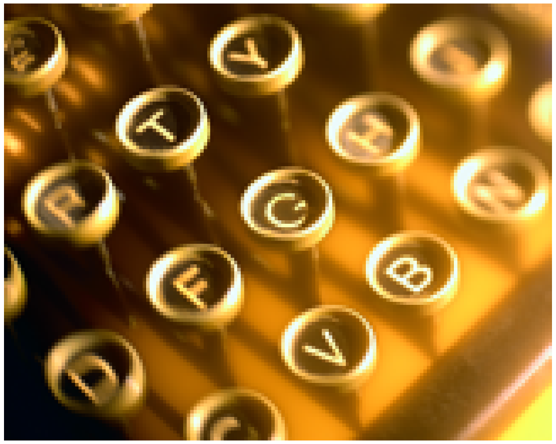
Learning Dynamics and Equilibrium Selection

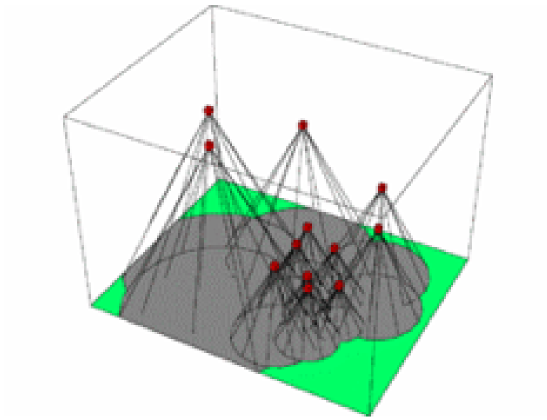
Jeff S Shamma
Georgia Institute of Technology

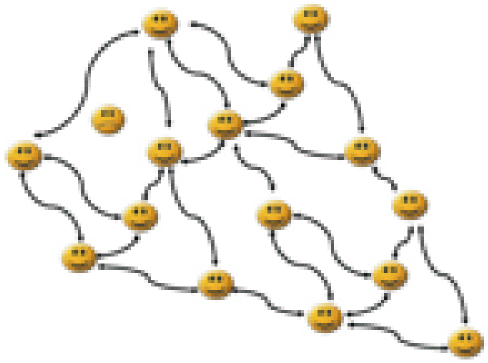
Joint work with
George Chasparis, Jason Marden, & Michael Fox

LCCC Workshop on Distributed Decisions via Games and Price Mechanisms
March 2010

Networked interaction: Societal, engineered, & hybrid







Equilibrium selection & dynamics

- How could agents converge to NE?

> *Arrow: "The attainment of equilibrium requires a disequilibrium process."*

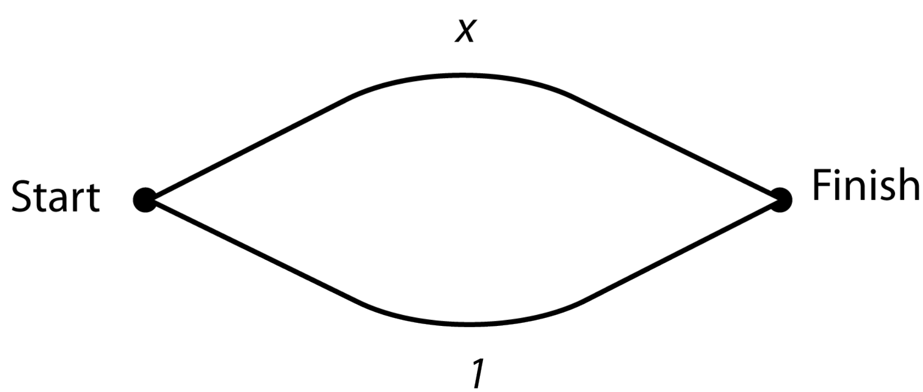
- Monographs:
  - Weibull, *Evolutionary Game Theory*, 1997.
  - Young, *Individual Strategy and Social Structure*, 1998.
  - Fudenberg & Levine, *The Theory of Learning in Games*, 1998.
  - Samuelson, *Evolutionary Games and Equilibrium Selection*, 1998.
  - Young, *Strategic Learning and Its Limits*, 2004.
  - Sandholm, *Population Dynamics and Evolutionary Games*, 2010.
- Surveys:
  - Hart, "Adaptive heuristics", *Econometrica*, 2005.
  - Fudenberg & Levine, "Learning and equilibrium", *Annual Review of Economics*, 2009.

- If agents self-organize to Nash equilibrium...
  - Price of Anarchy:

$$\frac{\text{Optimal global objective}}{\text{Worst case NE}} = \frac{\max_a G(a)}{\min_{a \in \text{NE}} G(a)}$$

  - Price of Stability:

$$\frac{\text{Optimal global objective}}{\text{Best case NE}} = \frac{\max_a G(a)}{\max_{a \in \text{NE}} G(a)}$$



What about dynamics?
Stable? Unstable? Stabilized? Destabilized?

	A	B
A	4,4	0,0
B	0,0	3,3

Typewriter Game

	S	H
S	3,3	0,1
H	1,0	1,1

Stag Hunt

- How to distinguish equlibria?
- Payoff based distinctions: Payoff dominance vs Risk dominance
- Evolutionary (i.e., *dynamic*) distinction
  - Young (1993) "The evolution of convention"
  - Kandori/Mailath/Rob (1993) "Learning, mutation, and long-run equilibria in games"
  - many more...
- Adaptive play:
  - "Two" players sparsely sample from finite history
  - Players either:
    - $*$ Play best response to selection
    - $*$ Experiment with small probability
  - **Young (1993):** Risk dominance is "stochastically stable"

Outline

- Dynamics & equilibrium selection theme continued...
  - Constrained log linear learning
  - Self assembly
  - Dynamic reinforcement dynamics
- "Prescriptive" issues & opportunities
  - What are implications of additional constraints?
  - How to exploit additional degrees of freedom?

- Setup:
  - Players: $\{1, ..., p\}$
  - Actions: $a_i \in \mathcal{A}_i$
  - Action profiles:

$$(a_1, a_2, ..., a_p) \in \mathcal{A} = \mathcal{A}_1 \times \mathcal{A}_2 \times ... \times \mathcal{A}_p$$

  - Payoffs: $u_i : (a_1, a_2, ..., a_p) = (a_i, a_{-i}) \mapsto \mathbf{R}$
  - Global objective: $G : \mathcal{A} \to \mathbf{R}$
- Action profile $a^* \in \mathcal{A}$ is a *Nash equilibrium* (NE) if for all players:

$$u_i(a_1^*, a_2^*, ..., a_p^*) = u_i(a_i^*, a_{-i}^*) \geq u_i(a_i', a_{-i}^*)$$

- Learning dynamics:
  - $t = 0, 1, 2, ...$
  - $\mathbf{Pr}\left[a_i(t)\right] = p_i(t), \quad p_i(t) \in \Delta(\mathcal{A}_i)$
  - $p_i(t) = \mathcal{F}_i($available info at time $t)$

Special class: potential games

- **Potential games:** For some $\phi : \mathcal{A} \to \mathbb{R}$

$$\phi(a_i, a_{-i}) - \phi(a_i', a_{-i}) > 0$$
$$\Leftrightarrow$$
$$u_i(a_i, a_{-i}) - u_i(a_i', a_{-i}) > 0$$

  i.e., potential function increases iff unilateral improvement.

- Features:
  - Typical of "coordination games"
  - Desirable convergence properties under various algorithms
  - Need not imply "cooperation" or $\phi = G$
  - Prescriptive opportunity: Potential game by design

- Distributed routing
  - Payoff = negative congestion. $c_r(\sigma_r)$
  - Potential function:

$$\phi = \sum_r \sum_{n=1}^{\sigma_r} c_r(n)$$

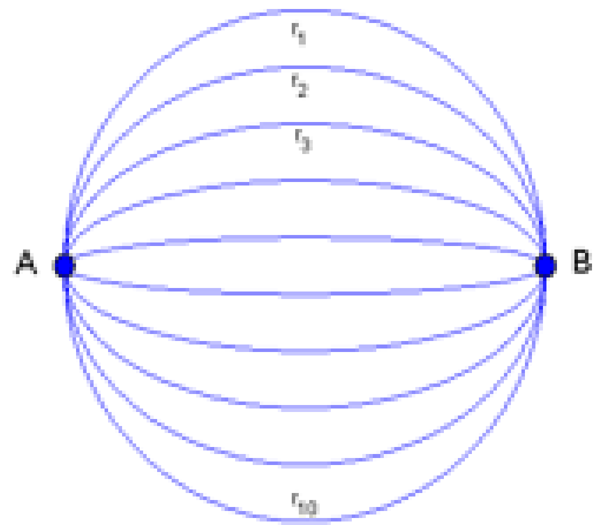
  - Overall congestion:

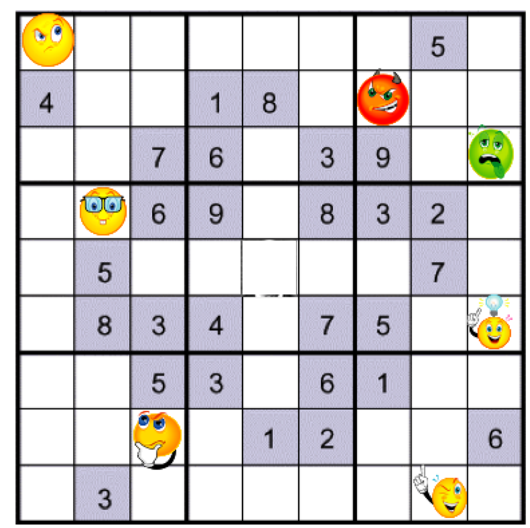
$$G = \sum_r \sigma_r c_r(\sigma_r)$$

  - **Note:** $\phi \neq G$
- Multiagent sudoku:

$u_i(a) =$#reps in row + #reps in column + #reps in sector

$$\phi(a) = \sum_i u_i(a)$$





- Preliminary: Gibbs distribution

$$\mathbf{Pr}\left[v_i\right] \propto e^{v_i/T}$$

As $T \downarrow 0$ assigns all probability to $\arg\max\{v_1, v_2, ..., v_n\}$

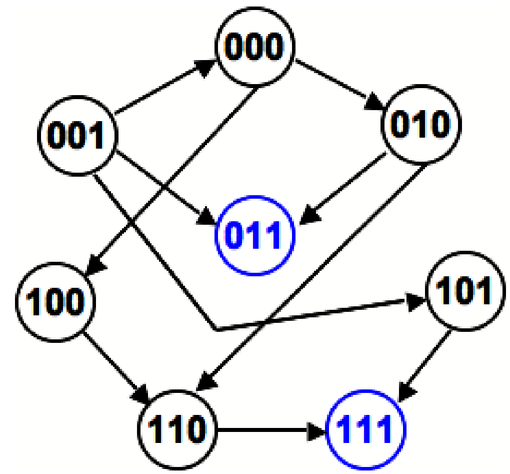
- At stage t
  - Player i is selected at random
  - Chosen player selects

$$\mathbf{Pr}\left[a_i(t) = j\right] \propto e^{u_i(j, a_{-i}(t-1))/T}$$

  - Interpretation: Noisy best reply to previous joint actions
- Fact: SAP results in a Markov chain over joint action space $\mathcal{A}$ with a unique stationary distribution, μ.
- **Blume (1993)**: In (cardinal) potential games, steady-state distribution satisfies

$$\mathbf{Pr}\left[a\right] \propto e^{\phi(a)/T}$$

  - Implication: As $T \downarrow 0$, concentrated at potential maximizer
  - i.e., Potential maximizer is "stochastically stable"



Log linear learning with constraints

- Impose constrained evolution:

$$a_i(t) \in C(a_i(t-1))$$

  - Limited mobility
  - Obstacles

- Mimicking log linear learning *alters* stochastically stable action profiles!
- Example: Identical interest game

	L	M	R
U	0	0	9
D	10	-10	-10

$$C_2(L) = \{L, M\} \quad C_2(M) = \{L, M, R\} \quad C_2(R) = \{M, R\}$$

  - Potential maximizer: (D, L)
  - Stochastically stable state: (U, R)
  - Intuition:
    * $(U, R) \rightarrow (D, L)$ “costs” 18 (used to cost 9)
    * $(D, L) \rightarrow (U, R)$ “costs” 10 (used to cost 10)

- At stage t:
  - Player i is selected at random
  - Chosen player compares $a_i(t-1)$ with randomly selected $a_i' \in C(a_i(t-1))$

$$\mathbf{Pr}\left[a_i(t)\right] \propto e^{u_i(a_i(t-1),a_{-i}(t-1))/T} \quad \text{vs} \quad e^{u_i(a_i',a_{-i}(t-1))/T}$$

- **Marden & JSS, 2008:** Under binary log linear learning, only potential function maximizers are stochastically stable.
- *No longer* characterize stationary distribution.
- Recall example:

	L	M	R
U	0	0	9
D	10	-10	-10

- Binary version: $(U, M) \to (U, L)$ now has zero resistance.

Payoff based log linear learning

- What if evaluation of $U_i(a', a_{-i}(t-1))$ no longer possible?
- New setup: Players can only measure $a_i(t)$ and $U_i(a(t))$
- Introduce *baseline action* $a_i^b(t)$ and *baseline utility* $u_i^b(t)$
- Action selection:

$$a_i(t) = a_i^b(t) \text{ with probability } (1-\epsilon)$$

$$a_i(t) \text{ is chosen randomly over } \mathcal{A}_i \text{ with probability } \epsilon$$

- Baseline action & utility update:

New baseline	*Keep baseline*
with probability $\sim e^{U_i(a(t))/T}$	with probability $\sim e^{u_i^b(t)/T}$
$\Downarrow$	$\Downarrow$
$a_i^b(t+1) = a_i(t)$	$a_i^b(t+1) = a_i^b(t)$
$u_i^b(t+1) = u_i(a(t))$	$u_i^b(t+1) = u_i^b(t)$

- **Marden & JSS, 2008:** Under payoff based linear learning, only potential function maximizers are stochastically stable.

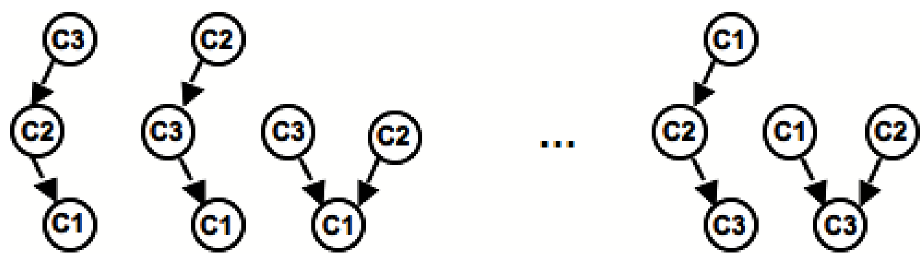
- Definition:
  - Let P^ϵ denote the transition probability matrix of an irreducible & aperiodic Markov chain.
  - Let μ^ϵ be the (unique) stationary distribution for P^ϵ
  - A state, x, is **stochastically stable** if

$$\liminf_{\epsilon \to 0} \mu^\epsilon(x) > 0$$

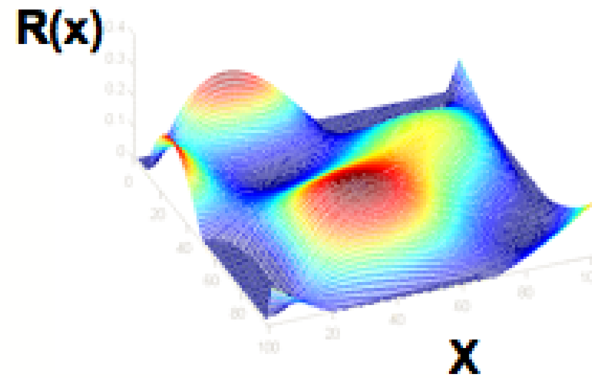
- **Young (1993)**: To determine stochastic stability
  - View learning dynamics as ϵ perturbation of reference ($\epsilon = 0$) Markov chain
  - Divide reference Markov chain into recurrence classes
  - Define *resistance* to transition between recurrence classes:

$$0 < \lim_{\epsilon \downarrow 0} \frac{P_{ij}^\epsilon}{\epsilon^{r(i \to j)}} < \infty$$

  - Form *stochastic potential* for each recurrence class
  - Minimal stochastic potential implies stochastic stability



- Combinatoric utilization vs pragmatic utilization

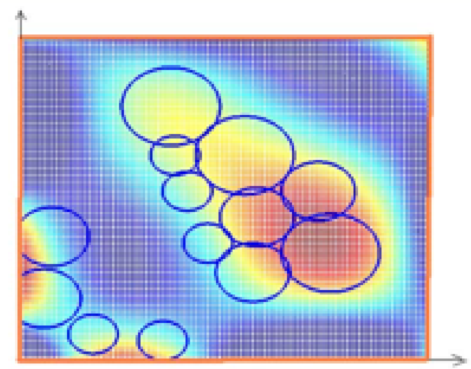


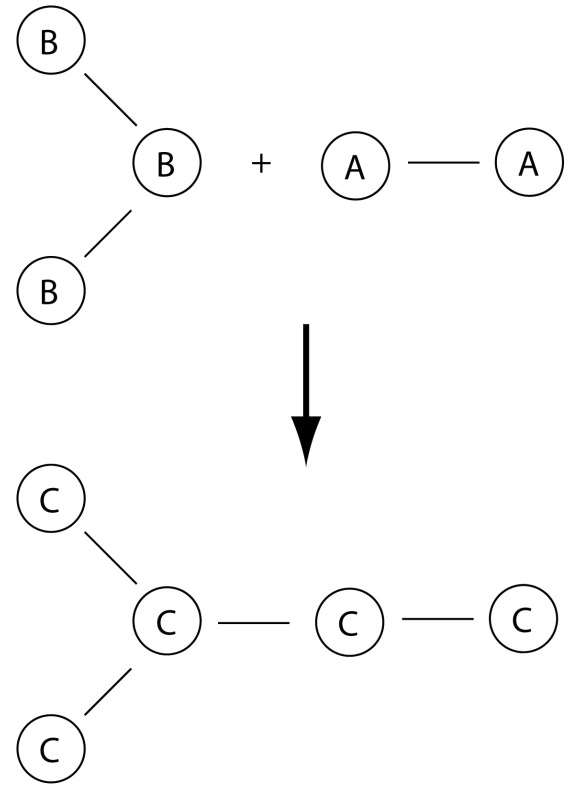
- Objective: Maximize expected reward

$$\phi(a) = \sum_{x} R(x) P(x,a)$$

$$P(x,a) = 1 - \Pi_{i=1}^{n}(1 - p_i(x, a_i))$$

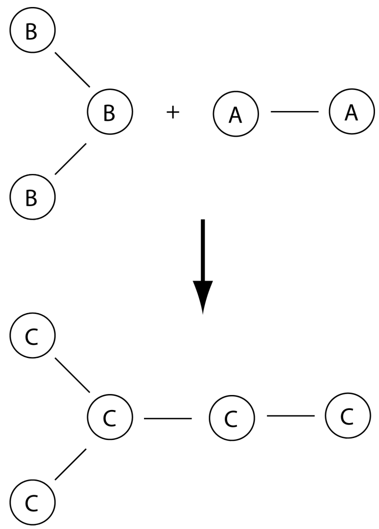
- Implementation:
  - Assign sensor utilities to induce potential game
  - Apply constrained binary log-linear learning

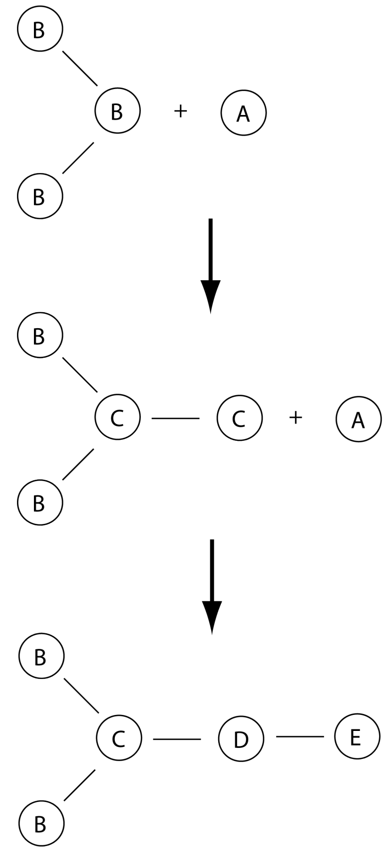




- Atoms form subassemblies.
- Subassemblies form complete assemblies.

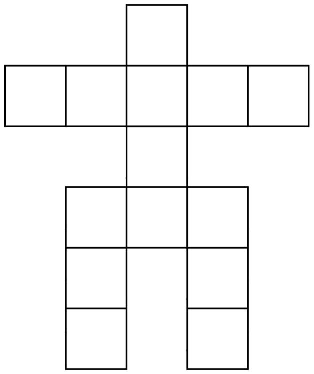
- References:
  - Yim, Shen, Salemi, Rus, Moll, Lipson, Klavins, & Chirikjian, “Modular self-reconfigurable robot systems: Challenges and opportunities for the future”, 2007.
  - Klavins, “Programmable self-assembly”, 2007.

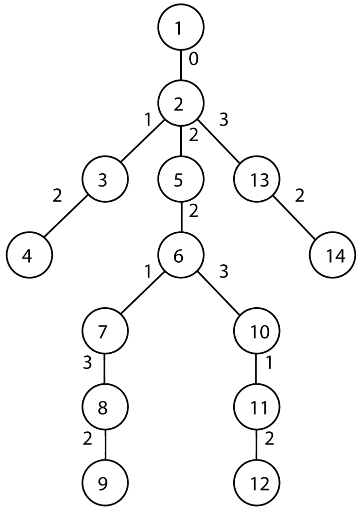


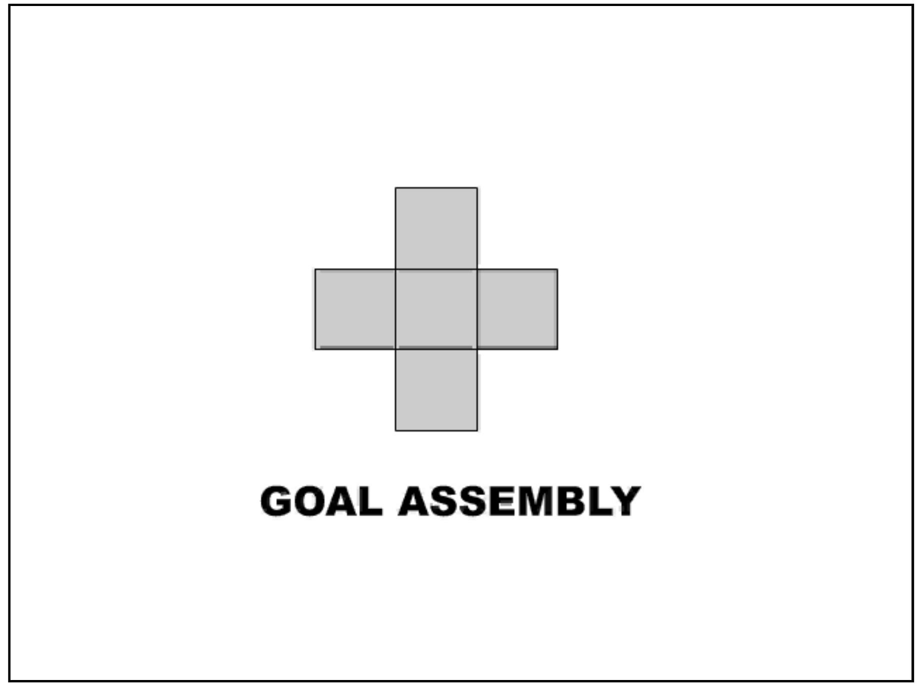


- General setup:
  - Nonlocal rules
  - Full “graph grammars”

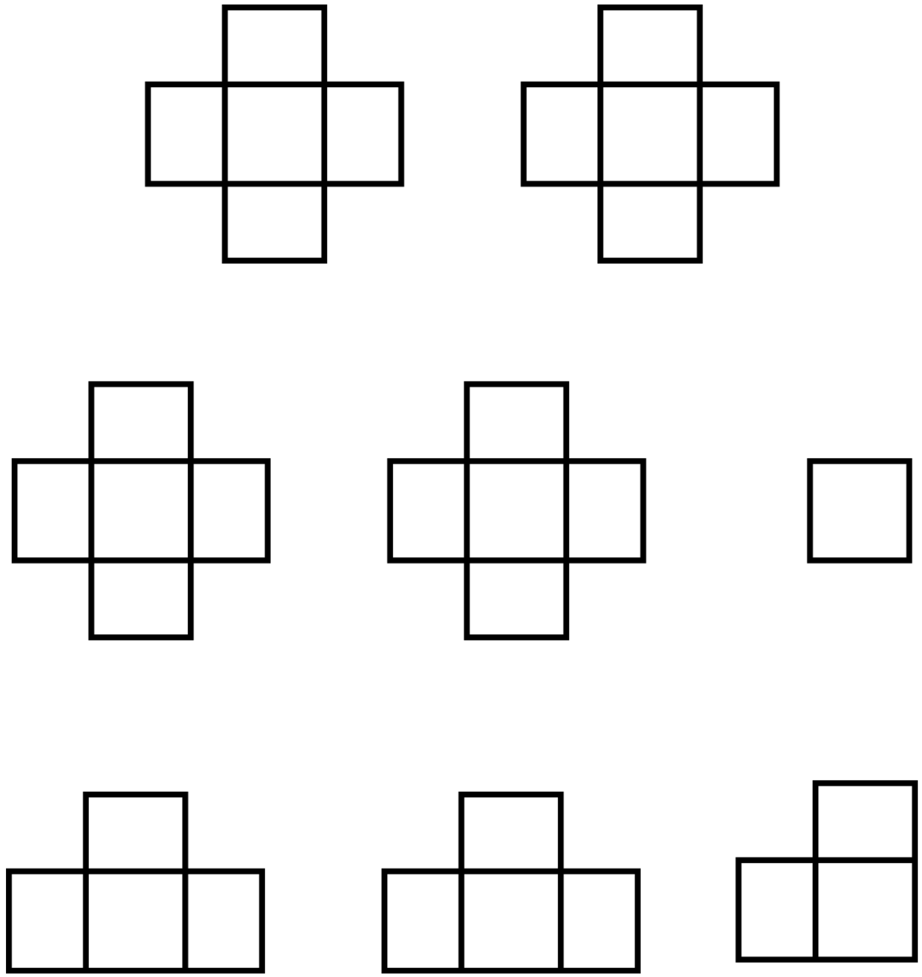
- Specialized setup:
  - Serial assembly
  - Local rules
  - Bond or break
  - Reversibility



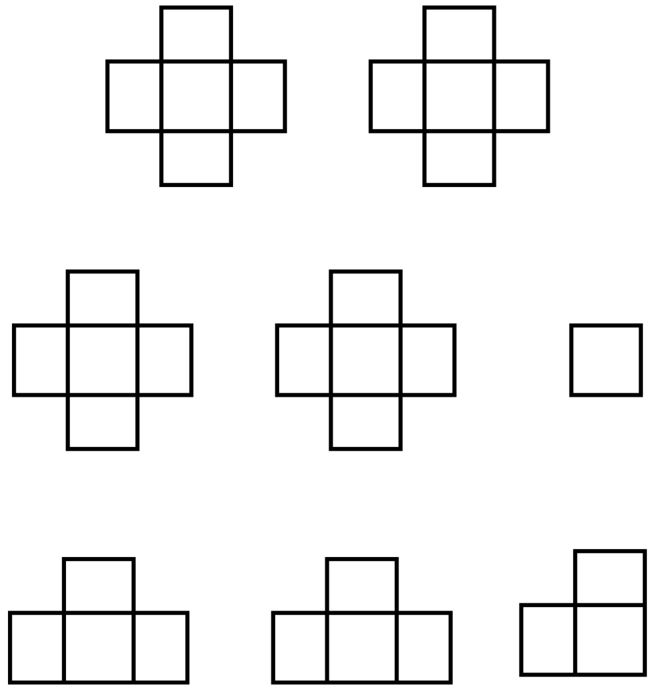




- Complete assembly = Acyclic weighted graph
- Node state: (Position, Vacancies)
- Nodes meet randomly
- If singleton meets vacancy: Active nodes update state
- Singletons break off with probability ϵ

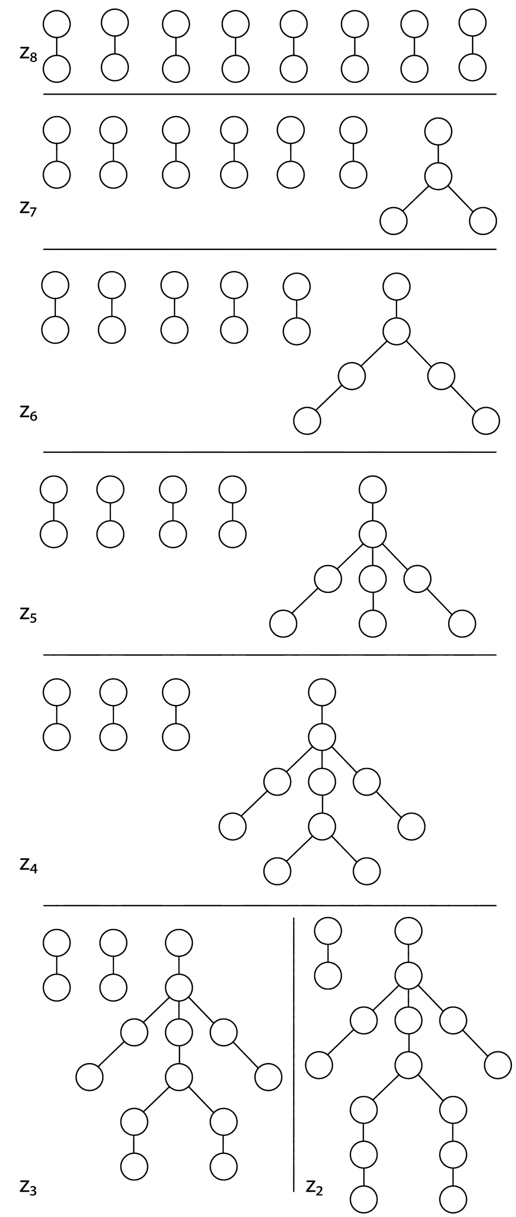


Critical case: #Atoms = Integer multiple of final assembly



- **Theorem (Fox & JSS, 2009):** A state is stochastically stable if and only if there is a minimal number of (sub)assemblies.
- Corollary: Let a complete assembly have N parts. The maximum number of incomplete assemblies is $N-1$. (For any number of atoms.)

- Form a “backbone” of states with m subassemblies
- Level down: Resistance of 1
- Level up: Resistance at least 2



- Reinforcement learning: $x_i =$ action propensities

$$x_i(t+1) = x_i(t) + \delta(t)(a_i(t) - x_i(t)), \quad \delta(t) = \frac{u_i(a(t))}{t+1}$$

$$p_i(t) = (1-\varepsilon)x_i(t) + \frac{\varepsilon}{N}\mathbf{1}$$

$$\delta_{\text{std}}(t) = \frac{u_i(a(t))}{\mathbf{1}^{\mathrm{T}}U_i(t) + u_i(a(t))}$$

Interpretation: Increased probability of utilized action.

- *Dynamic* reinforcement learning: Introduce running average

$$y_i(t+1) = y_i(t) + \frac{1}{t+1}(x_i(t) - y_i(t))$$

$$p_i(t) = (1-\varepsilon)\Pi_\Delta\left[x_i(t) + \underbrace{\gamma(x_i(t) - y_i(t))}_{\text{new term}}\right] + \frac{\varepsilon}{N}\mathbf{1}$$

Marginal foresight dominance

- **Chasparis & JSS (2009):** The pure NE a^* has positive probability of convergence iff

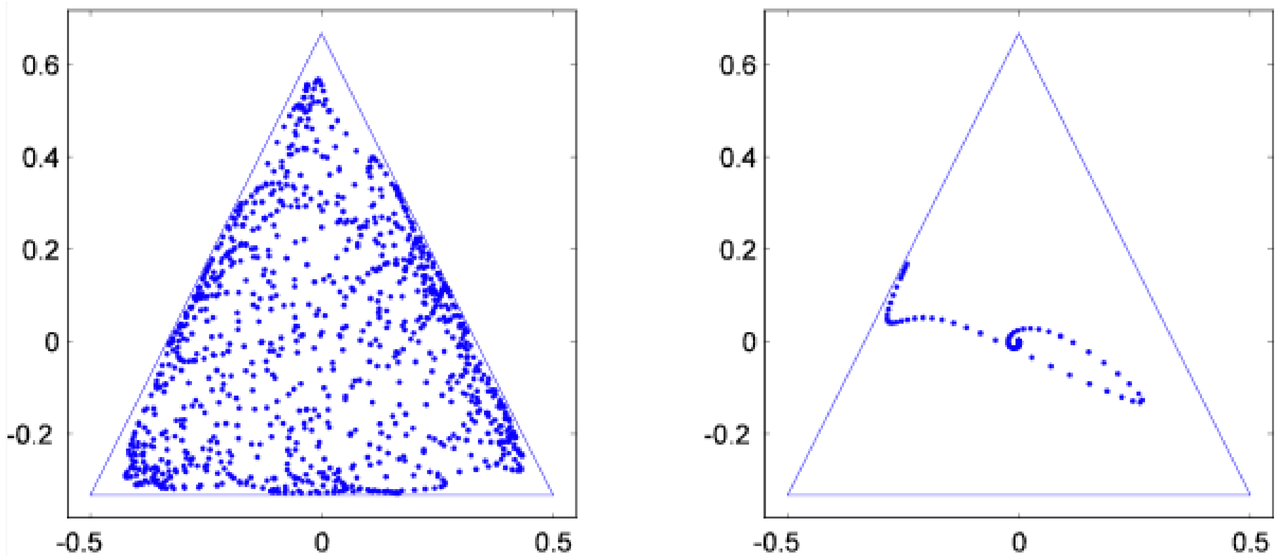
$$0 < \gamma_i < \frac{u_i(a_i^*, a_{-i}) - u_i(a_i', a_{-i}^*) + 1}{u_i(a_i', a_{-i}^*)}, \quad \forall a_i' \neq a_i^*$$

  (as opposed to all pure NE)
  Proof: ODE method of stochastic approximation.

- Implication:
  - Introduction of "forward looking" agent can destabilize equilibria
  - Surviving equilibria = equilibrium selection

- For 2×2 symmetric coordination games
  - RD & not PD $\Rightarrow$ foresight dominance
  - RD & PD & Identical interest $\Rightarrow$ foresight dominance
  - RD & PD together $\not\Rightarrow$ foresight dominance

- Similar ideas can *stabilize* equilibria (Arslan & JSS)
- Illustration: Perturbed RPS & Marginal foresight replicator dynamics



$$\dot{q}_1^j = \left(e_j^\top M_{12}(q_2 + \gamma\dot{r}_2) - q_1^\top M_{12}(q_2 + \gamma\dot{r}_2)\right) q_1^j$$

$$\dot{q}_2^j = \left(e_j^\top M_{21}(q_1 + \gamma\dot{r}_1) - q_2^\top M_{21}(q_1 + \gamma\dot{r}_1)\right) q_2^j$$

$$\dot{r}_1 = \lambda(q_1 - r_1)$$

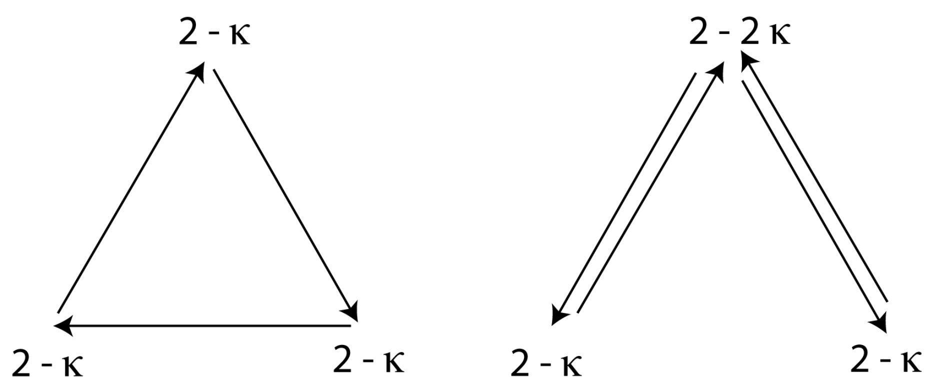
$$\dot{r}_2 = \lambda(q_2 - r_2)$$

$$\max_i \frac{a_i}{a_i^2 + b_i^2} < \frac{1}{\max_i a_i}$$

- Setup:
  - Agents form costly links with other agents
  - Benefits inherited from connectivity

$$u_i(a(t)) = \Big(\text{\# of connections to } i\Big) - \kappa \cdot \Big(\text{\# of links by } i\Big)$$

- Properties:
  - Nash networks are "critically connected"
  - Wheel network is unique *efficient* network
  - **Chasparis & JSS (2009):** The wheel network is foresight dominant.



Final remarks

- Recap:
  - Dynamics & equilibrium selection
  - Prescriptive agenda influence
- Future work:
  - Convergence rates
  - Fully exploit prescriptive agenda
  - Agent dynamics