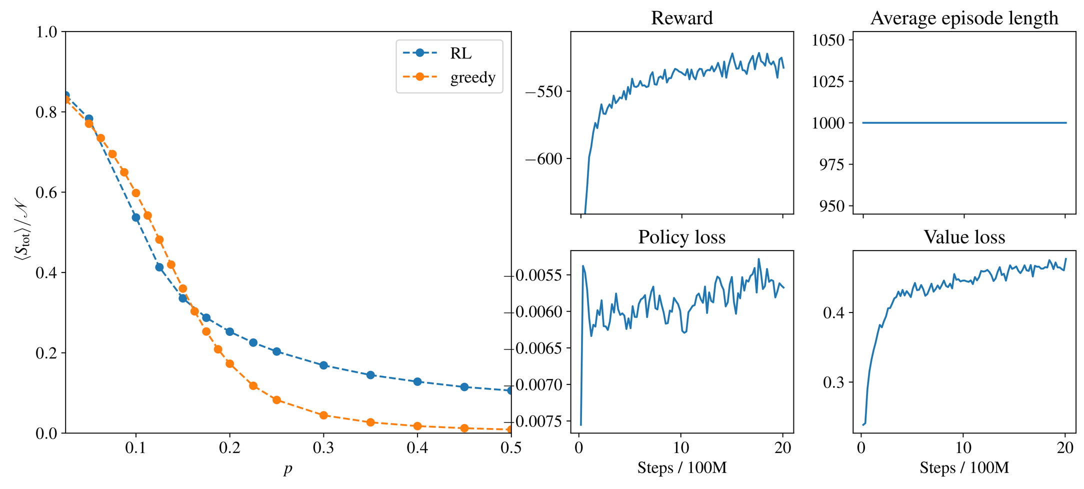
4. **Value Function Update**
$L^{\mathrm{value}}(\phi) = \mathbb{E}_t[(R_t - V_\phi(s_t))^2]$
Update ϕ minimizing $L^{\mathrm{value}}(\phi)$

RL application



Trained at $p = 0.15 \approx p_c$:

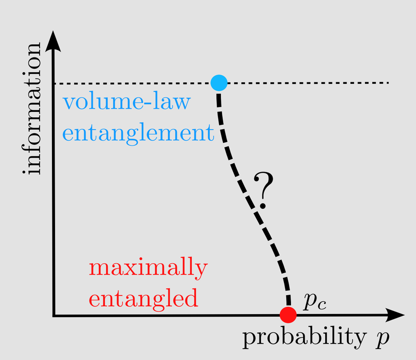
- The RL algorihtm converges at a *different* strategy
- The RL policy outperforms the greedy policy for $p > p^{\mathrm{critical}}$



Trained at $p = 0.1 < p_c$:

- The discrepancy between RL and greedy policy is smaller for $p < p_c$
- Suggests greedy policy is optimal for $p < p_c$

- The greedy strategy is close (but not equal) to the optimal strategy.
- The dynamics of the system is similar in both cases: RL and greedy.
- What is the optimal policy with less information?
less information means not all rows of $\mathcal{T}$ are available.
There is no straightforward way of generalizing the greedy strategy
→ RL is the right tool for this
- Is complete information necessary to choose optimal local actions?
Is there a minimum amount of information s.t. an optimal action can be taken?



We expect two possible scenarios:

- jump from greedy-like to random-like
- continous transformation from greedy-like to random-like

References

[1] R. Morral-Yepes, A. Smith, S. L. Sondhi, F. Pollmann, PRX Quantum 5, 010309 (2024)
[2] D. Gottesman, Physical Review A 57.1 (1998)
[3] S. Aaronson, D. Gottesman, Physical Review A 70.5 (2004)
[4] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov, arXiv:1707.06347 (2017)
[5] A. Nahum, J. Ruhman, S. Vijay, J. Haah, Physical Review X 7.3 (2017)
[6] Y. Li, X. Chen, M. P. A. Fisher, Physical Review B 100.13 (2019)

Outlook

- Characterize the dynamical properties at critical point
- Finite size scaling to study behaviour in thermodinamic limit
- Use the RL model to extrapolate between complete and zero information, aiming to study how the phase transition changes

Phase Transitions in Quantum Games

mpipks

Two agents against each other, each following a *policy*:

random, greedy, RL

$\rightarrow$ who will win, and why? Find out at my poster :)