



Pioneering Excellence in Education,
Research & Innovation



Computational Ethics



Janvi, Karthik, Nikitha and Jayati
Prof. Srinath Srinivasa

Nov 16th-18th, 2022





Overview

Introduction to Computational Transcendence

Applications of CT

Supply Chain

Traffic
Management

Computational Ethics

Utilitarianism

Deontology

Virtue Ethics

Intrinsic Motivation

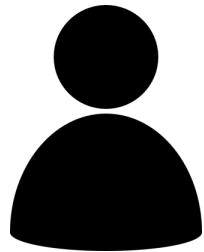
Affiliation

Achievement

Power

Computational Transcendence

Dilemma of Responsibility



Individually beneficial
With a collective / long-term cost

Irresponsible Choice

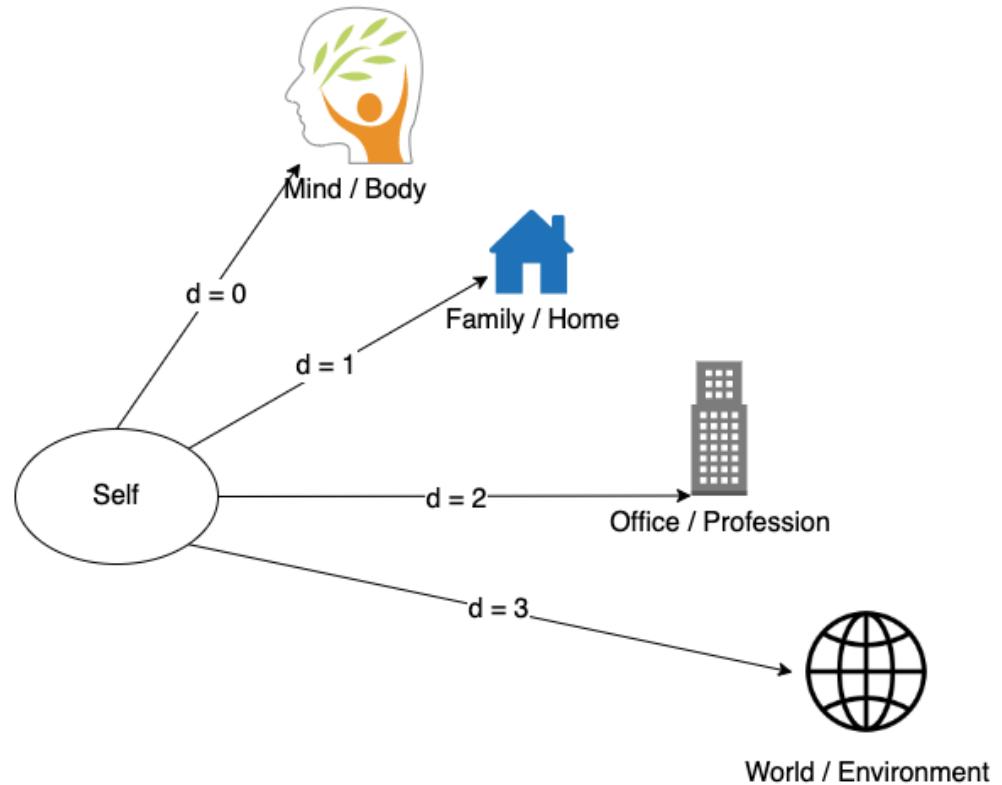


Individually suboptimal / costly
Collectively good

Responsible Choice

Elastic Sense of Self

We humans have elastic identities, we care not just about ourself, but also about our family, team members, environment, world etc.

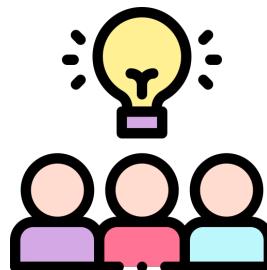


Computational Transcendence

Transcendence can be defined as the capability of an autonomous agent to virtually extend its identity to multiple aspects like agents, groups and concepts to different extents in a system.



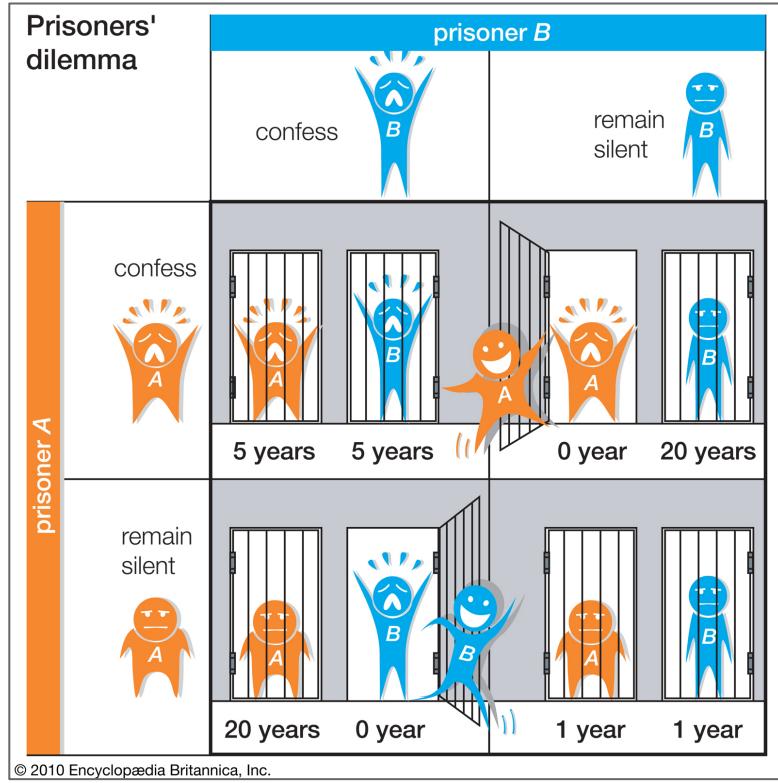
Rational Association



Association of Identity

vs

Prisoners' Dilemma

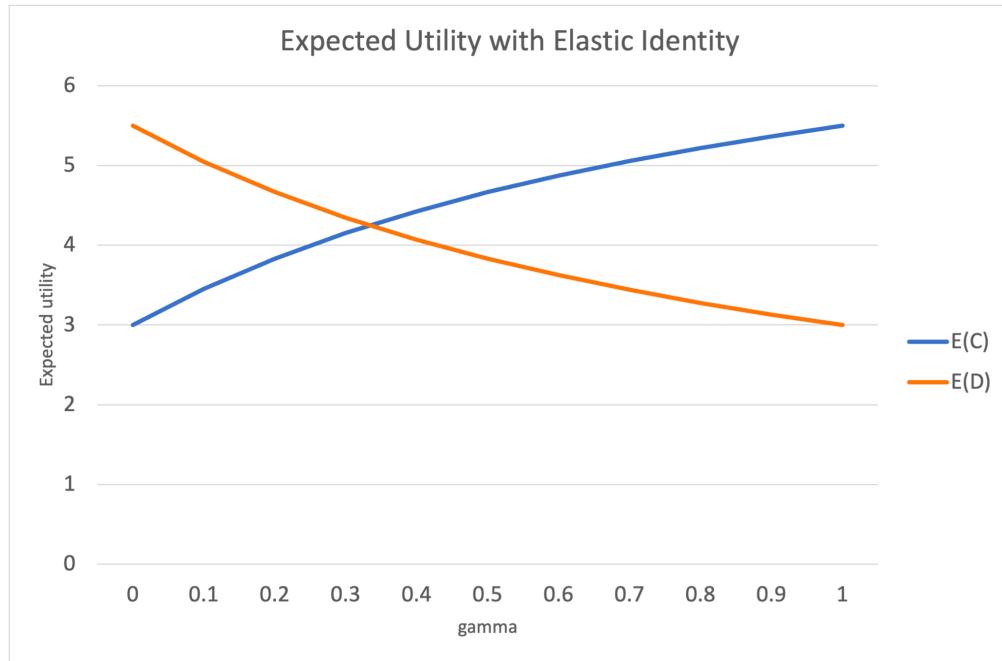


Prisoners' dilemma is popular in game theory. It demonstrates the dilemma of an agent to either:

Cooperate for mutual reward
Or
Defect for individual reward

This represents the dilemma of responsibility.

Transcendence in Prisoners' Dilemma

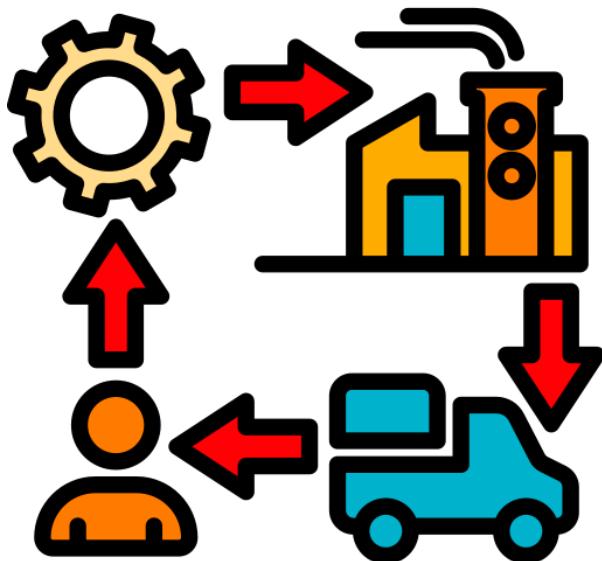


More an agent cares about others (transcends), the more it demonstrates responsible behaviour (cooperates).

C - Cooperate, **D** - Defect, **gamma** - extent to which the agent cares about others

Transcendence in Supply Chain

Supply Chain Setup



Customer: Places orders

Industry: Generates products

Intermediary: Ships orders

Dilemma in a Supply Chain



OR



Wait to accumulate more orders?

Ship all the orders received so far?

Supply chain Efficiency Indicators

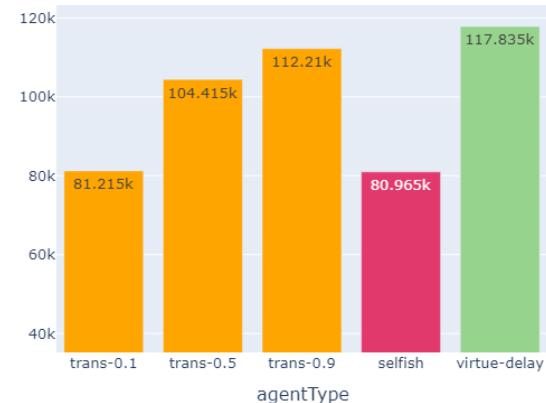
- **Total Cost** - storage cost, transportation cost
- **Total Utility** - profit on fulfilling orders
- **Average Delay Per Order** - time taken by orders in supply chain

Types of Agents

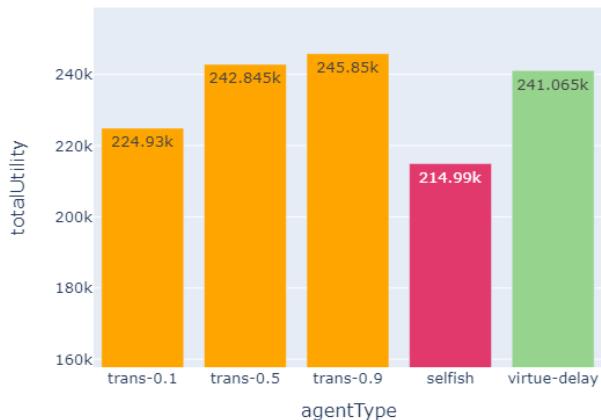


| Selfish Agent | Virtue-Delay Agent | Transcended Agent |
|----------------------|----------------------------|---|
| Minimizes its costs. | Minimizes delay of orders. | Accounts for other agents in its neighbourhood. |

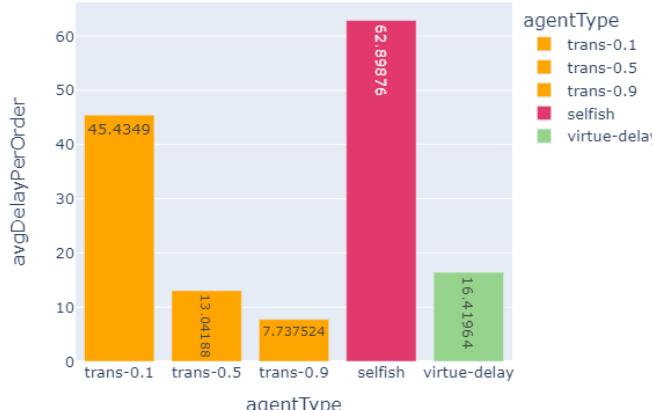
Results



Total Cost



Total Utility



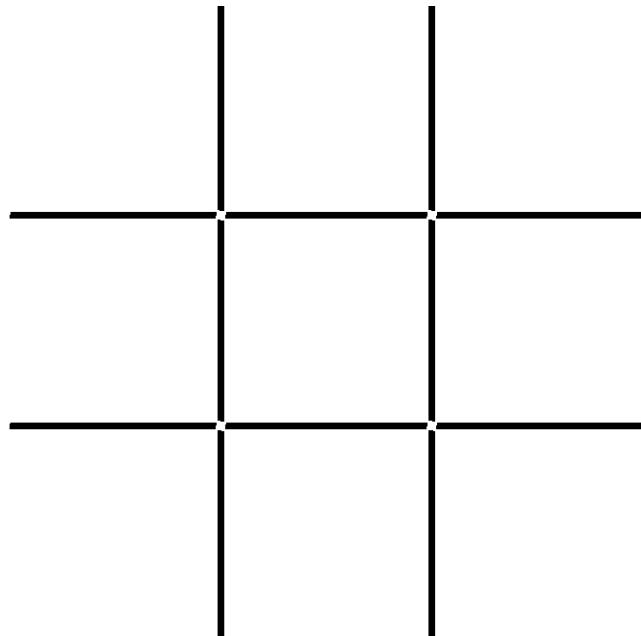
Avg delay per order

Although transcended agents incur high cost, they also gain higher utility and have lower average delay per order.

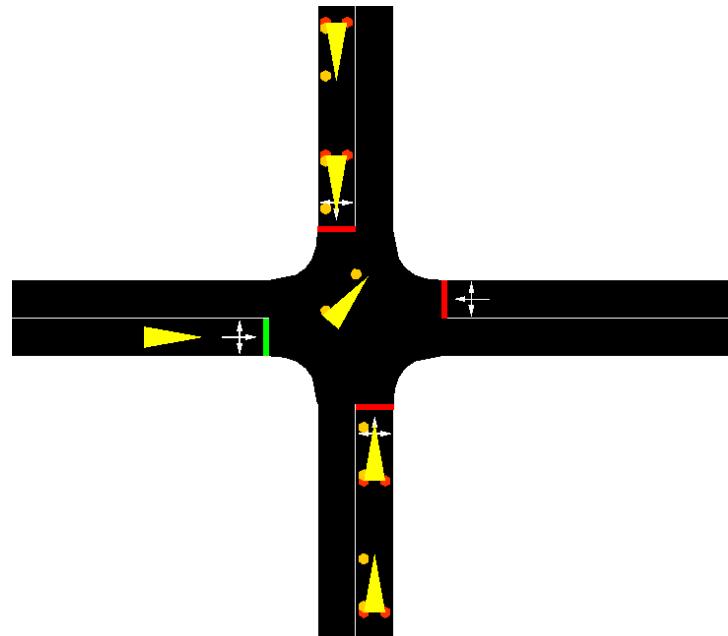
Transcendence for Traffic Management

Traffic Light Setup

A 4x4 road grid with traffic lights at each intersection.



A zoomed in view of an intersection.

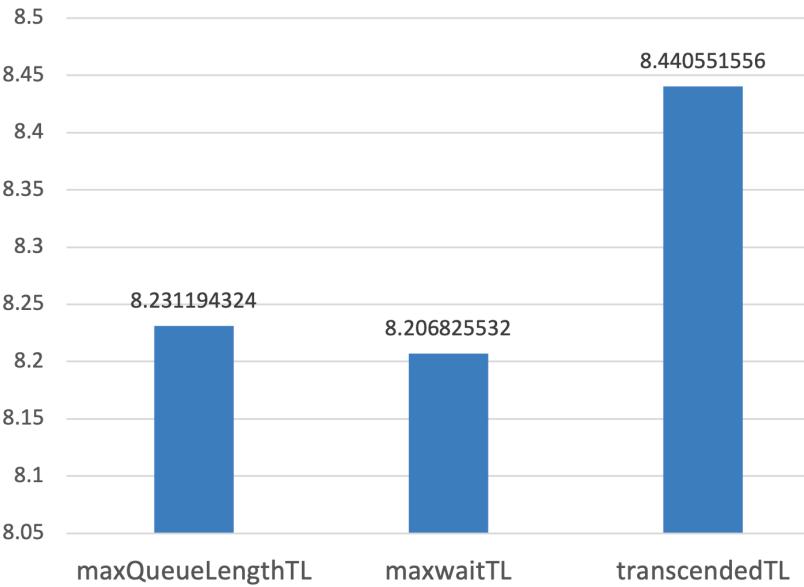


Types of Traffic Lights

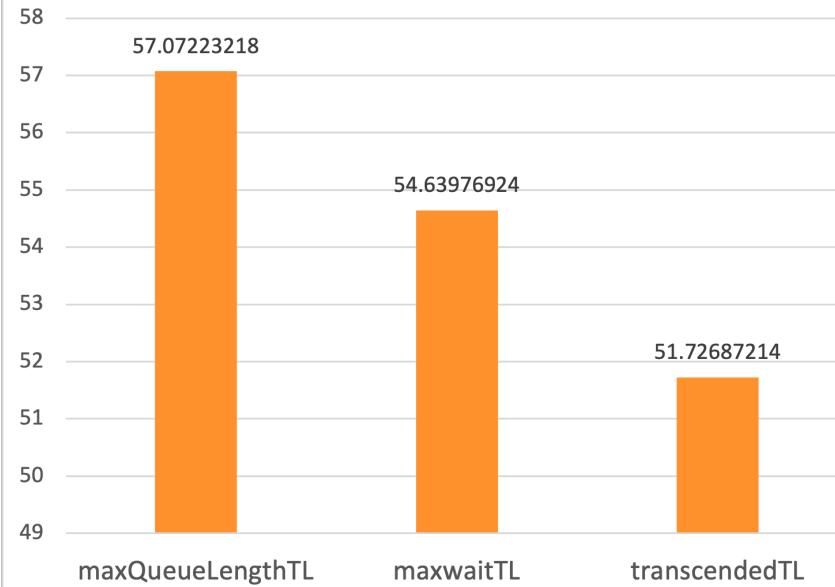
| Type | Description |
|-------------------------|--|
| Cyclic | Turn green for roads cyclically in a fixed loop. |
| Max Queue Length | Turn green for the road with max queue length. |
| Max Waiting Time | Turn green for the road with max waiting time. |
| Transcended | Turn green so as to ensure smooth traffic flows at own and neighbouring intersections. Care about the waiting time of vehicles passing through the intersection. |

Results

average vehicular speed (m/s)



average vehicular waiting time (s)



Transcended traffic lights perform better than other types of traffic lights!

Computational Ethics

Ethical Agents

| Icon | Ethical Agent Type | Description |
|------|--------------------|--|
| | Utilitarian | Maximizes collective utility of all stakeholders. |
| | Deontic | Learns from others and responds accordingly. |
| | Virtuous | Upholds context specific virtues. |
| | Transcended | Accounts for others in the system and then responds. |
| | Adversarial | Acts selfishly. |

Ethical Dilemma of a Doctor in a pandemic

Utilitarianism

Number of lives saved at home vs hospital.

Deontology

Get the population opinion and act accordingly.

Virtue Ethics

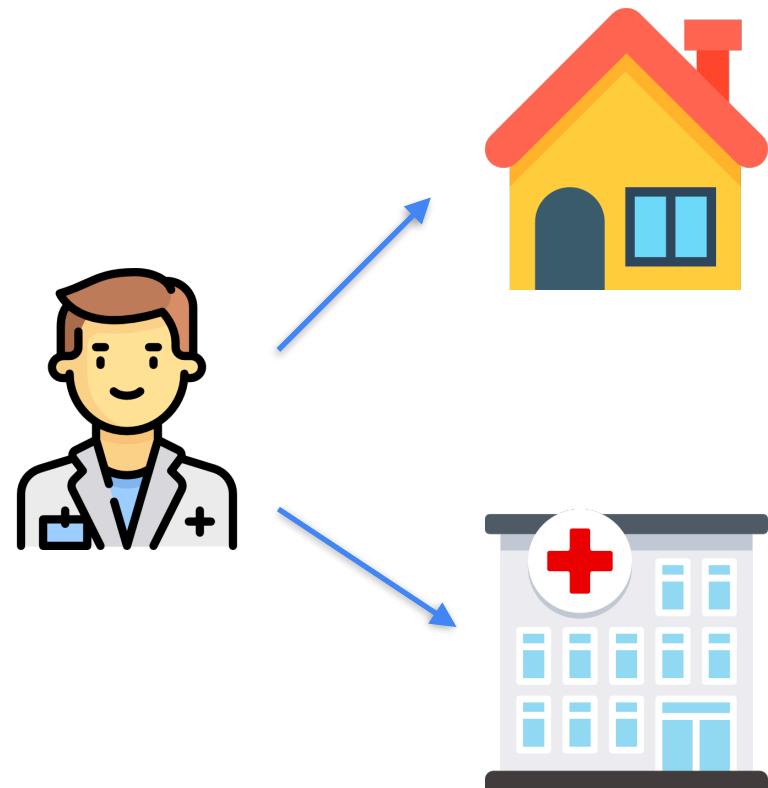
Uphold virtue of care.

Transcendence

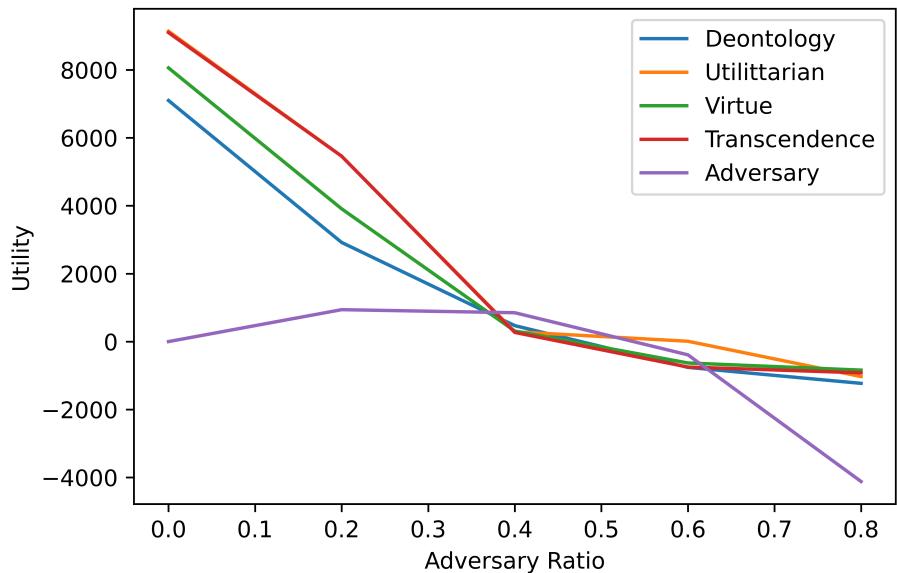
Account for family, patients and hospital.

Adversarial

Only care about self and family.



Results



As the adversary ratio increases:

Utility of all ethical agents decreases!

After a certain threshold, even the utility of the adversarial agents decreases (purple line).

Intrinsic Motivation

The idea of Intrinsic Motivation..



Soldier



Civilian

In the situation of a war



People can perceive same external situation in different manner.

That's where intrinsic motivation comes in!

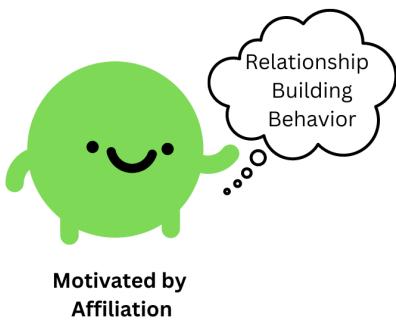
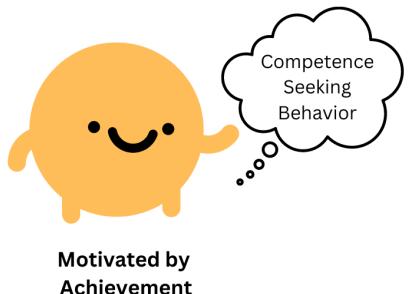
Depending on the intrinsic motivation people react to external situations!

We studied three intrinsic motivations..



Also known as “Influential Trio”!

They display a diverse response to a particular situation.



They played prisoners' dilemma game in a network.

It was observed that..

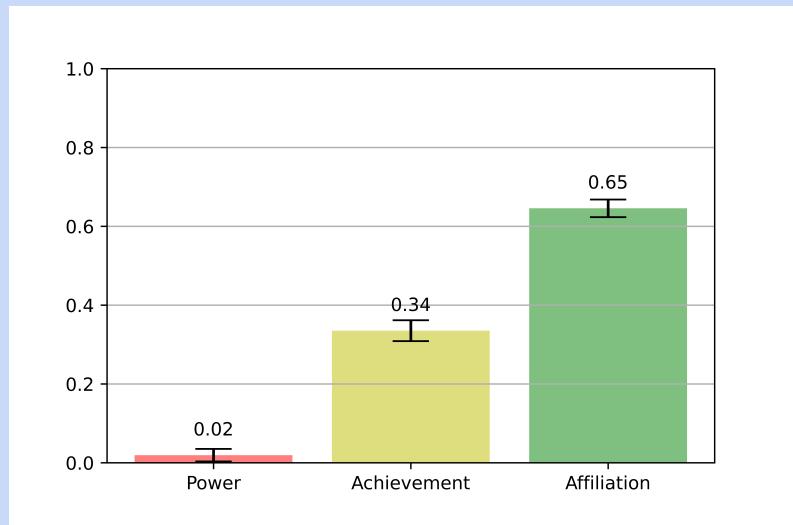
For the network to settle :

Affiliation: Majority (60-70)% , who are willing to cooperate.

Power : Least (0-10)% , who display controlling behaviour towards external incentive.

Achievement : Moderate (30-40)% , who want to be treated fairly.

Hence, a lot of affiliated agents (cooperators) are required to stabilize the network!



Settled proportion

Our Related Publications

- Jayati Deshmukh and Srinath Srinivasa. “Computational Transcendence: Responsibility and Agency.” *Frontiers in Robotics and AI* 9:977303. doi: 10.3389/frobt.2022.977303 [[link](#)]
- Srinath Srinivasa and Jayati Deshmukh. [“The Evolution of Computational Agency.” Novel Approaches to Information Systems Design.](#) IGI Global, 2020. 1-19. ISBN13: 9781799829751 [[link](#)]
- Jayati Deshmukh, Srinath Srinivasa and Sridhar Mandyam, “What Keeps a Vibrant Population Together?,” *Complex Systems*, 30(3), 2021 pp. 347–373. [[link](#)]
- Srinath Srinivasa, Jayati Deshmukh. AI and the Sense of Self. arXiv:2201.05576 [[link](#)]



Acknowledgements

Team Details

Prof. Srinath Srinivasa

Jayati Deshmukh

Janvi Chhabra

Sama Sai Karthik

Nikitha Adivi

Center for Internet of Ethical Things (CIET)



Contact

sri@iiitb.ac.in

