Evolutionary Dynamics of Artificial Agents, Exploration and Learning in Games

Brian Mintz,
advised by Professor Feng Fu

Dartmouth College, April 2024

Table of Contents

1. Background
2. Exploration / Exploitation
3. Opinion Dynamics
4. Conclusion

A biological basis for morality?

https://www.youtube.com/watch?v=meiU6TxysCg&t=78s
(Excerpt from Frans de Waal's TED talk).

Why is cooperation so widespread?

Models of cooperation based on the Prisoner's Dilemma and the Snowdrift game, by Michael Doebeli and Christoph Hauert, Ecology letters, 2005.

One solution

Evolutionary Game Theory uses mathematical models of evolution to understand the origins of behavior, allowing a rigorous commentary on profound questions like how to sustain international cooperation and the origin of multicellular life.

Beyond this intuitive appeal, it features a number of fascinating mathematical phenomena, including tipping points, complex nonlinear interactions, hysteresis, and chaos.

Like modeling in ecology, the goal is primarily **qualitative insights** rather than **quantitative predictions**.

Mathematical Approaches

Many fundamental models of evolution are given by **Markov chains**, a type of memory-less stochastic processes. These can often be summarized through a transition matrix which determines the long term behavior of the system, or the probability of getting stuck in some state.

$$\vec{x}_t = M^t \vec{x}_0$$

Various limits can transform these into deterministic models such as **systems of ordinary differential equations**.

$$\frac{d}{dt}\vec{x} = F(\vec{x})$$

Thesis at a glance

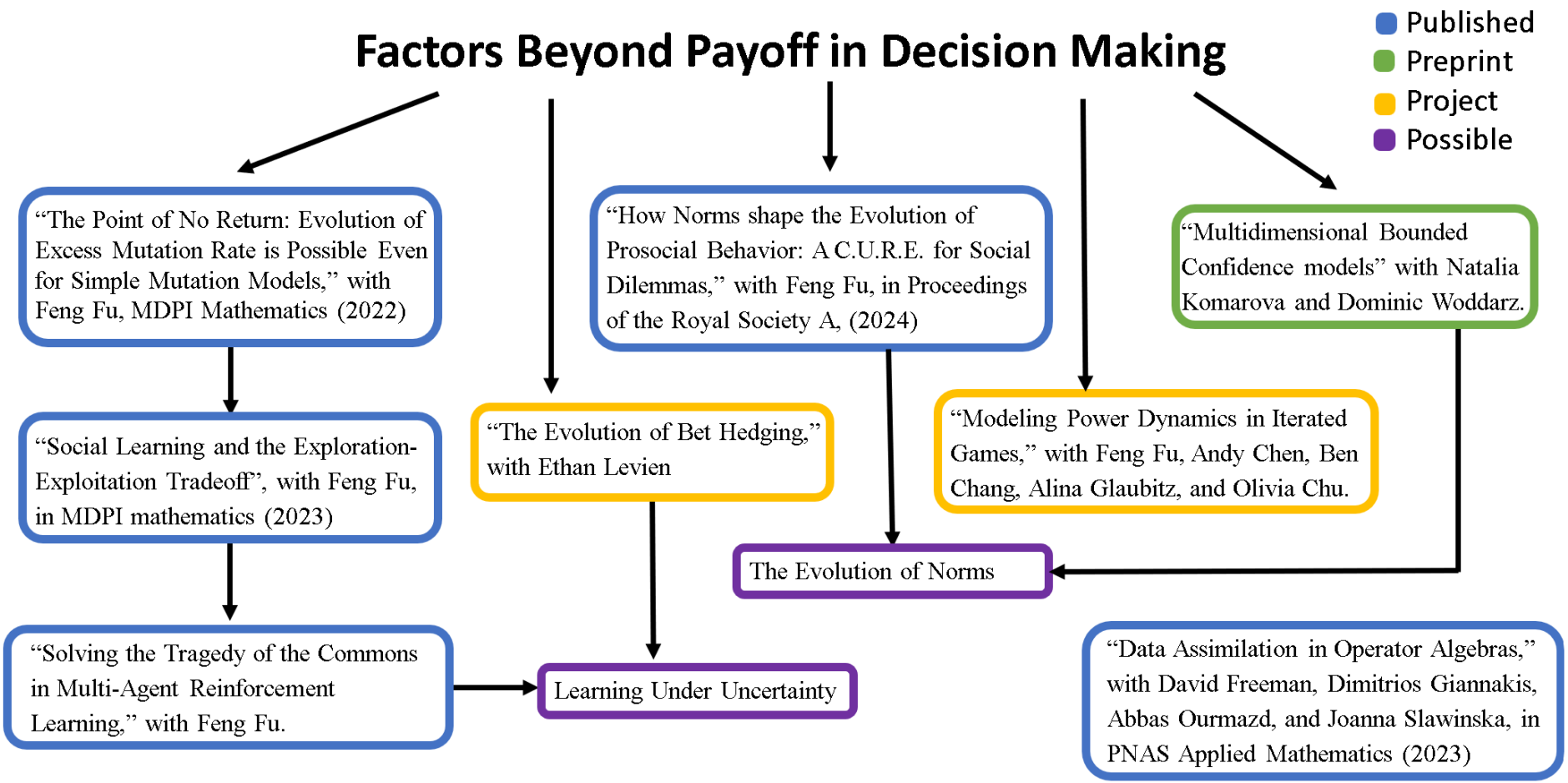


Table of Contents

1 Background

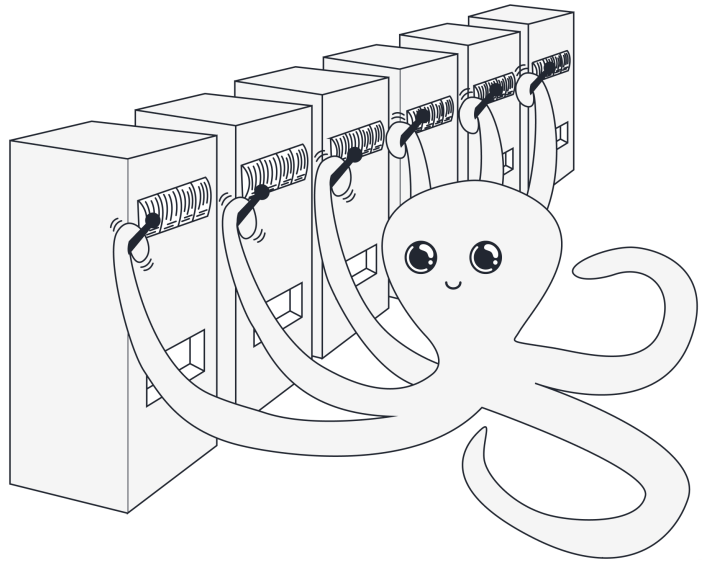
2 Exploration / Exploitation

3 Opinion Dynamics

4 Conclusion

The Multi-Armed Bandit

Decisions rarely happen in isolation. We often face the same choices over and over again: what to order at a restaurant, who to ask for advice, which drugs to continue developing, ...



We need to **explore** our options, but also **exploit** the best ones. The tension between these has been a longstanding question in computer science [GGW11, Whi80, S+19].

Reinforcement Learning

Reinforcement Learning is a Machine Learning paradigm that consists of an agent learning the optimal action depending on the state of an environment. The core idea is simple: actions with beneficial consequences will be repeated more.

Practice far outpaces theory with this technique, especially with Multi-Agent RL as this creates a dynamic environment. We sought to improve understanding of exploration by combining evolutionary and learning dynamics, focusing on the evolution of temperature, an exploration parameter.

Q-learning

We consider a foundational model of RL with strong theoretical guarantees [WD92, LP22]. Agents keep a table of values for each action. Under the Boltzmann mechanism, their probability of choosing action i is

$$x_i(t) = \frac{\exp(Q_i(t)/T)}{\sum_i \exp(Q_i(t)/T)}$$

These Q-values are then updated by

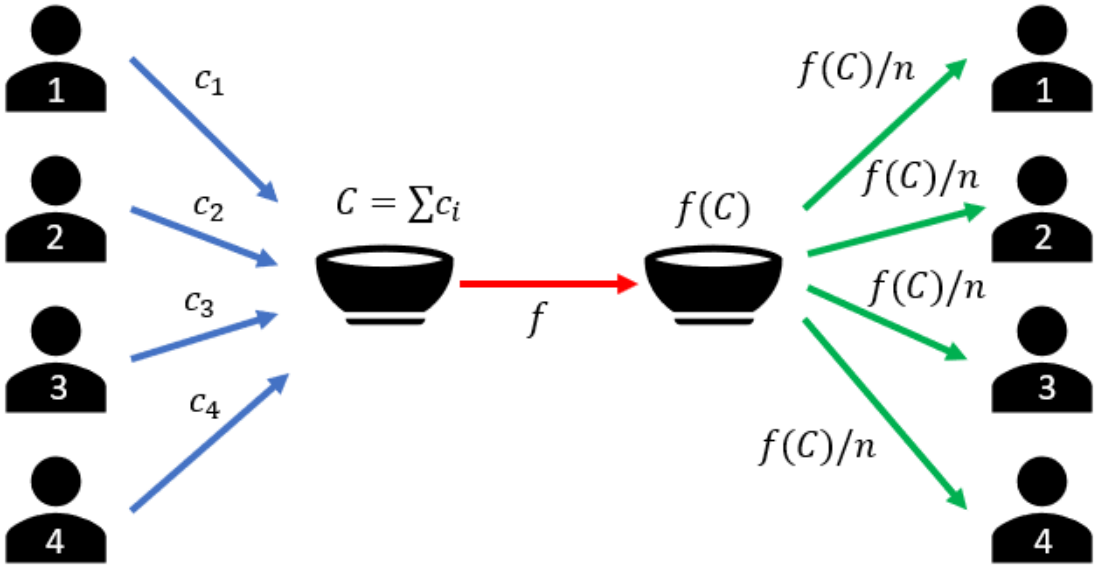
$$Q_i(t+1) = Q_i(t) + \alpha[r_i(t) - Q_i(t)] \tag{1}$$

where $r_i(t)$ is the reward of choosing action i at time t, and α is the learning rate.
Taking the time derivative then rearranging and scaling time by α/T, we find

$$\frac{\dot{x}_i}{x_i} = \left[r_i - \sum_k x_k r_k \right] - T \sum_k x_k \ln \frac{x_i}{x_k} \tag{2}$$

Public Goods Games

Each of N individuals contributes some amount c_i to a pot, which is scaled by a function $f(x)$ then distributed evenly among the players.



Free-Rider problem / Tragedy of the Commons: individuals benefit from contributing less, but this hurts the collective.

Stochastic Model

Our model consists of a population of N agents following reinforcement learning: stateless Q-learning with parameters α and T. Each time step, a group of k individuals is selected to interact, and receives payoffs from the public goods game based on the set of actions they choose. Then each individual has probability r of independently being replaced by another proportional to their fitness (averaged over interactions).

Rather than study individuals evolutionary trajectories, we investigate the **fixation probability** among traits, their probability of replacing all individuals with the resident trait. These would allow us to simulate evolutionary trajectories, assuming the mutation rate is low enough that only two types occur at a time.

Deterministic Model

Letting x be the probability of an agent contributing, their strategy dynamics are given by

$$\dot{x} = x(1-x)\left(r_C - r_D - T \ln \frac{x}{1-x}\right) \tag{3}$$

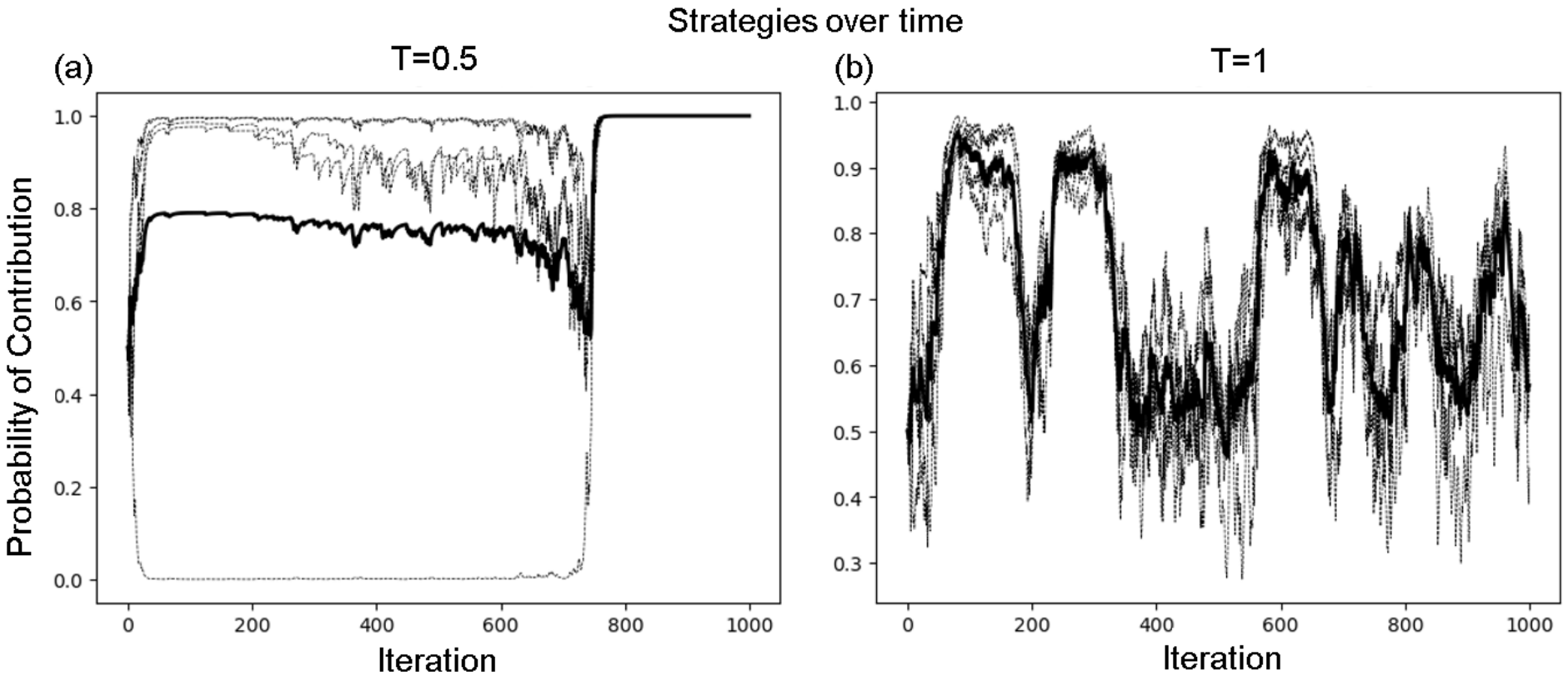
Assuming all agents follow the strategy x, the averaged rewards are

$$\text{cooperation: } r_C(x) = -1 + \sum_{i=0}^{k-1} \binom{k-1}{i} x^i (1-x)^{k-1-i} f(i+1) \tag{4}$$

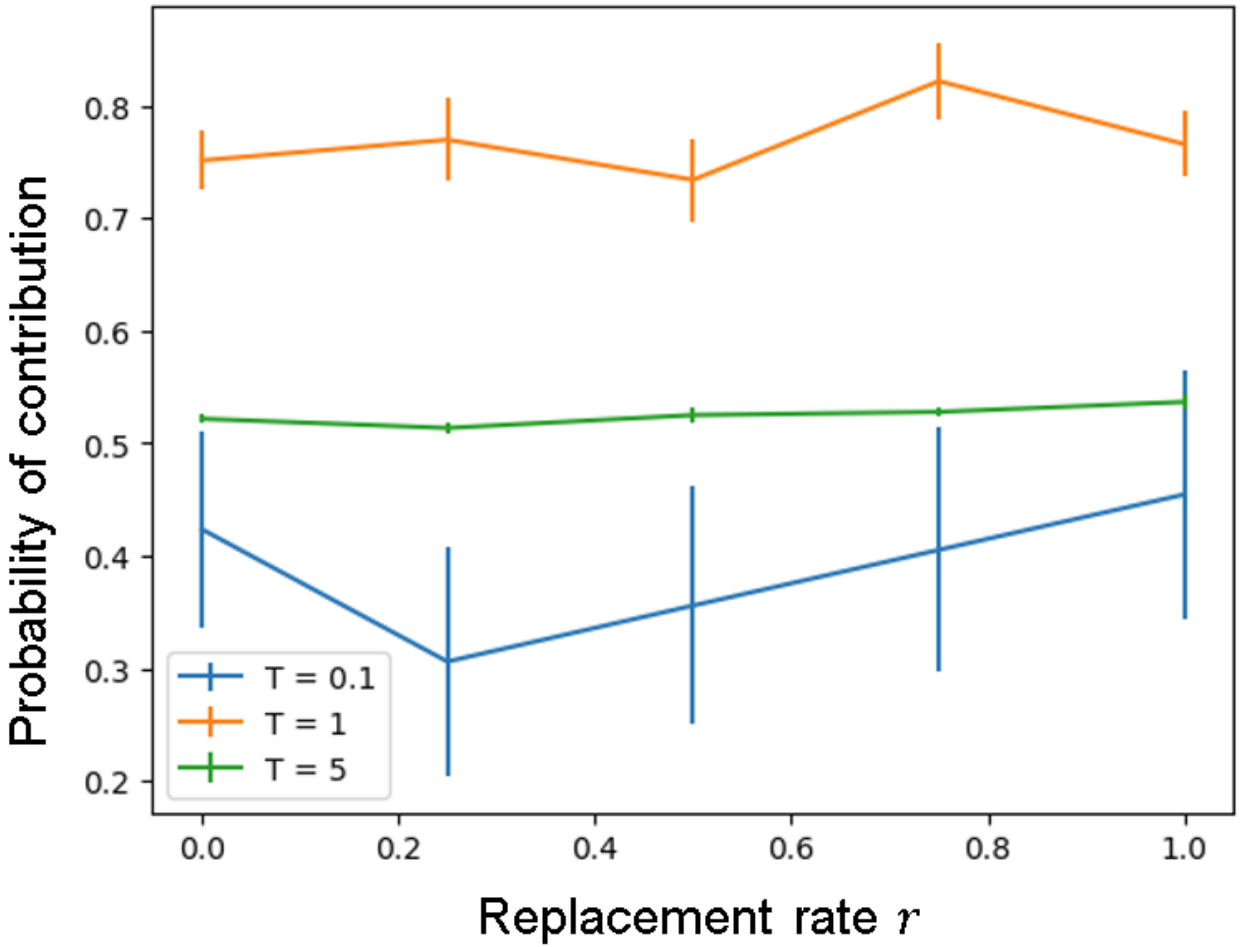
$$\text{defection: } r_D(x) = \sum_{i=0}^{k-1} \binom{k-1}{i} x^i (1-x)^{k-1-i} f(i) \tag{5}$$

We study a proxy of fixation probability, the **invasion fitness** $p(m,r) - p(r,r)$ where $p(m,r)$ is the (equilibrium) payoff to the mutant if one individual follows the mutant strategy m while the rest follow resident strategy r.

Learning Trajectories

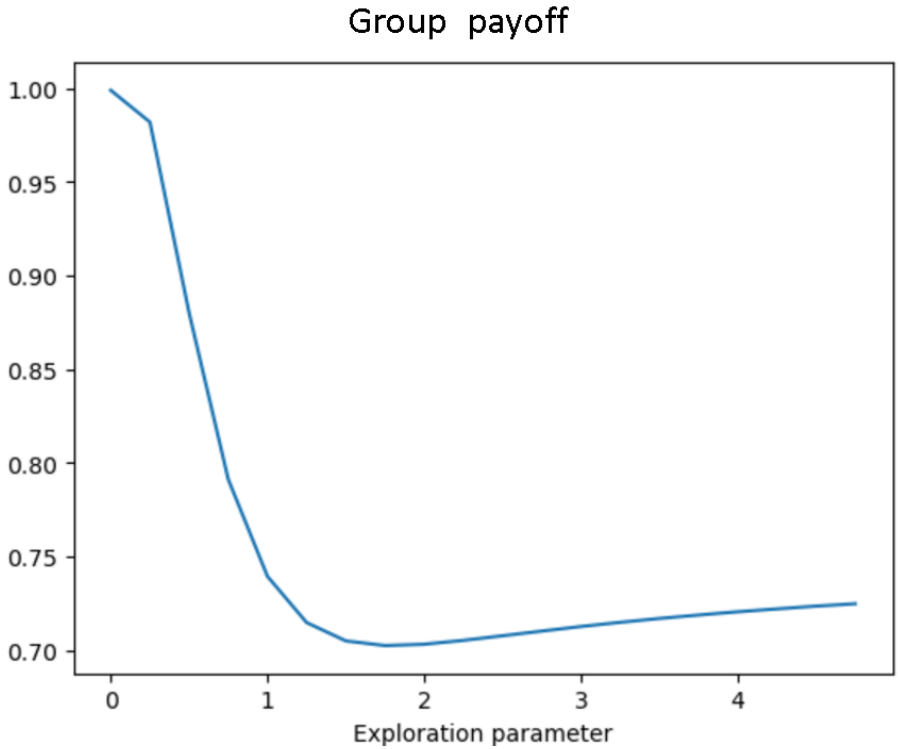


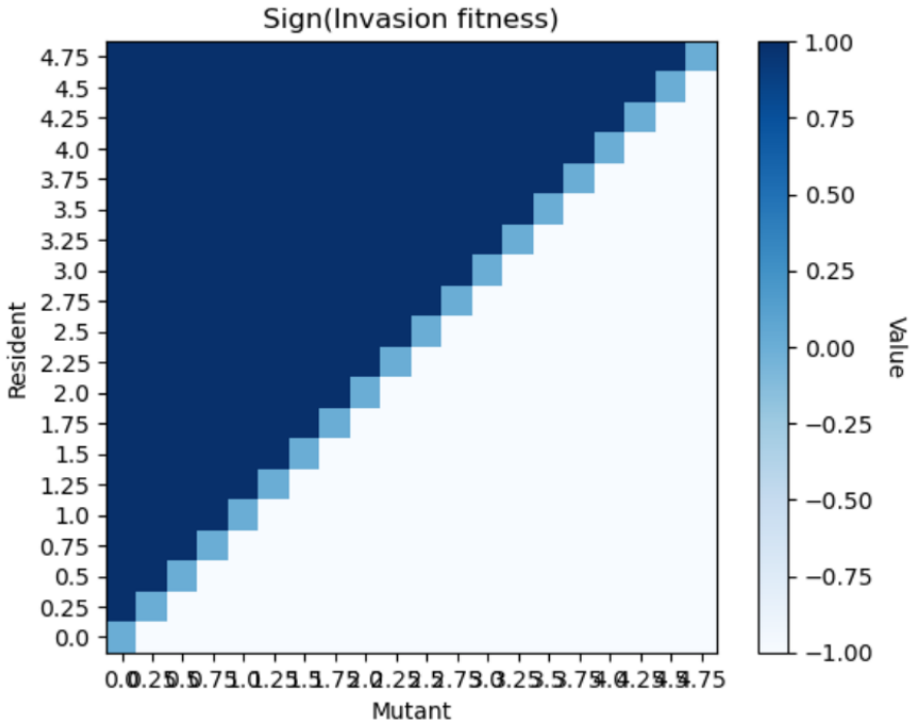
Replacement rate and Temperature



Temperature has a non monotonic effect on contribution levels, and replacement generally increases contribution.

Selection effects

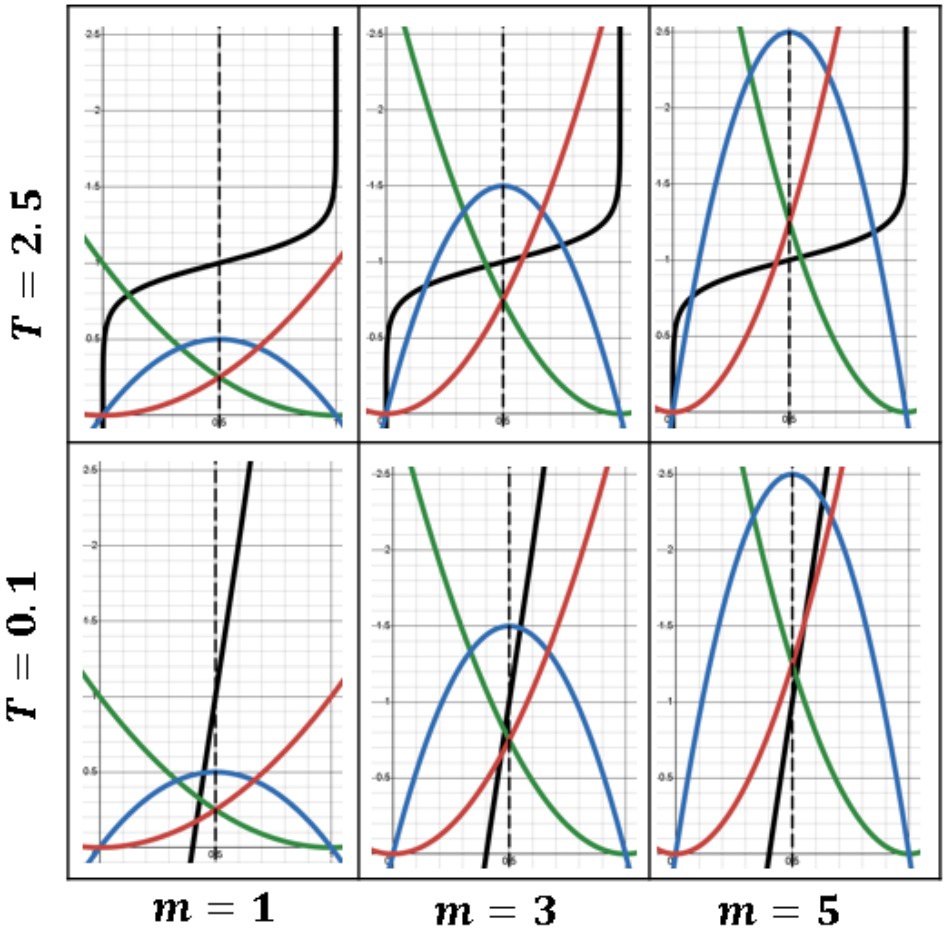


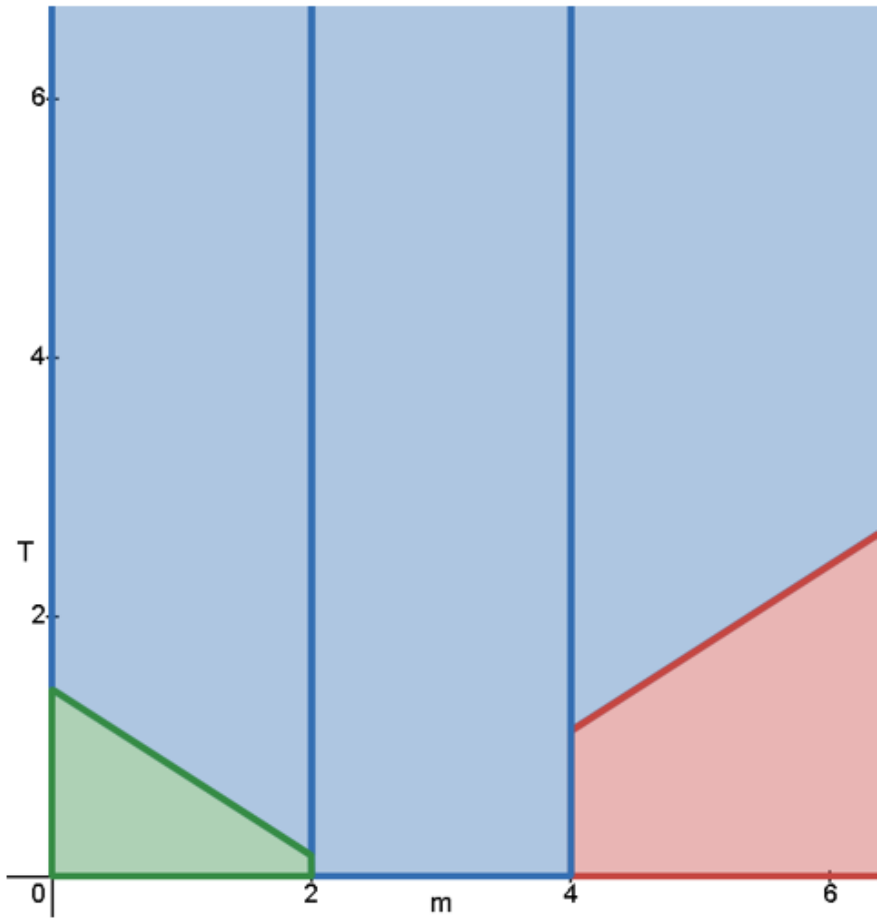


By varying the reward function, we found cases where the temperature evolved up, down, and to a stable value. The above plot uses rewards $[1, 1, 1, 1, 5]$. Since the system was solved numerically, there was also some nuance here.

Optimal Control

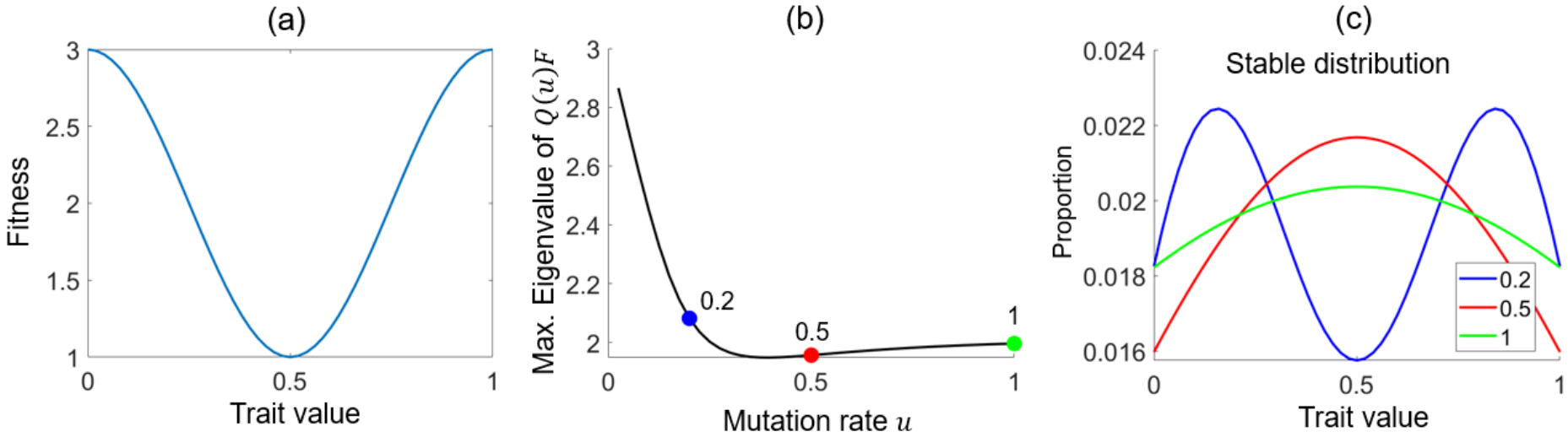
We also determine when each boundary case optimizes the level of contribution.





Prior Investigations

This work extended our previous studies into a model with exploration by simple diffusion. There we had a closed form for the invasion fitness as the maximum eigenvalue of a certain matrix.



We found the counterintuitive result that these mutation rates could evolve upwards even under constant selection! This effect was robust to a variety of local mutation models.

Summary

Results:

- Our work extended MARL through dynamic groups, similar to the transition from Classical to Evolutionary Game Theory. We investigated this model through stochastic simulations and a deterministic system of ODE's.
- Restricting to a small case, we characterized when selection would be positive or negative and the optimal reward function.

Next steps:

- Add a number of rounds, or continuation probability, parameter to investigate iterated games.
- Use more sophisticated learning methods, such as neural networks or deep Q-learning.
- Extend the analysis to groups with heterogeneous exploration rates.

Table of Contents

1 Background

2 Exploration / Exploitation

3 Opinion Dynamics

4 Conclusion

Motivation

Question: How do local interactions shape global trends in opinions?

Applications: polarization and echo chambers, the spread of misinformation, adoption of new technologies[WSP+24], optimal advertising in politics or marketing [AKPT18], ...

Complex systems: these are an interesting example of local interactions creating global effects. Translating between these micro and macro scales can reveal unexpected emergent phenomena.

Background

Opinion dynamics studies the degree of fragmentation caused by a wide variety of local update rules[SLST17, Lor07]. Of particular interest are consensus and polarization states.

- Statistical physics approaches[CFL09] study phase transitions, like in the **voter models** where individuals adopt opinions from their neighbors at random[CS73, HL75].
- The **DeGroot** model[DeG74] updates the vector of opinions through multiplication by an influence matrix.
- The **Hegselmann–Krause** model uses a dynamic influence matrix where all opinions within some distance of one's own opinion are given equal weight[RK02].
- Many other models ...

Stochastic Model (extension of Axelrod's work)

- A population of N individuals each have J independent binary opinions (0 or 1), where issue i has weight $w_i \in \mathbb{R}^+$.
- Iteratively, a focal and target individual are chosen, as well as one of the J issues, all uniformly at random. If the similarity

$$\sigma(s,t) = \sum_{i=1}^{J} w_i \delta(s_i, t_i) \tag{6}$$

of their opinions s and t is above a friendship threshold α_f, the focal individual adopts the target's opinion on the issue discussed. If the similarity is below an enemy threshold α_e, they adopt the opposite opinion.

- For the majority of this work we use $\alpha_f = \alpha_e$ for simplicity.

Deterministic model

To better understand these dynamics, we take a large population limit to obtain a system of differential equations

$$\frac{d}{dt}x_s = -x_s\left[\sum_t L_{s,t}x_t\right] + \sum_{i=1}^{J} x_{f_i(s)}\left[\sum_t G^i_{s,t}x_t\right]$$

$$L_{s,t} = \begin{cases} 1 - \frac{d(s,t)}{J} & \sigma(s,t) < \alpha_e \\ 0 & \alpha_e < \sigma(s,t) < \alpha_f \\ \frac{d(s,t)}{J} & \alpha_f < \sigma(s,t) \end{cases}$$

$$G^i_{s,t} = \begin{cases} \frac{1}{J} & \sigma(f_i(s),t) < \alpha_e \text{ and } t_i \neq s_i \\ \frac{1}{J} & \sigma(f_i(s),t) > \alpha_f \text{ and } t_i = s_i \\ 0 & \text{otherwise} \end{cases}$$

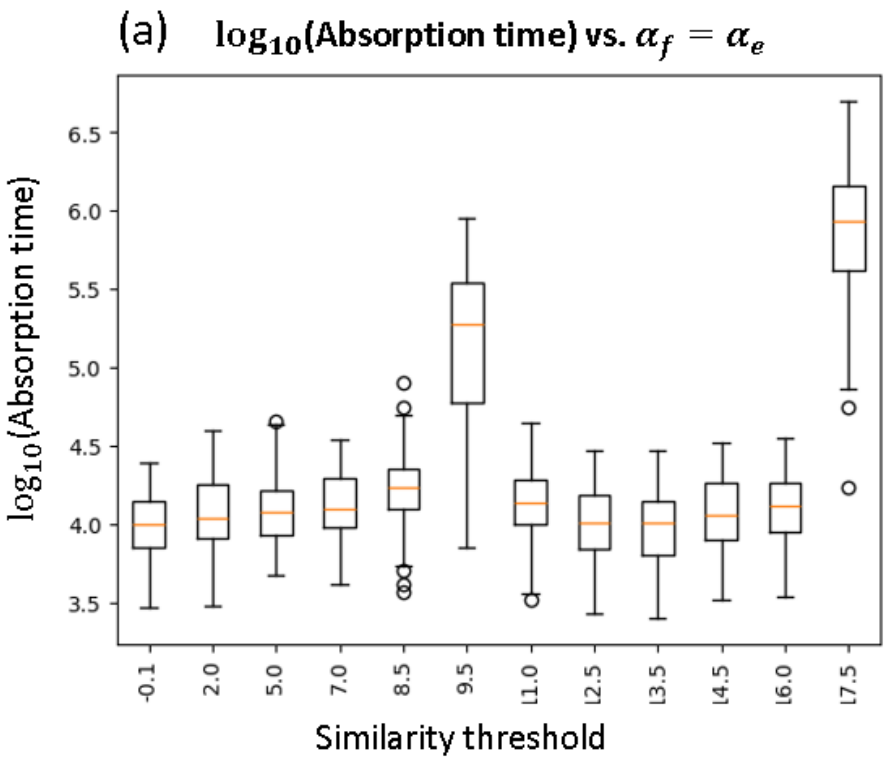
Example

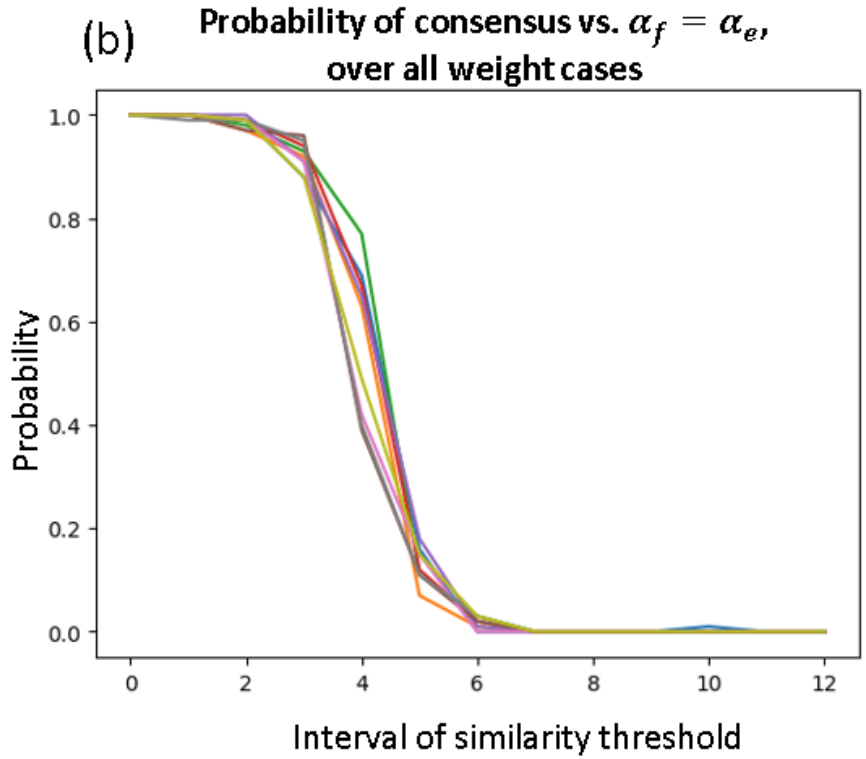
For $\alpha_f = \alpha_e = 9.5$ with weights $[9, 8, 6, 4]$, we have $4\frac{d}{dt}x_{0000}$ is

$$
\begin{aligned}
&- x_{0000}(x_{0001} + x_{0010} + 2x_{0011} + x_{0100} + 2x_{0101} + 2x_{0110} + x_{0111} + \dots \\
&\dots + x_{1000} + 2x_{1001} + 2x_{1010} + x_{1011} + 2x_{1100} + x_{1101} + x_{1110}) \\
&+ x_{1000}(x_{0000} + x_{0001} + x_{0010} + x_{0100} + x_{1111}) \\
&+ x_{0100}(x_{0000} + x_{0001} + x_{0010} + x_{1000} + x_{1111}) \\
&+ x_{0010}(x_{0000} + x_{0001} + x_{0100} + x_{1000} + x_{1111}) \\
&+ x_{0001}(x_{0000} + x_{0010} + x_{0100} + x_{1000} + x_{1111})
\end{aligned}
$$

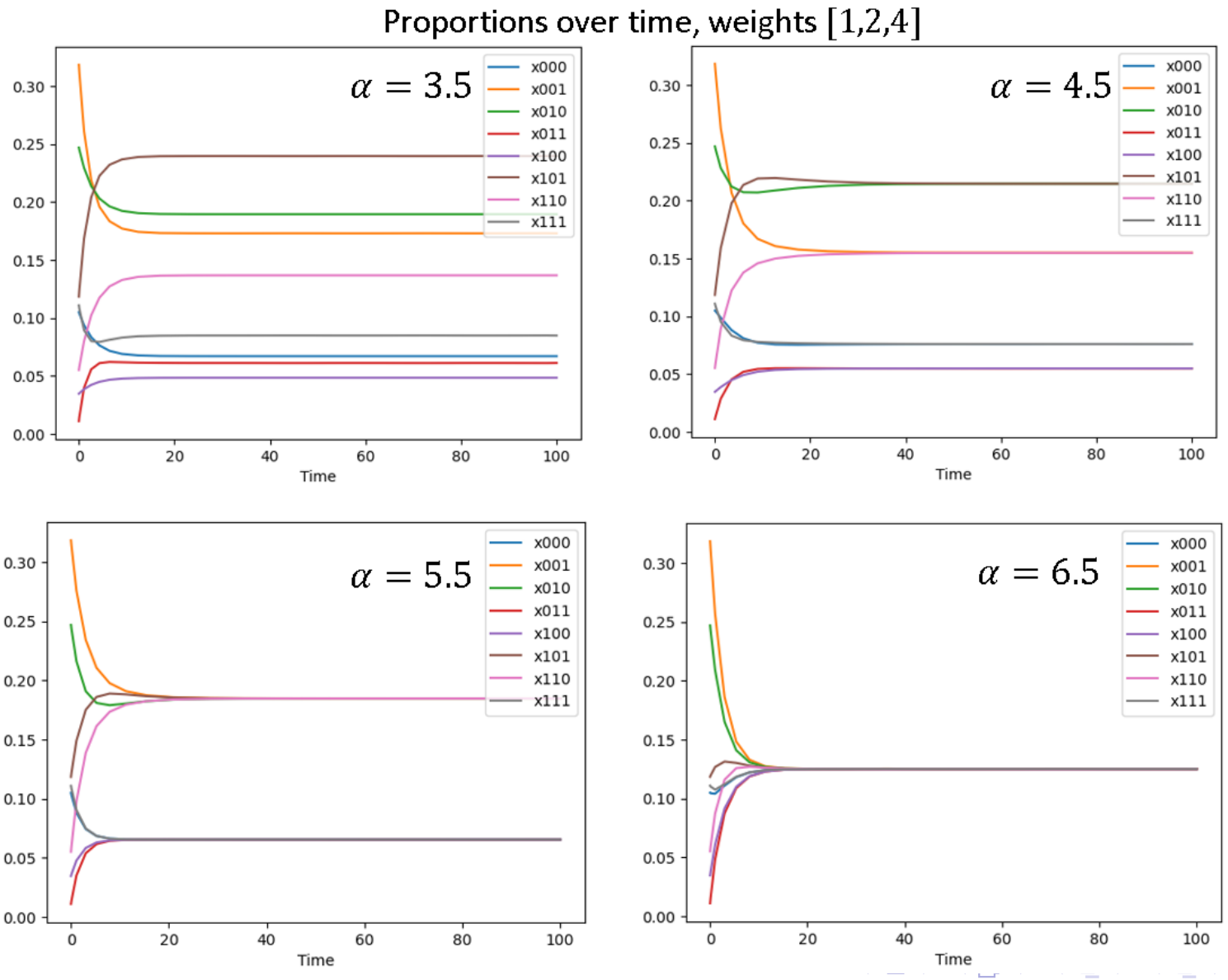
We are trying to understand the behavior of this nonlinear system of equations in 2^J variables with respect to $J + 2$ parameters.

Convergence times and Probability of consensus



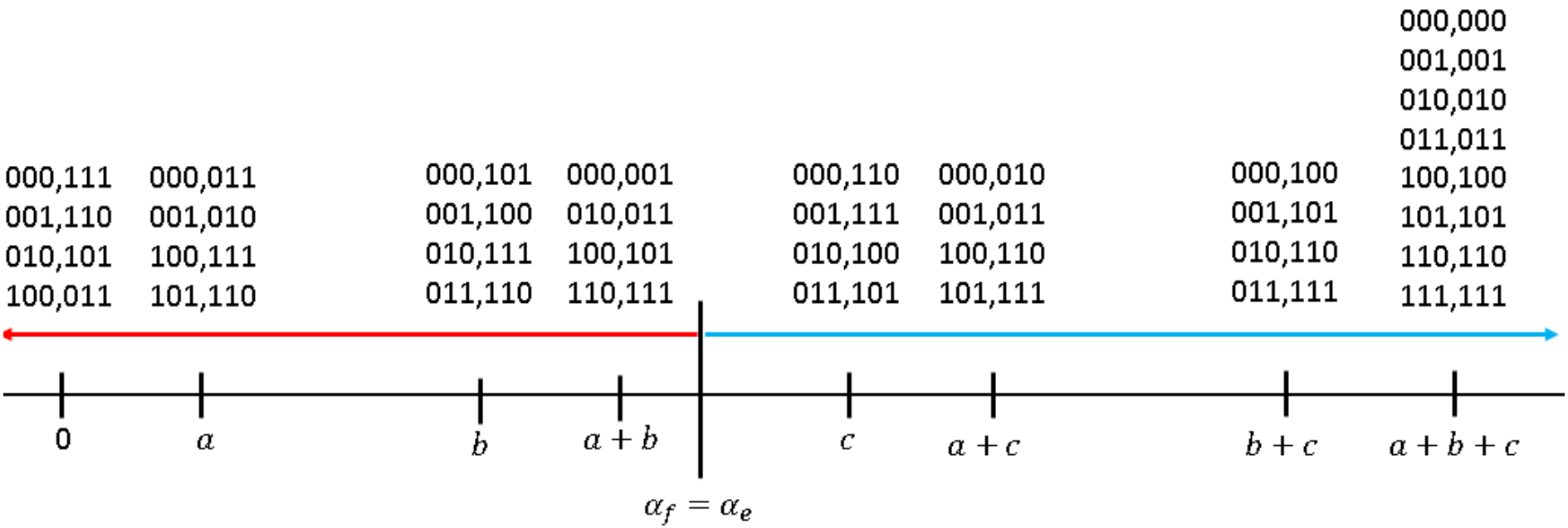


Symmetries in ODE model



Classification of α

This model only depends on the relationships between opinions. There are only finitely many values for the similarity of two opinions, given by a sum of some subset of the weights.



Any two thresholds in an interval between adjacent subset sums will produce the same model.

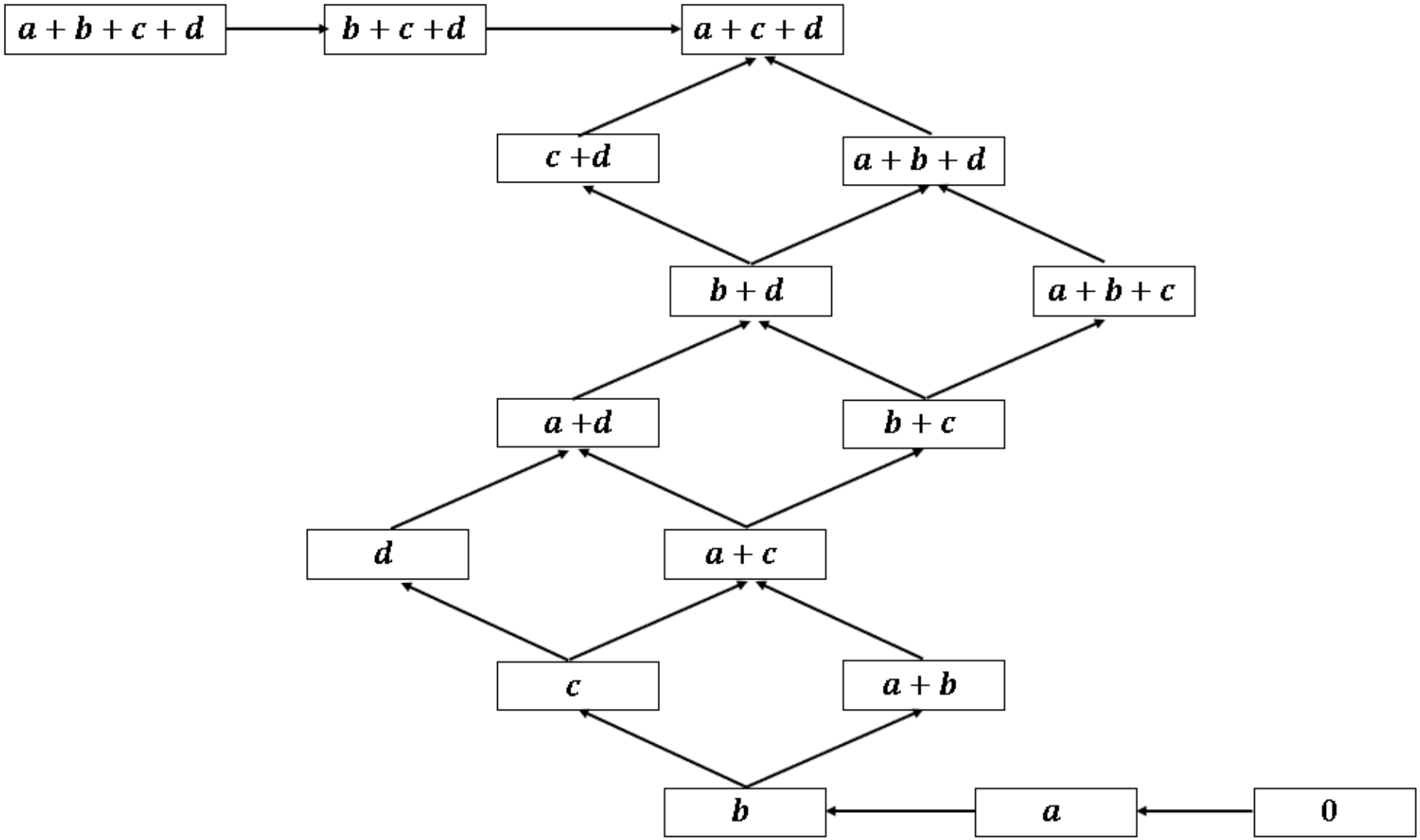
Classification of weights

If two weight lists have these subset sums in different orders, then they can produce different relationships between the opinions and therefore different models. Thus it suffices to determine all possible orderings of these subset sums. These are completions of the partial order of subsets sums into a total order.

We can do this by the **topological sorting** algorithm: at each iteration, one chooses a minimal element of the partially ordered set as the next smallest, removes it from the partially ordered set and possibly updates the ordering.

The number of these cases for each J is 1,1,2,14,516,124187,214580603,... (OEIS A009997).

Hasse Diagram for $J = 4$

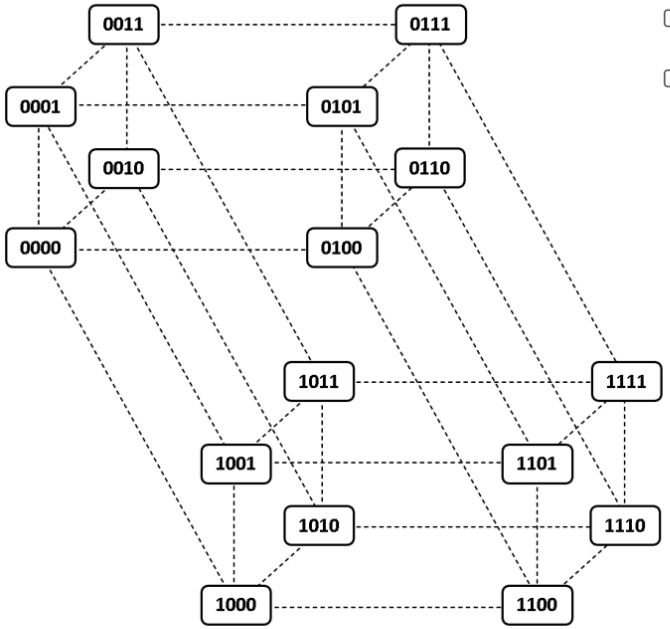


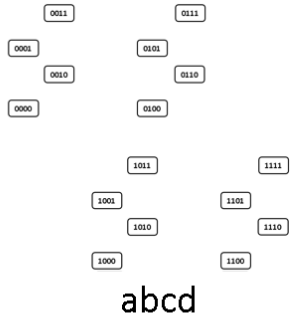
Symmetry Table

Conditions on weights $[a,b,c,d]$, $d<c<b<a$				
$d+c<b$	$d+b<a$	$a<b+c$	$a+d<b+c$	
			$b+c<a+d$	
		$b+c<a$	$a<d+b+c$	
			$d+b+c<a$	
	$a<d+b$	$a+d<b+c$		
		$b+c<a+d$		
$b<d+c$	$d+c<a$	$d+b<a$	$a<b+c$	$a+d<b+c$
				$b+c<a+d$
			$b+c<a$	$d<a+b+c$
				$a+b+c<d$
		$a<d+b$	$a+d<b+c$	
			$b+c<a+d$	
	$a<d+c$	$a+d<b+c$		
		$b+c<a+d$		

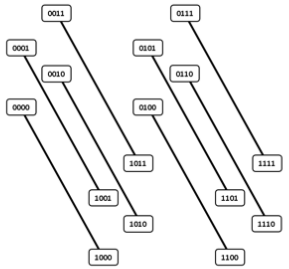
Interval for Similarity threshold $\alpha_f = \alpha_e$																
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
xyzt	xyzt	xyz*	xy**	xy**	xy**	xyz*	S	S	S	S	C	C	****	****	****	****
xyzt	xyzt	xyz*	xy**	xy**	xy**	xyz*	S	D	S	S	C	C	****	****	****	****
xyzt	xyzt	xyz*	xy**	xy**	xy**	xyz*	xyzt	D	D	S	C	C	****	****	****	****
xyzt	xyzt	xyz*	xy**	xy**	xy**	xyz*	xyzt	xyzt	xyzt	S	C	C	****	****	****	****
xyzt	xyzt	xyz*	xy**	xy**	xy**	C	S	S	S	C	C	C	****	****	****	****
xyzt	xyzt	xyz*	xy**	xy**	xy**	C	S	S	S	C	C	C	****	****	****	****
xyzt	xyzt	xyz*	xy**	x***	xy**	xyz*	S	S	S	S	C	****	****	****	****	****
xyzt	xyzt	xyz*	xy**	x***	xy**	xyz*	S	D	S	S	C	****	****	****	****	****
xyzt	xyzt	xyz*	xy**	x***	xy**	xyz*	xyzt	D	D	S	C	****	****	****	****	****
xyzt	xyzt	xyz*	xy**	x***	xy**	xy**	xyzt	xyzt	D	S	C	****	****	****	****	****
xyzt	xyzt	xyz*	xy**	x***	xy**	xy**	S	S	S	C	C	****	****	****	****	****
xyzt	xyzt	xyz*	xy**	x***	xy**	C	S	D	S	C	C	****	****	****	****	****
xyzt	xyzt	xyz*	xy**	x***	****	C	S	S	S	C	****	****	****	****	****	****
xyzt	xyzt	xyz*	xy**	x***	****	C	S	S	S	C	****	****	****	****	****	****

Symmetry diagrams



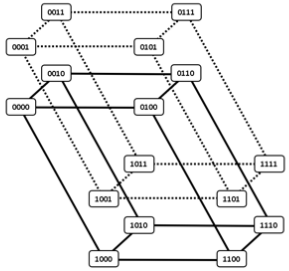


abcd

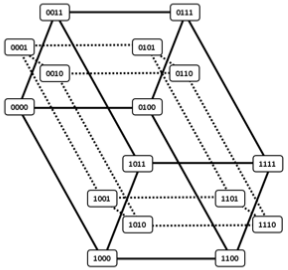


*bcd

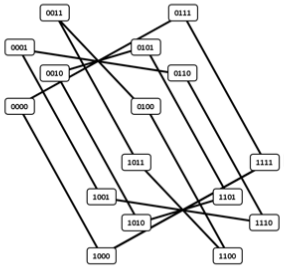
**cd



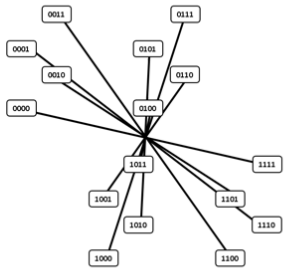
***d



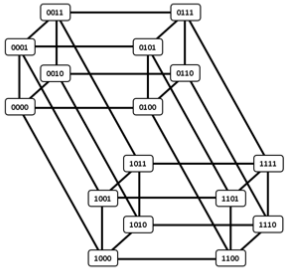
C



s

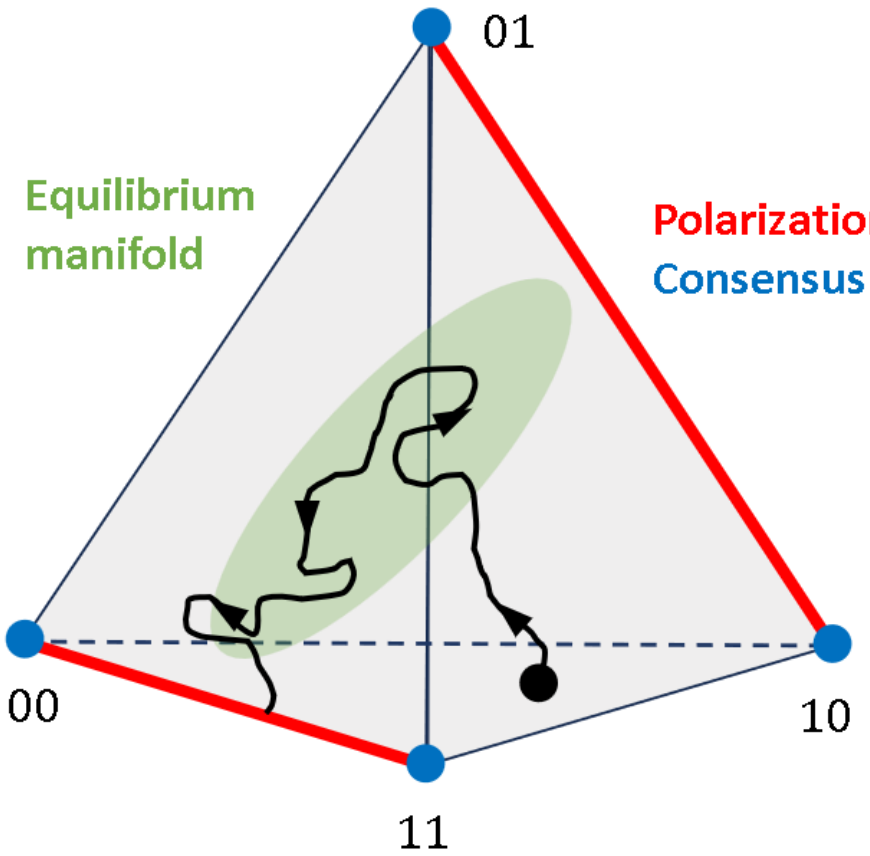


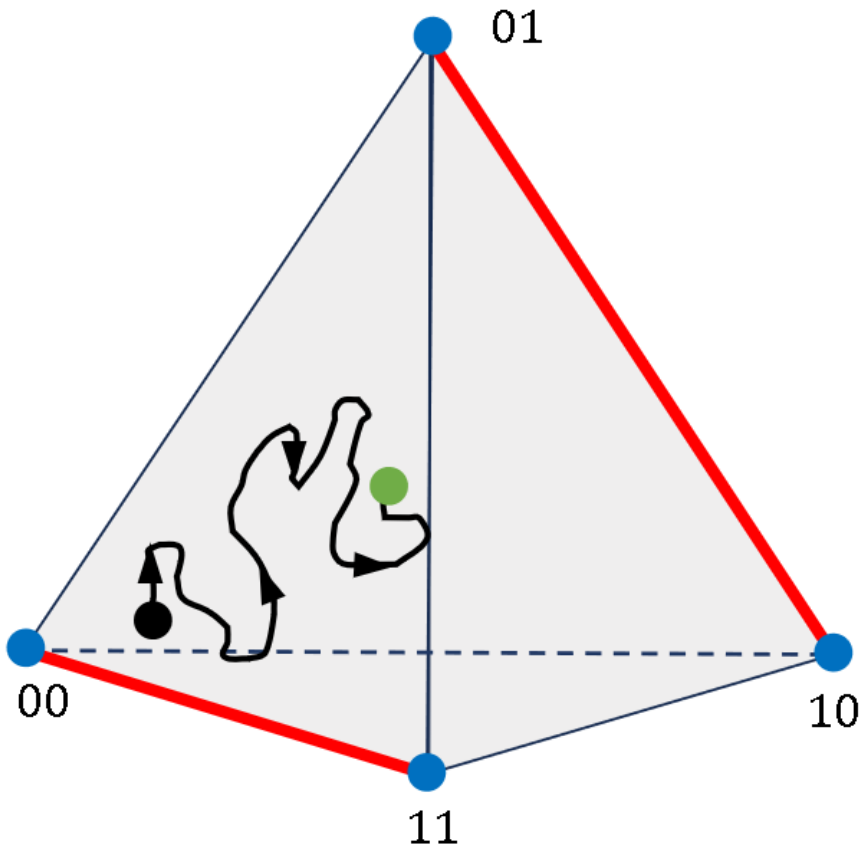
D



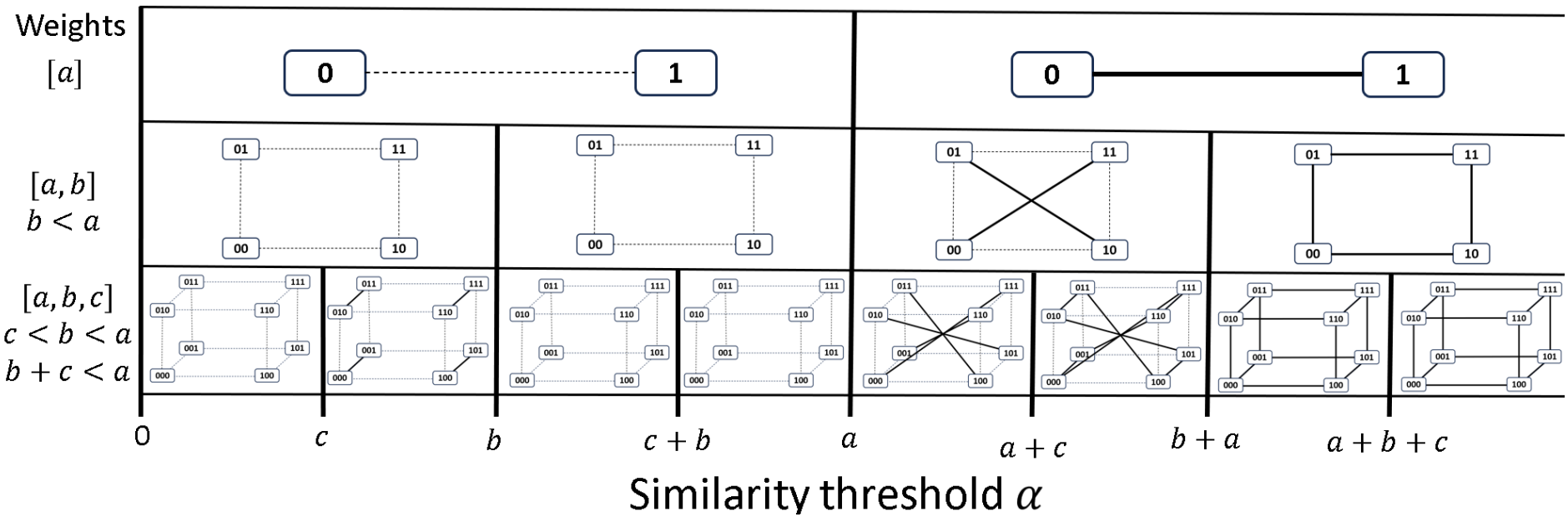
****

Explanatory Diagram





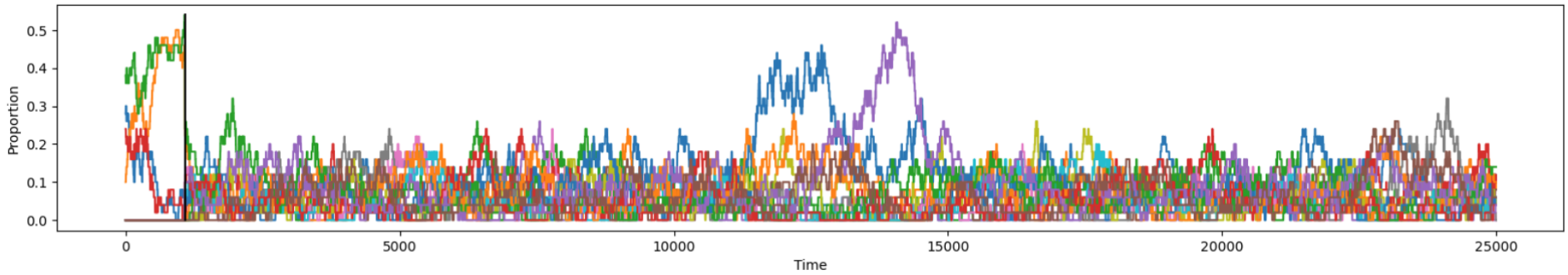
How symmetries build



We can change the symmetry type by introducing an issue, changing the properties of the opinion dynamics.

Small issues can have huge effects

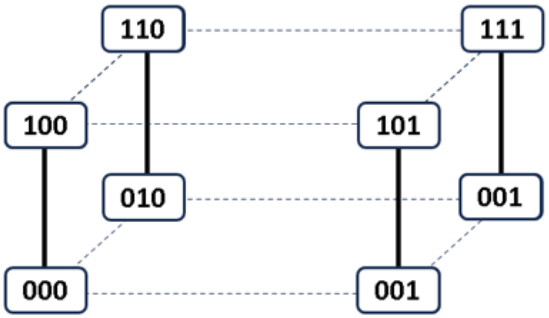
Adding the last weight to $[10002, 10001, 10000, 5]$ with $\alpha_f = \alpha_e = 10003$ can take a converged population and effectively prevent it from reaching convergence again.

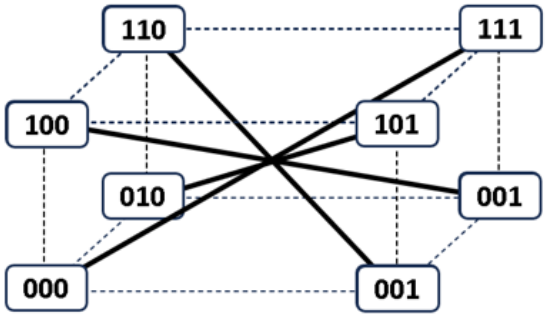


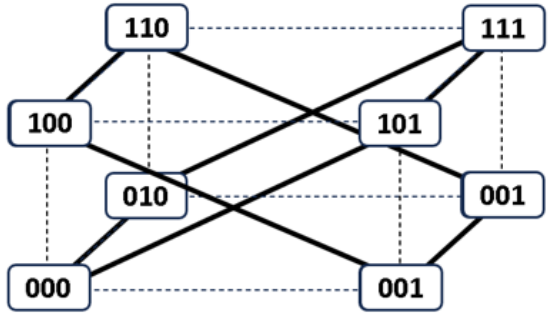
Similarly, adding the last weight to $[10008, 10004, 10000, 2]$ can create consensus if $\alpha = 10001$ or mitigate polarization if $\alpha = 10005$.

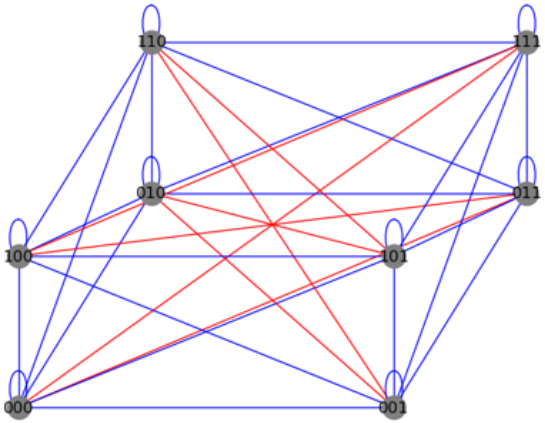
Respectively, the same effects can occur with homogeneous weights, going from $J = 3$ to 4 with $\alpha_f = \alpha_e = 1.5$.

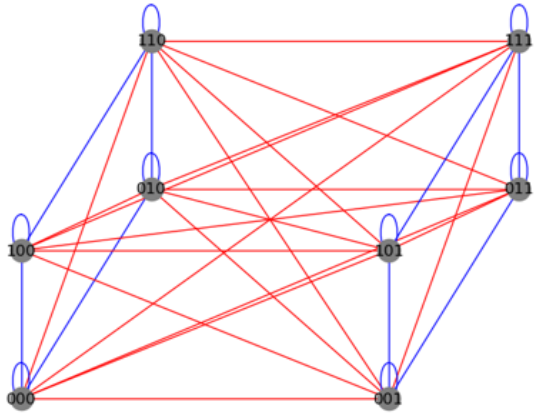
Answer 1: Relationship networks

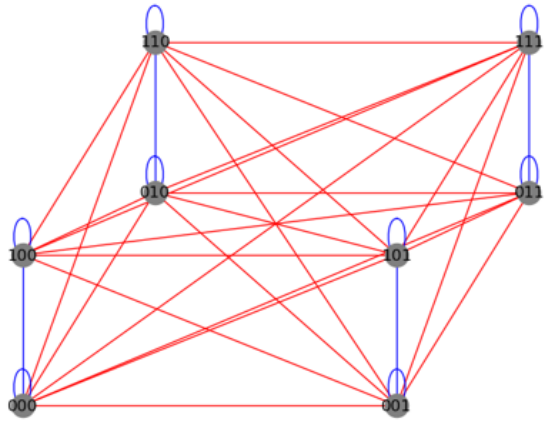


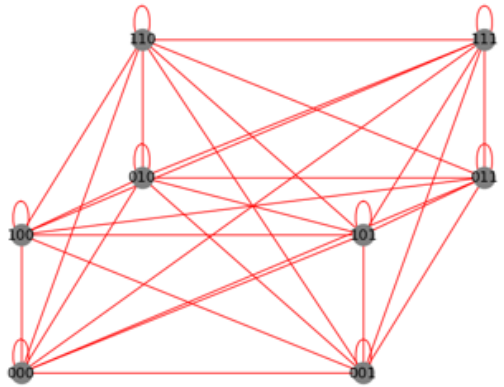


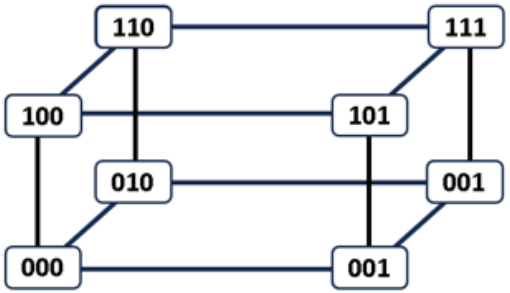


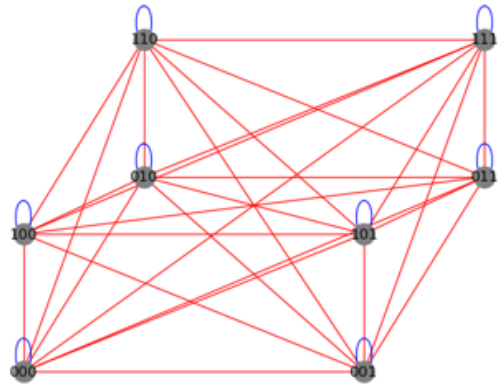


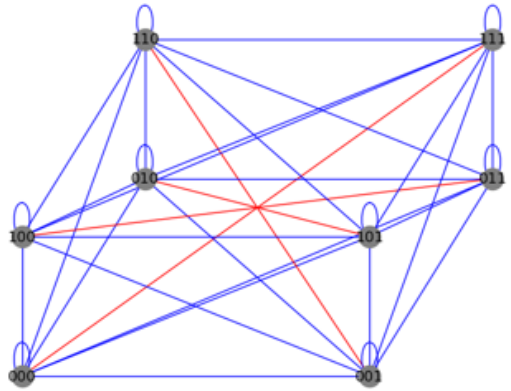


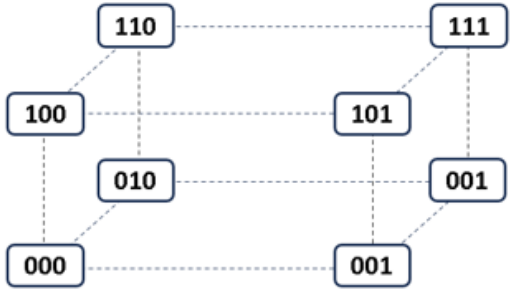


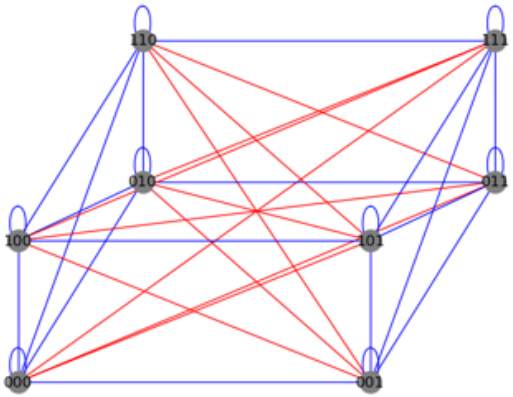

110
111
010
011
100
101
000
001

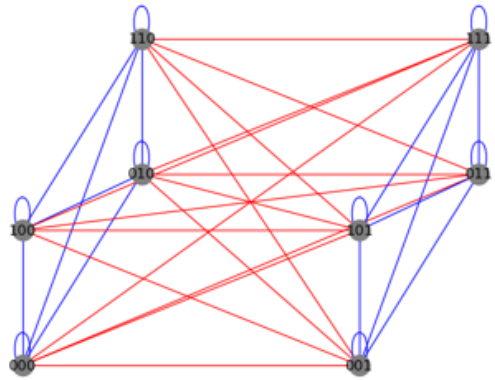

110
111
100
101
010
001
000
001


110
111
010
011
100
101
000
001


110
111
010
011
100
101
000
001


110
111
100
101
010
001
000
001


110
111
010
011
100
101
000
001


110
111
010
011
100
101
000
001

Answer 2: Why are there symmetries?

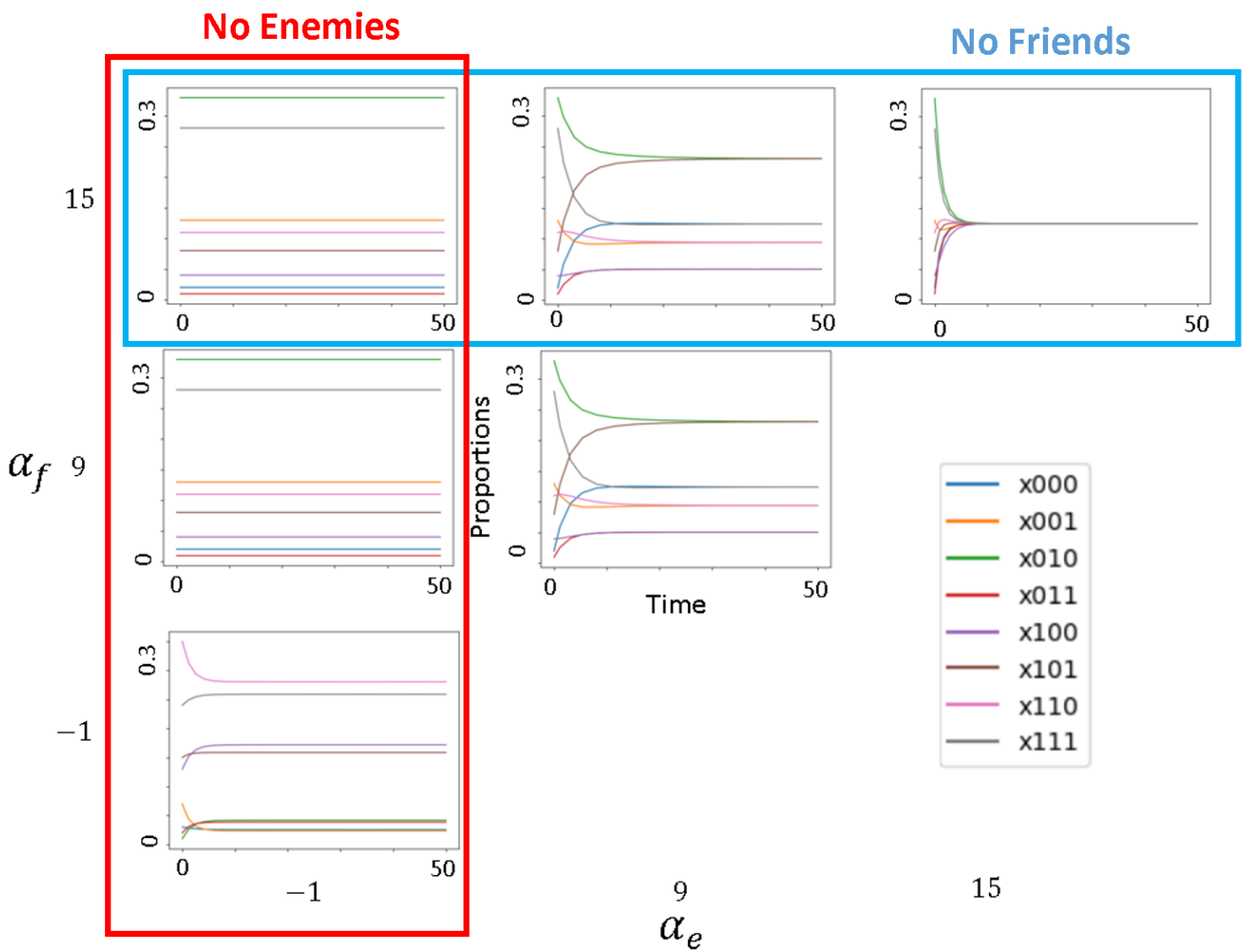
These symmetries in the equilibria appear to result from symmetries in the systems of equations, which depend non-trivially on the relationship network.

If $f(\sigma(\vec{x})) = -f(\vec{x})$ for some function $f(\vec{x})$ and permutation σ, then $\vec{x} = \sigma(\vec{x})$ implies

$$f(\vec{x}) = f(\sigma(\vec{x})) = -f(\vec{x}) \implies f(\vec{x}) = 0$$

This means points fixed by anti-symmetries of the dynamics are equilibria. We observe this in some cases but not all. Further, we find $F(\sigma(\vec{x})) = \sigma F(\vec{x})$ for the bit flip permutations, which quickly shows the uniform distribution is always an equilibrium.

Answer 3: Antagonism creates symmetry



Summary

Results:

- The similarity threshold has a non-monotonic relationship with absorption time, and a strikingly clear pattern across all cases: sympathetic interaction promote consensus.
- This results from symmetries in the equilibria of the deterministic model, caused by symmetries in the relationship network.
- A complete classification showed introducing arbitrarily small issues can effectively prevent convergence or promote consensus.

Next Steps:

- Consider issues proportional to weight, or multiple alternatives per issue.
- Further investigate when and why symmetries occur.
- Heterogeneous thresholds to see effects of antagonistic individuals.
- Allowing the number of opinions or weights to change over time.

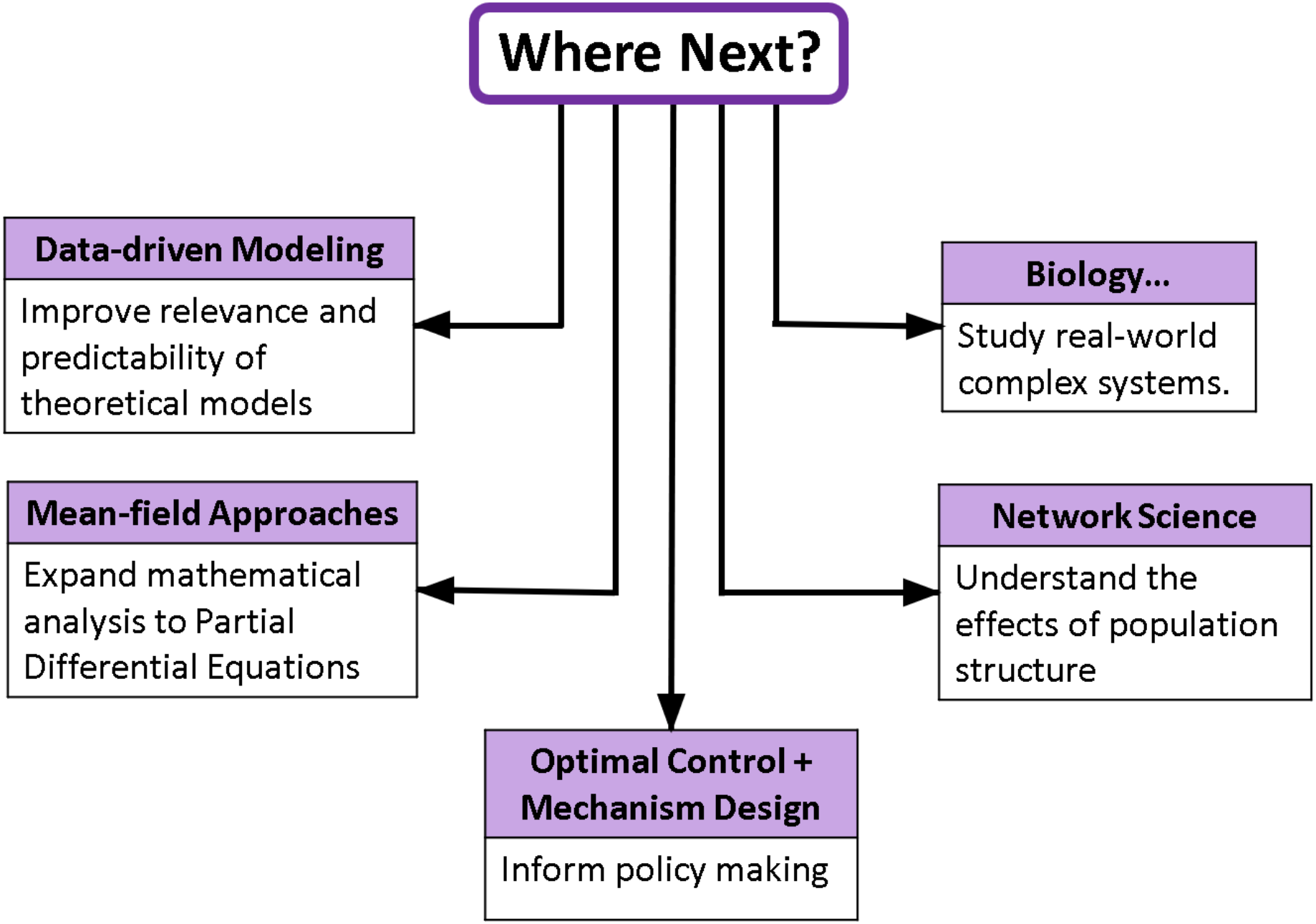
Table of Contents

1 Background

2 Exploration / Exploitation

3 Opinion Dynamics

4 Conclusion


Where Next?
Data-driven Modeling
Improve relevance and predictability of theoretical models
Biology...
Study real-world complex systems.
Mean-field Approaches
Expand mathematical analysis to Partial Differential Equations
Network Science
Understand the effects of population structure
Optimal Control + Mechanism Design
Inform policy making

Thank you to my collaborators!

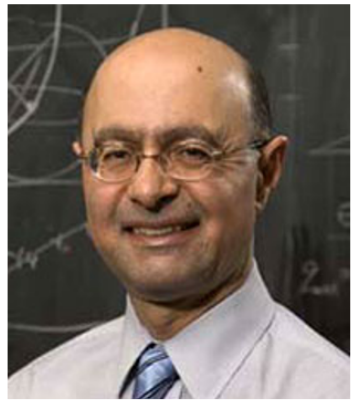

Natalia Komarova


Feng Fu


Dimitris Giannakis


David Freeman


Pin-Hao Andy Chen


Abbas Ourmazd


Joanna Slawinska


Dominik Wodarz


Feng-Chun Ben Chou


Olivia Chu


Alina Glaubitz


Ethan Levien

Thank you for coming to my defense!

(Website with this talk and other fun math stuff)

References I

Rediet Abebe, Jon Kleinberg, David Parkes, and Charalampos E Tsourakakis, *Opinion dynamics with varying susceptibility to persuasion*, Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018, pp. 1089–1098.

Claudio Castellano, Santo Fortunato, and Vittorio Loreto, *Statistical physics of social dynamics*, Reviews of modern physics **81** (2009), no. 2, 591–646.

Peter Clifford and Aidan Sudbury, *A model for spatial conflict*, Biometrika **60** (1973), no. 3, 581–588.

Morris H DeGroot, *Reaching a consensus*, Journal of the American Statistical association **69** (1974), no. 345, 118–121.

John Gittins, Kevin Glazebrook, and Richard Weber, *Multi-armed bandit allocation indices*, John Wiley & Sons, 2011.

References II


Richard A Holley and Thomas M Liggett, *Ergodic theorems for weakly interacting infinite systems and the voter model*, The annals of probability (1975), 643–663.


Jan Lorenz, *Continuous opinion dynamics under bounded confidence: A survey*, International Journal of Modern Physics C **18** (2007), no. 12, 1819–1838.


Stefanos Leonardos and Georgios Piliouras, *Exploration-exploitation in multi-agent learning: Catastrophe theory meets game theory*, Artificial Intelligence **304** (2022), 103653.


Hegselmann Rainer and Ulrich Krause, *Opinion dynamics and bounded confidence: Models, analysis and simulation*, Journal of Artificial Societies and Social Simulation **5** (2002), no. 3.


Aleksandrs Slivkins et al., *Introduction to multi-armed bandits*, Foundations and Trends® in Machine Learning **12** (2019), no. 1-2, 1–286.

References III


Alina Sîrbu, Vittorio Loreto, Vito DP Servedio, and Francesca Tria, *Opinion dynamics: models, extensions and external effects*, Participatory sensing, opinions and collective awareness (2017), 363–401.


Christopher JCH Watkins and Peter Dayan, *Q-learning*, Machine learning **8** (1992), 279–292.


Peter Whittle, *Multi-armed bandits and the gittins index*, Journal of the Royal Statistical Society: Series B (Methodological) **42** (1980), no. 2, 143–149.


Els Weinans, Patrick Steinmann, Elisa Perrone, Ahmadreza Marandi, and George AK van Voorn, *An exploration of drivers of opinion dynamics*, Journal of Artificial Societies and Social Simulation **27** (2024), no. 1, 5.