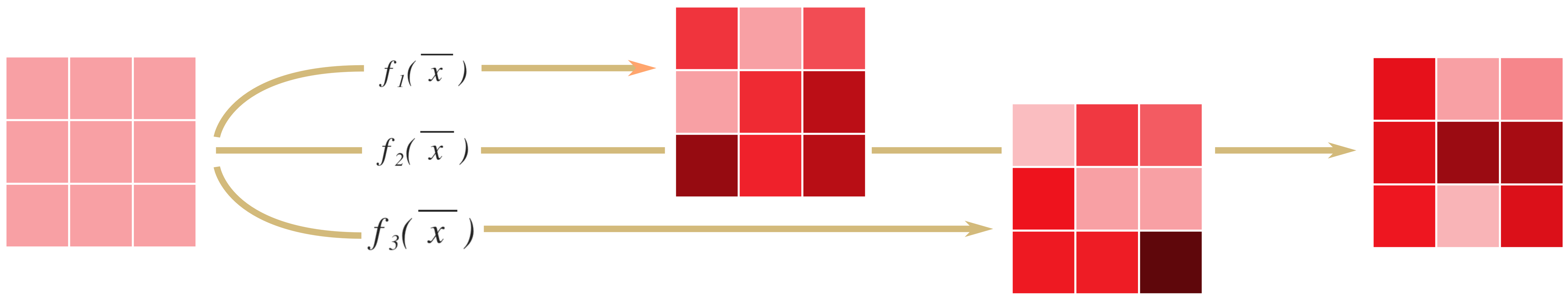



INDIAN INSTITUTE OF SCIENCE
भारतीय विज्ञान संस्थान

# UTILITY FUNCTIONS WITH COMPOUNDING RETURNS LEAD TO THE EVOLUTION OF COOPERATIVITY IN MULTI-ARMED BANDIT NETWORKS


*Preetham Venkatesh, Bhaskar Kumawat*
*preethamv@iisc.ac.in, kbhaskar@iisc.ac.in*
*Undergraduate Department, Indian Institute of Science - 560012*


## ABSTRACT


The transition to multi-cellular ensembles from single cells is one of the most interesting problems in contemporary evolutionary biology. Modeling organisms as multi-armed bandit (MAB) networks that can perform a fixed number of tasks, we have developed biologically inspired utility functions that lead to sharing of information when social utility of the cluster is optimised. The multi-armed bandit problem in algorithmic game theory deals with the exploration-exploitation tradeoff in context of allocating a limited set of resources, ie, the genome size in our system. Current solutions algorithms to the MAB problem with infinite arms have several assumptions which fail when applied to organisms that can share information, and thus optimization of the network has been achieved using a genetic algorithm. We have also studied the effects of various parameters that the organisms can modulate, such as export efficiency and the value of reward for completing low complexity tasks, as well as externally determined factors such as number of cells and genomic size on the evolution of cooperation. In some cases, interesting patterns emerge - such as appearance of organisms that specialize to perform a particular task or cooperate to a larger extent - a phenomena akin to differentiation in real biological systems. These results have a natural parallel in evolution, and can also be extended to other problems such as long-term stability of organizations and administrative structures.


## MOTIVATION

Cells in a multi-cellular organisms[2][4]

People in a workplace

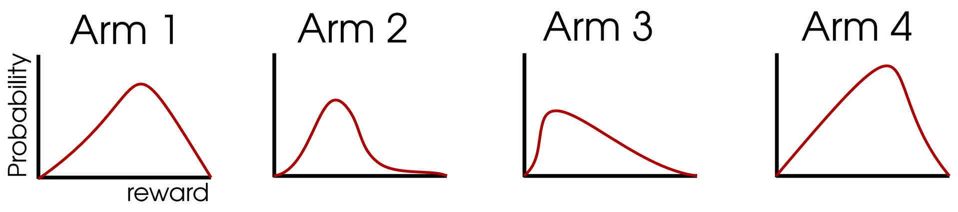
Need to share information and co-operate for proper working of the organisation/organism

How to decide the salary/rewards for individual agents to ensure this?

## MULTI-ARMED BANDITS

Exploration vs Exploitation

**Classically[1],**



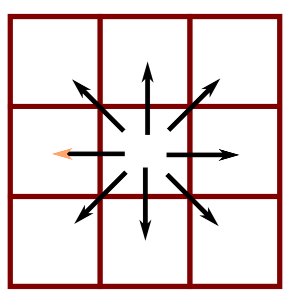

**In our case,**

For each agent if,
Work capacity = 6
Possible tasks = 10

$10^6$ possible task vectors[5]

Uncertainity provided by other agents, not by intrinsic stochasticity

## FRAMEWORK



Players in a grid (periodic boundaries):
- Limited computational capacity.
- Perform tasks to gain resources.
- Share resources with neighbors.

Doing task i —Fetches→ resources of type i

Player "Genome" —Execution generates→ Resource vector

Resource vector —MECHANISM→ Utility

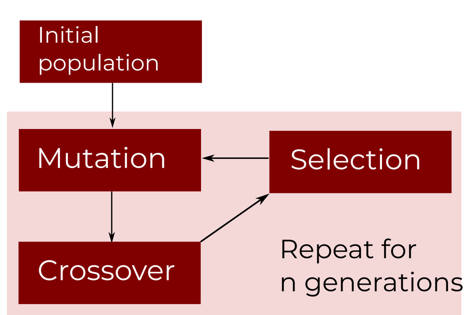
## DEFINITIONS

**Utility function (UF):** A function that returns a utility for a given information vector.
**Social welfare:** Sum total of all individual utilities in the network.
**Co-operativity:** Fraction of genome dedicated to sharing (as opposed to doing tasks) for each agent.
**Export efficacy:** Fraction of information about a task distributed to neighbors on sharing.
**Insight:** Marginal increase in reward for a higher level task on completion of a lower level task.

## LEARNING

Genetic Algorithm[3] with following parameters:

Population size: 100
Mutation probability: 0.2
Crossover probability: 0.1
Generations: 3000
To maximize Social Welfare




## UTILITY FUNCTIONS

**UF-0** (Non-compunding returns)

$$\sum_{i=1}^{nTasks} info(i)$$

```
info = [i1,i2,i3,i4...]
Sum { i1
      i2
      i3
      i4
```

**UF-1** (Unweighted compunding returns)

$$\sum_{i=1}^{nTasks} \{info(i) + I\cdot \sum_{j=1}^{i-1} sgn[info(j)]\}$$

```
Sum { i1
      i2+I*[sgn(i1)]
      i3+I*[sgn(i2)+sgn(i1)]
      i4+I*[sgn(i3)+sgn(i2)+sgn(i1)]
```

**UF-2** (Weighted compunding returns)

$$\sum_{i=1}^{nTasks} \{info(i) + I\cdot sgn[info(i)]\cdot \sum_{j=1}^{i-1} [info(j)]\}$$

```
Sum { i1
      i2+I*[info(i1)]
      i3+I*[info(i2)+info(i1)]
      i4+I*[info(i3)+info(i2)+info(i1)]
```

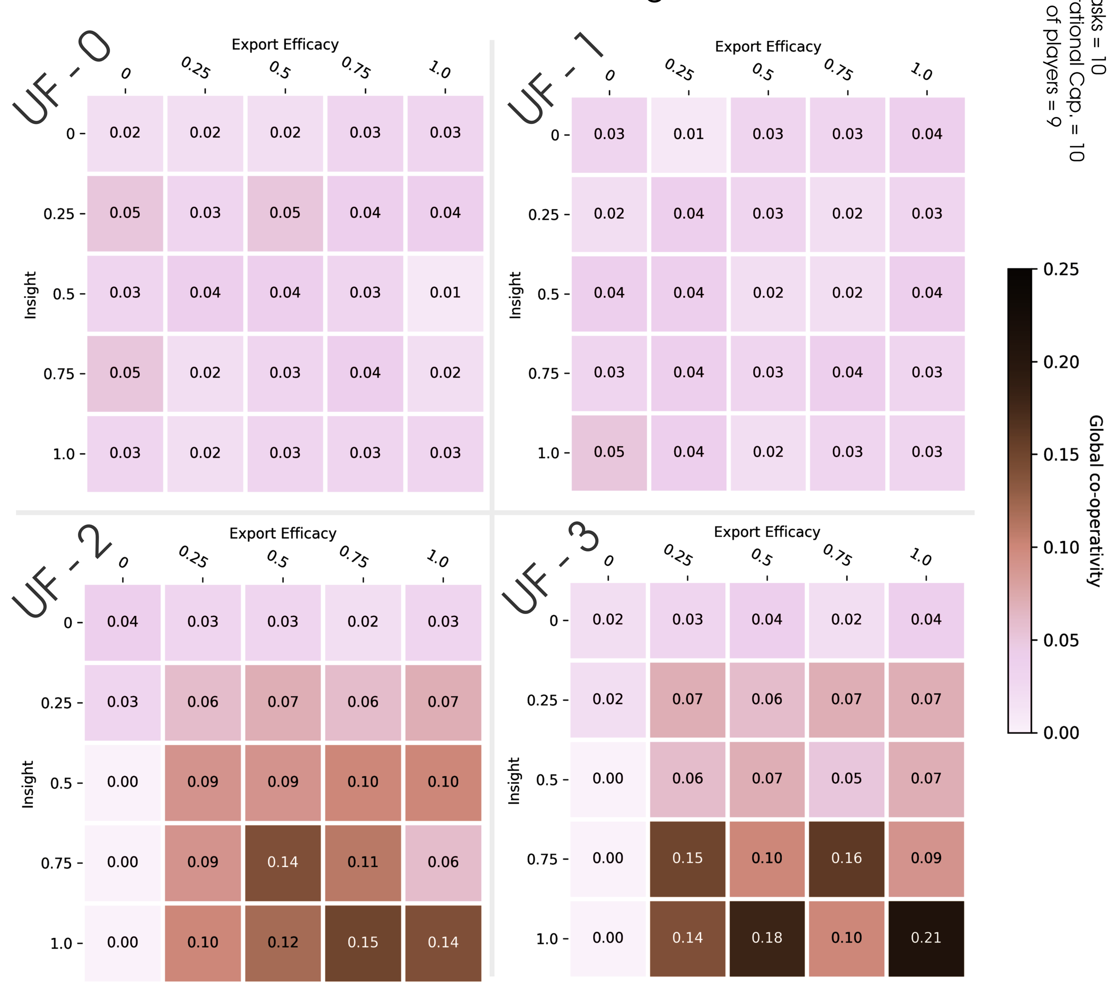
**UF-3** (Polynomial compounding returns)

$$\sum_{i=1}^{nTasks} \{info(i) + \sum_{j=1}^{i-1} sgn[info(j)] \sum_{k=1}^{j} I^{i-k}\, info(k)\}$$

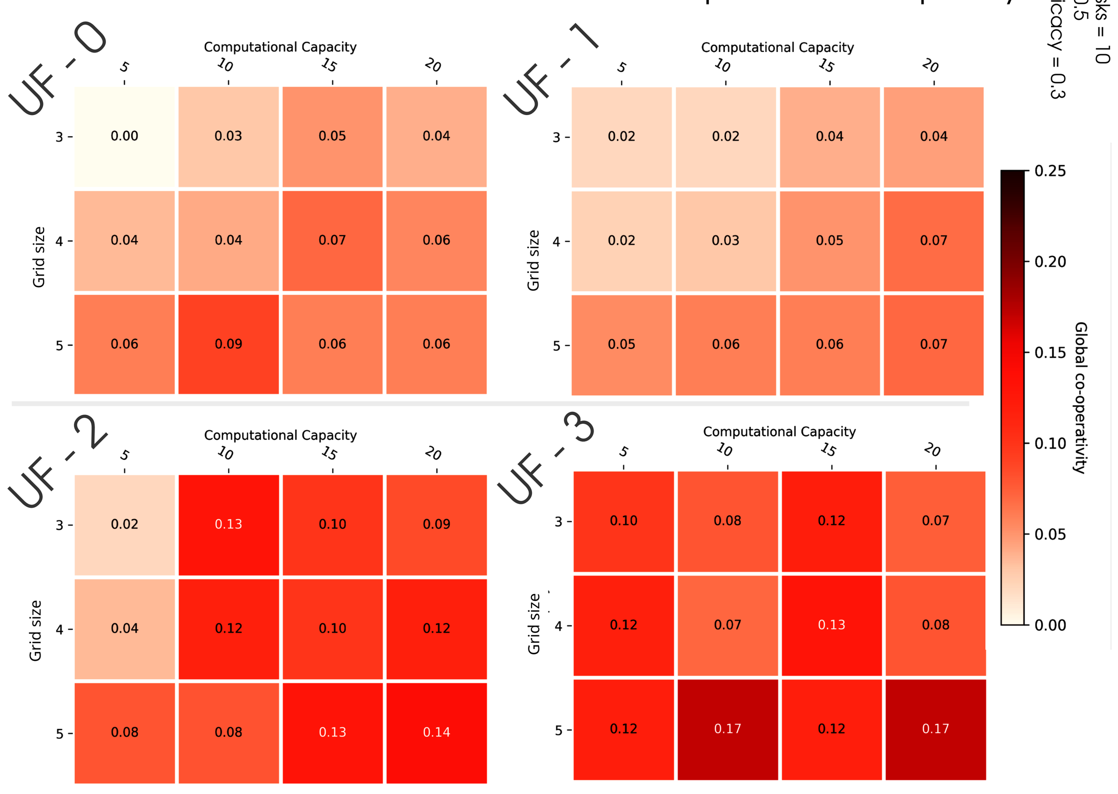
```
Sum { i1
      i2+I*info(i1)
      i3+I*info(i2)+I^2*info(i1)+I^2*info(i1)
      ...[1]
```

## RESULTS

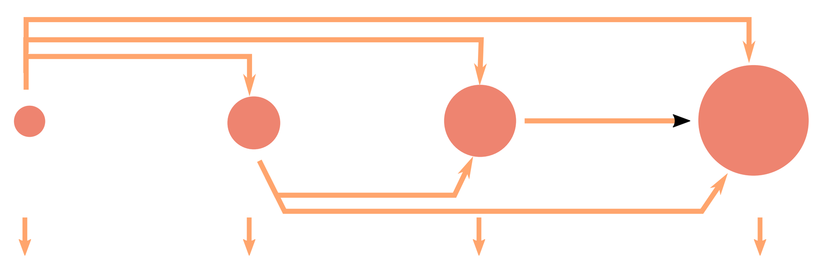
Variation in soft parameters → Export Efficacy, Insight




Variation in hard parameters → Number of players, Computational capacity




The term "compounding returns" lends to the fact that both of the utility functions 2 and 3 are designed such that the benefit from a higher complexity task is maximized when all the lower level tasks are also being performed at least once.




These simulations optimize social welfare of the group using a genetic algorithm but do not maintain spatial connections as organisms are place randomly at every iteration. They also do not implement independent optimization by the agents, a behaviour similar to multi-cellular organism.

EvoAgentNet

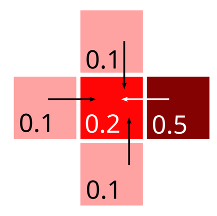
EvoAgentNet is a C++ program developed by us to simulate similar multi-armed bandit dynamics but in a network where spatial localisation is important and evolution and selection occurs in a more conventional single-cell fashion. The spatial reproductive processes ensure retention of inter-agent links over evolutionary time.

## FRAMEWORK

### Algorithm

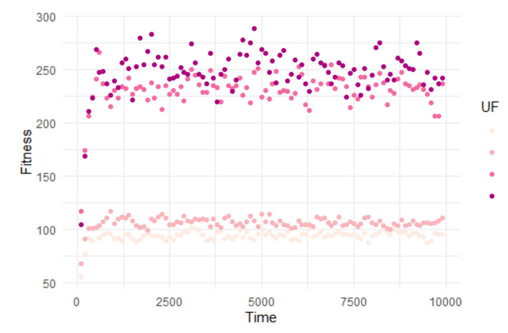
1. Calculate resource vectors for all agents and corresponding utility (fitness) after task execution and sharing. Here, sharing is limited to the four nearest neighbors.


2. For each agent in the network, compare fitness with four nearest neighbors.
3. With a probability weighted by the four neighbor's and self fitness, replace this agent with either of the five (with mutations).



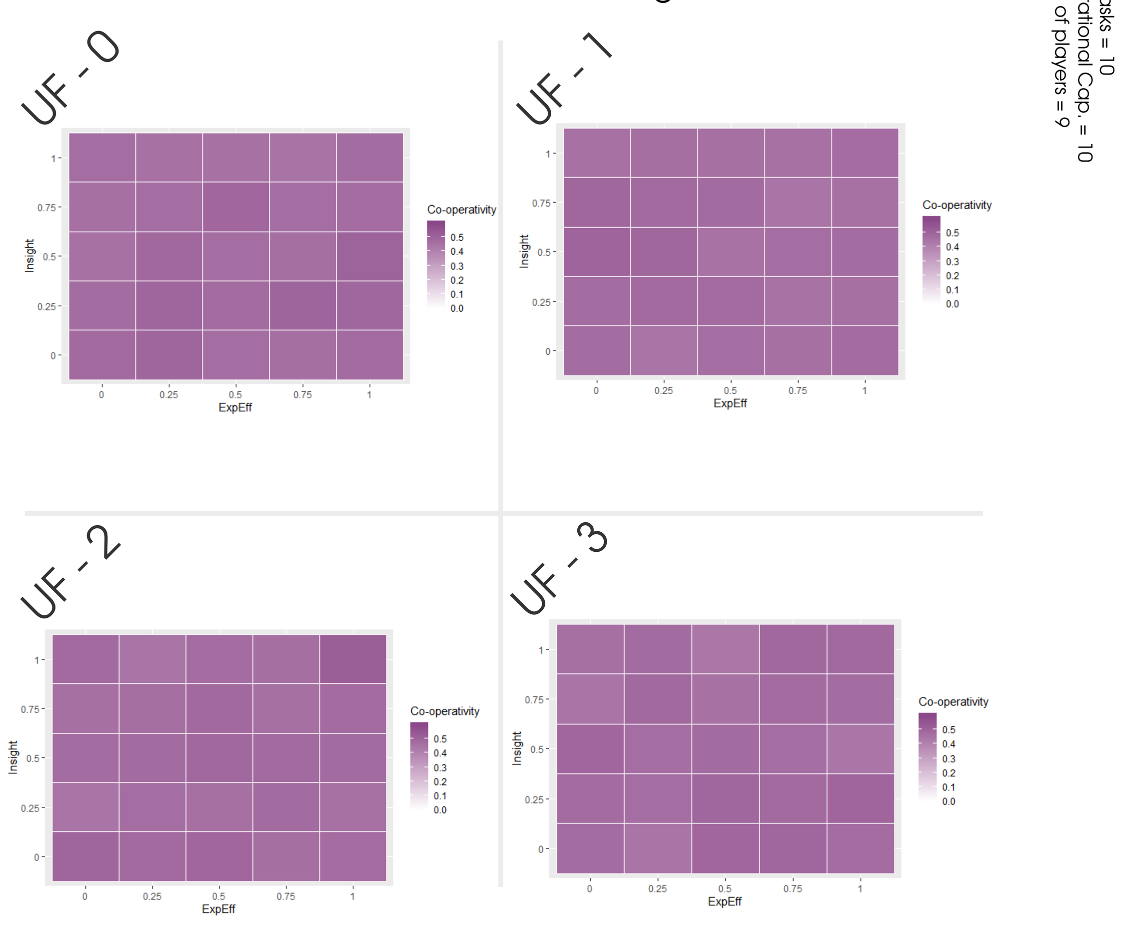

4. Calculate the Krackhardt E/I ratio and mutual information to quantify cooperation and division of labour respectively
5. Repeat this for set number of updates.

## PRELIMINARY RESULTS

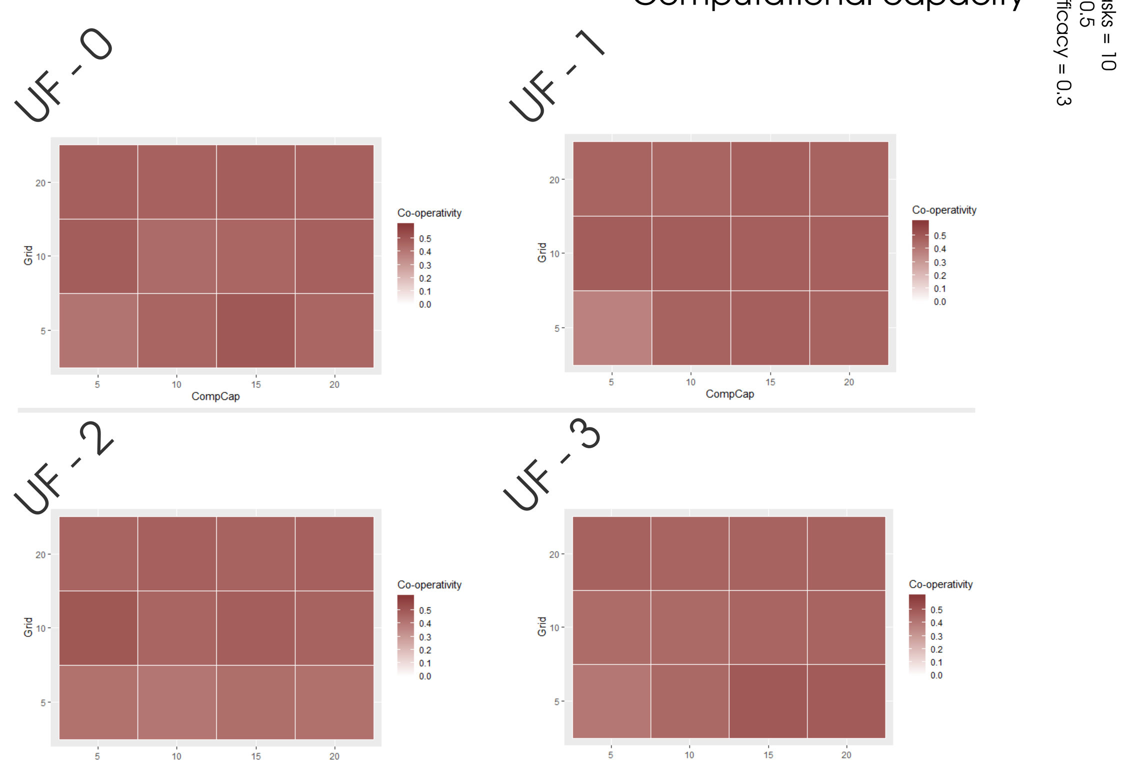



Average fitness of the network saturates within the first 1000 updates. Each utility function run has its own terminal fitness depending on the parameters.

Variation in soft parameters → Export Efficacy, Insight




Variation in hard parameters → Number of players, Computational capacity




## FUTURE DIRECTIONS

The preliminary results for EvoAgentNet indicate that there is no stable state for cooperativity in our system, and the final state appears to be mutation induced. Our future work would include reconfiguring the parameters, and developing new utility functions for the selfish agent network. We also propose to do a comparative study of the two systems.

## CONCLUSIONS

When optimized for social welfare, utility functions with compounding returns led to the evolution of cooperation in agents. Cooperativity also showed an increasing trend with export efficiency, insight and computational capacity.

## REFERENCES


[1] Jain, Shweta, et al. "Mechanisms with learning for stochastic multi-armed bandit problems." Indian Journal of Pure and Applied Mathematics 47.2
[2] Allen, Benjamin, Jeff Gore, and Martin A. Nowak. "Spatial dilemmas of diffusible public goods." Elife 2 (2013): e01169
[3] Koulouriotis and A. Xanthopoulos. "Reinforcement learning and evolutionary algorithms for non-stationary multi-armed bandit problems." Applied Mathematics and Computation 196.2 (2008): 913-922.
[4] Nowak, Martin A. "Five rules for the evolution of cooperation." science 314.5805 (2006): 1560-1563 .
[5] Wang, Yizao, Jean-Yves Audibert, and Rémi Munos. "Algorithms for infinitely many-armed bandits." Adv. in Neural Information Processing Sys.. 2009.
[6] Krackhardt, David, and Robert N. Stern. "Informal networks and organizational crises: An experimental simulation." Social psychology quarterly (1988): 123-140.
[7] Goldsby, Heather J., et al. "Task-switching costs promote the evolution of division of labor and shifts in individuality." Proceedings of the National Academy of Sciences 109.34 (2012): 13686-13691.