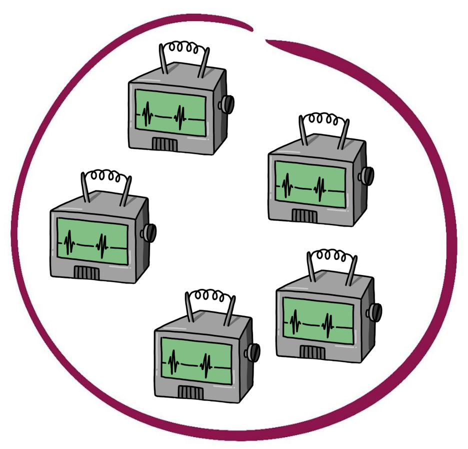
PICK YOUR BATTLES:
INTERACTION GRAPHS AS POPULATION-LEVEL OBJECTIVES FOR STRATEGIC DIVERSITY

Marta Garnelo, Wojciech Marian Czarnecki, Siqi Liu, Dhruva Tirumala, Junhyuk Oh, Gauthier Gidel, Hado van Hasselt, David Balduzzi

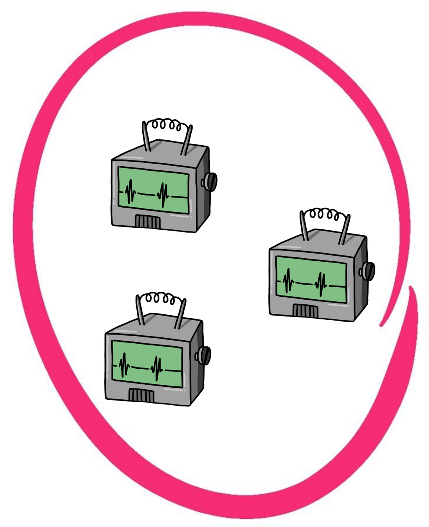
MOTIVATION

Strategic diversity of populations is important on three levels of agent training:



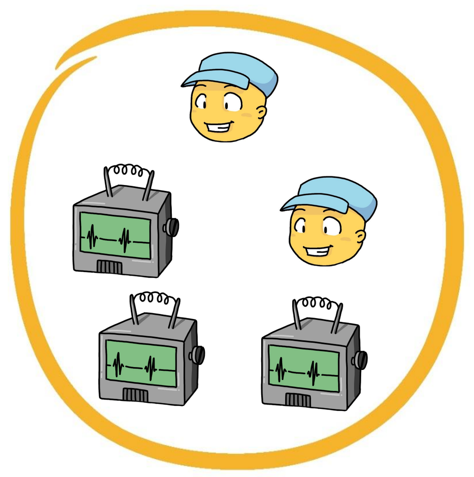
TRAINER POPULATION

Better set of training adversaries



LEARNER POPULATION

Better performing final population

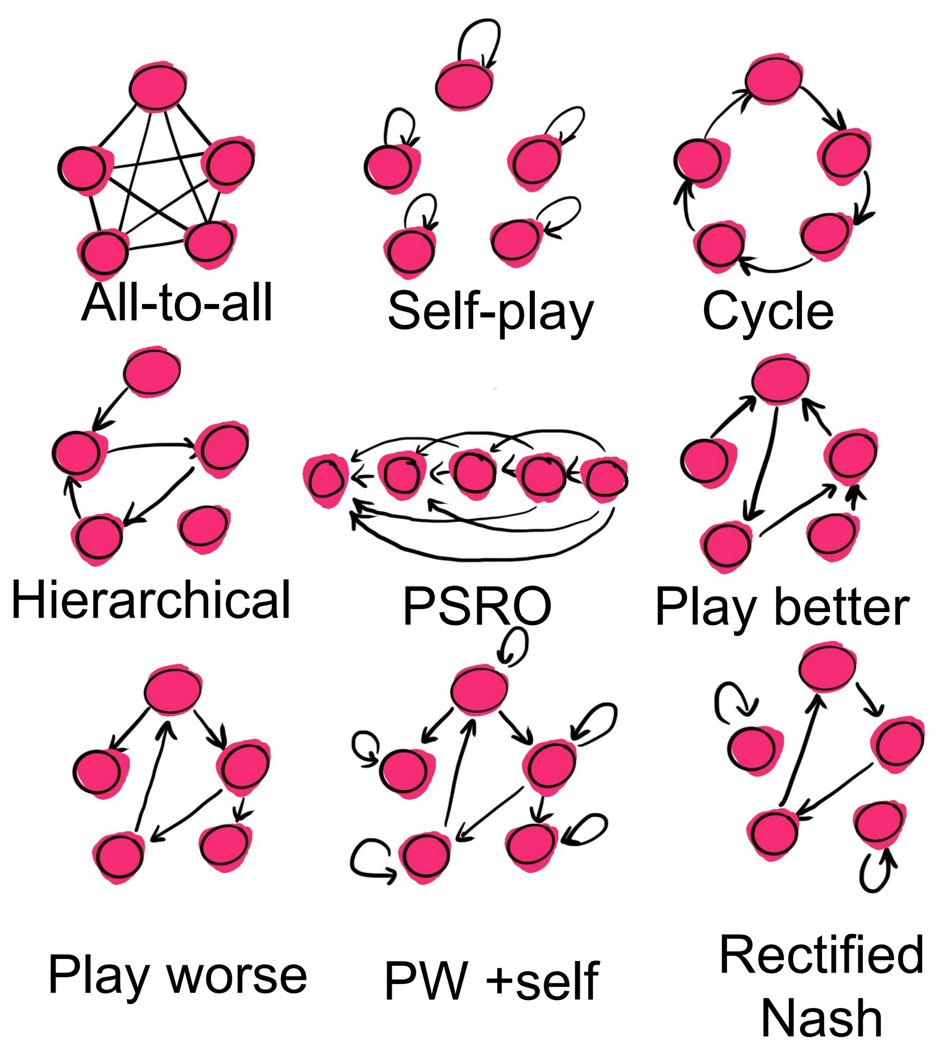


EVALUATOR POPULATION

Better performance estimation

How can we train diverse populations of agents from scratch (i.e. trainer population = learner population)?
→ **Don't train everyone in the population against everyone else.**

How can we express this match-making?
→ **Interaction graphs**

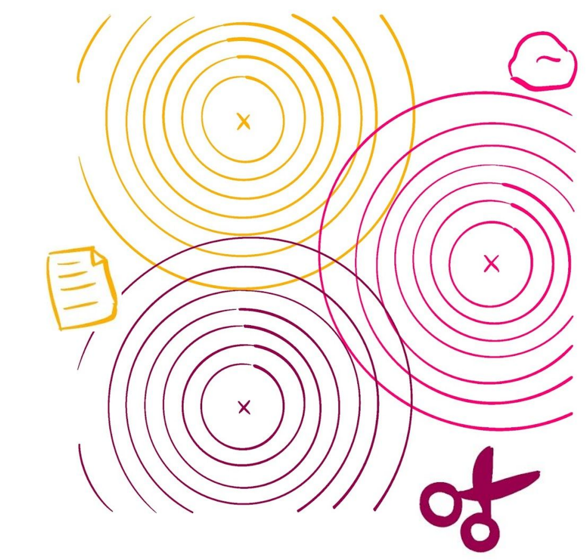


Fixed interaction graphs are defined at the beginning and remain fixed (e.g. self-play).

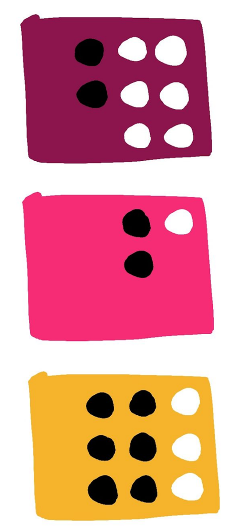
Adaptive interaction graphs have edges that are repeatedly updated as a function of their relative performance (e.g. play worse).

ENVIRONMENTS

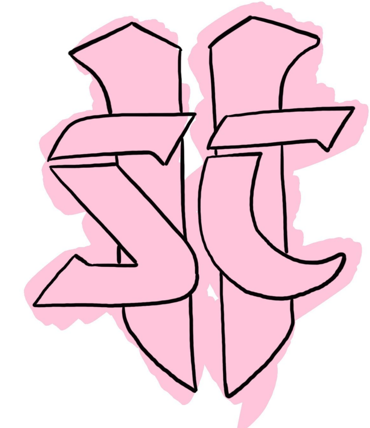
GMM ROCK-PAPER-SCISSORS



BLOTTO

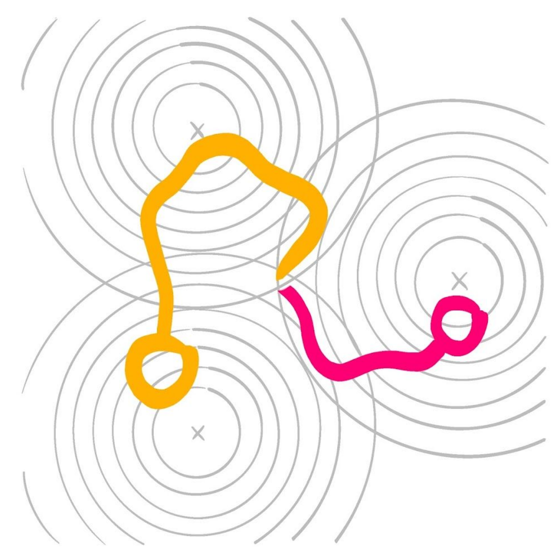


STARCRAFT II

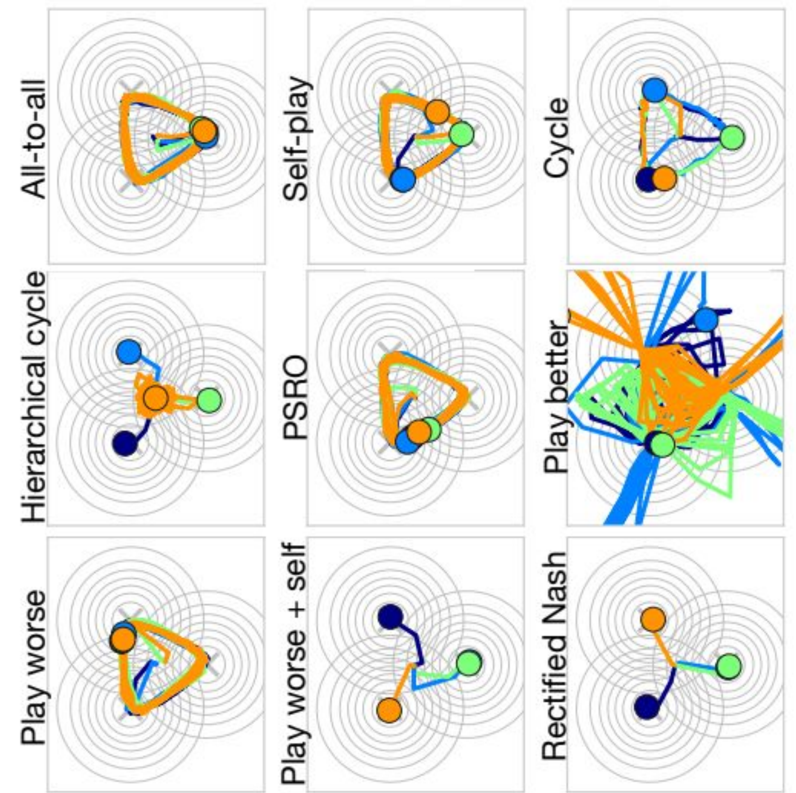


MAIN FINDINGS

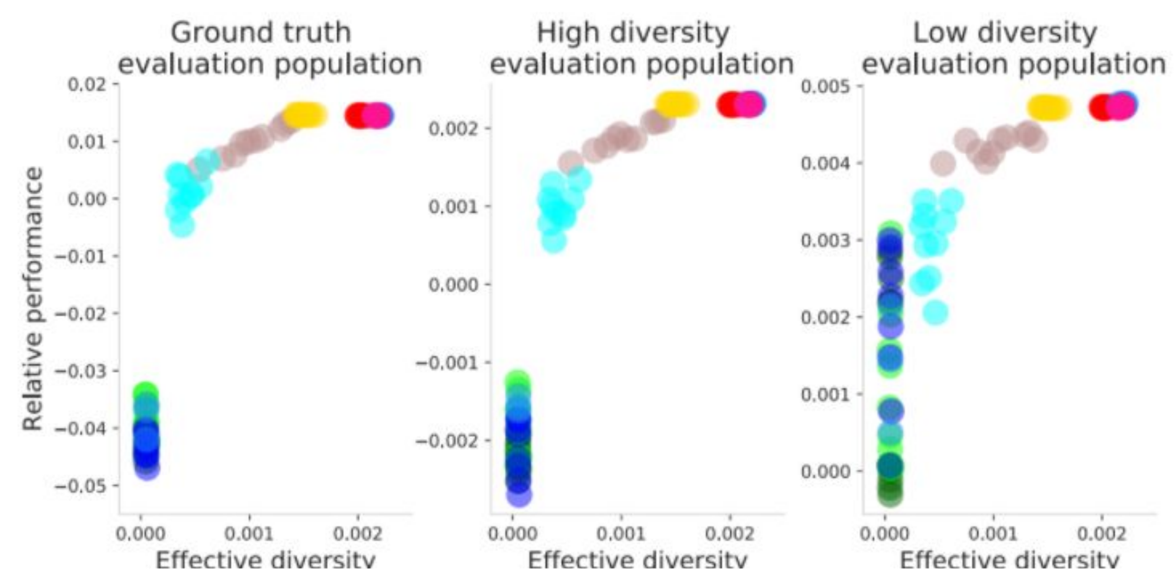
The type of graph has an effect on the training dynamics and the agents in a population. Graphs with cycles encourage specialisation.



Training trajectories in strategy space over time.

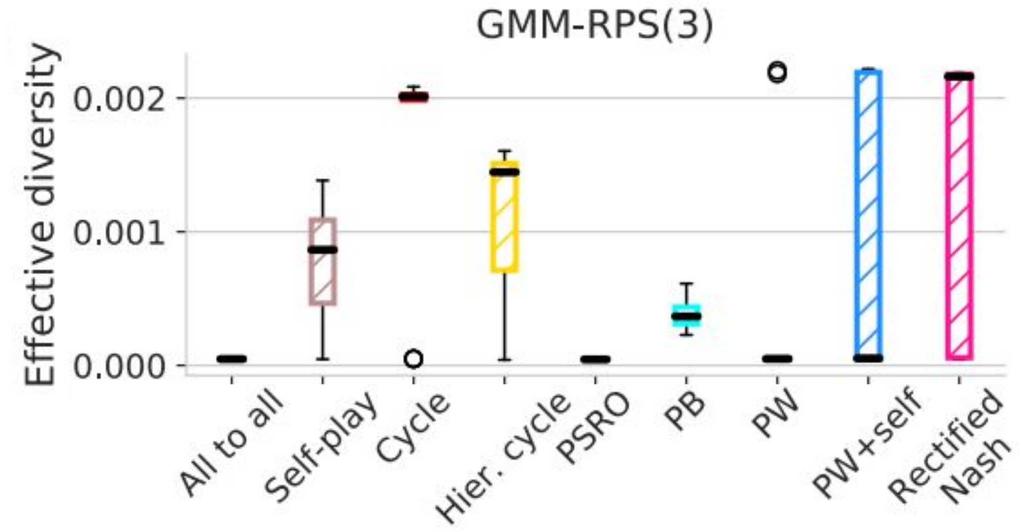


Diverse **evaluator** populations estimate performance better and diverse **learner** populations perform better.

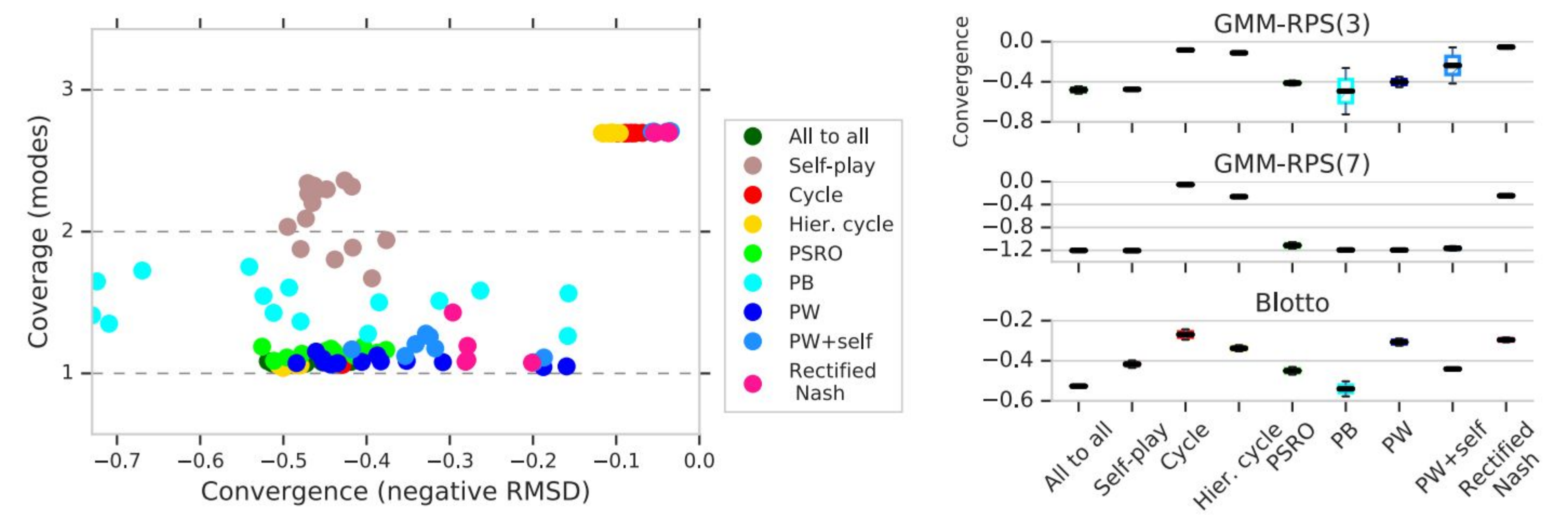


Focusing on agents that are **better than you** makes you less exploitable and focusing on those that are **worse than you** makes you a better exploiter.

The **type of graph** has an effect on the diversity and he agents in a population. Graphs with **cycles** encourage diversity.

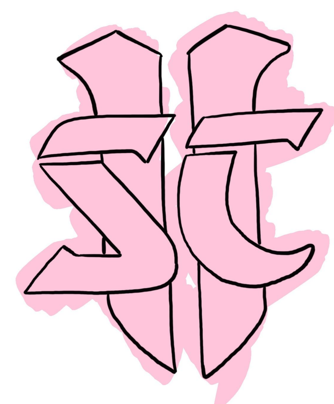


Individual **convergence** is not as important as population level convergence for diversity and coverage.



When moving to significantly **more complex environments** some fundamental insights hold, but some do not.

Method	Performance	Coverage	Diversity
All-to-all	47%	89%	39%
Hier. Cycle	46%	74%	29%
Cycle	30%	74%	12%
Rect. Nash	43%	63%	3%
Self-play	44%	53%	0%



CONCLUSION

- In multi-agent training there is **not a fixed objective.**
- It is necessary to **design algorithms** to search the strategy space effectively.
- We explore **population-level** learning algorithms where objectives of agents are specified by interaction graphs.