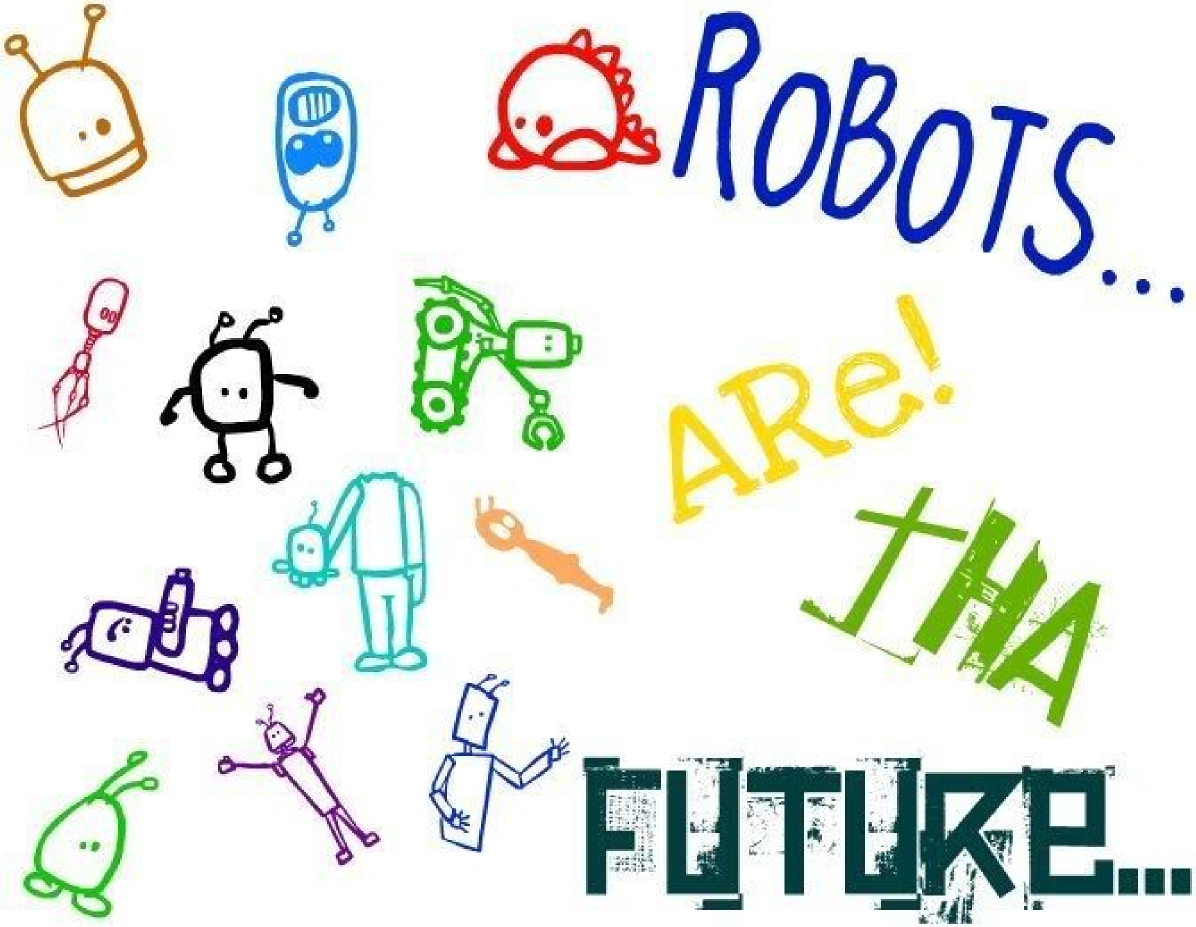
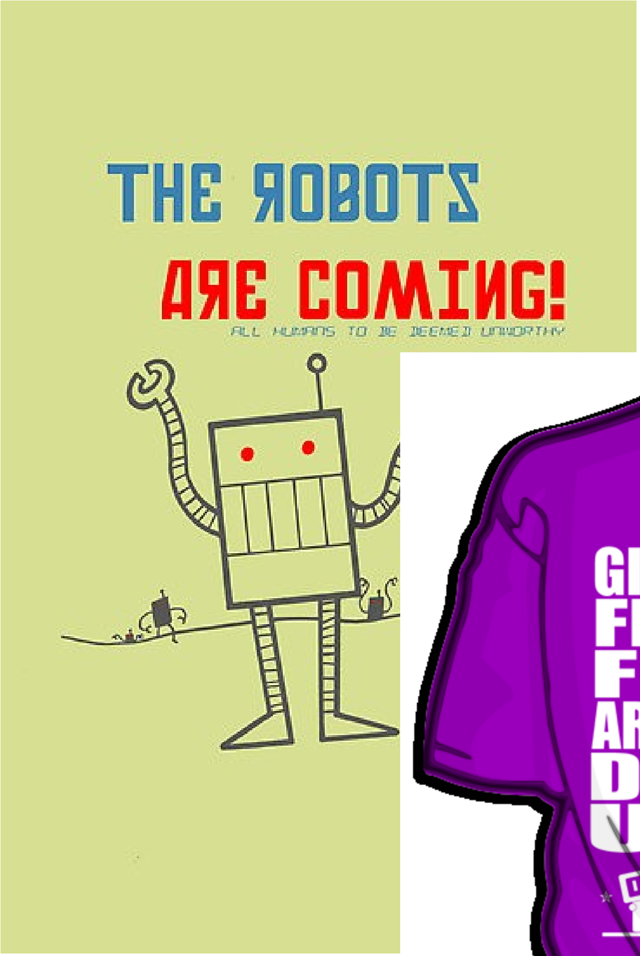
Games Robots Play

Hamming Seminar Series
23 November 2011

Subramanian Ramamoorthy
School of Informatics
The University of Edinburgh

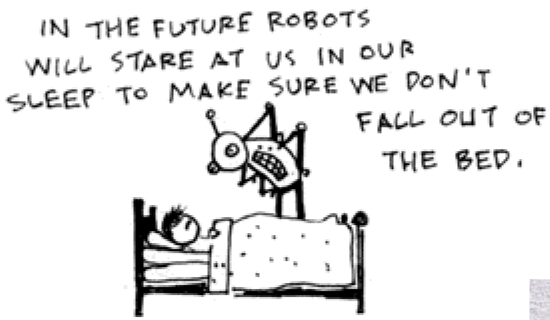

ROBOTS...
ARe!
tHA
FUTURe...

What Kind of a Future?



What did the Future Look Like in the Past?

What Could Robots be Used For in Future?



Pub of future will know your order

Microchips embedded under the skin will enable drinkers to order their favourite tipple before they reach the bar in the future, a technology specialist has said.

Ben Hourahaine, a future trends analyst, claims that in the 23rd century, cash will be obsolete because as soon as a customer walks into a pub, a reader above the door will scan their microchip, note their order and deduct a payment.

Beer pumps will pour a perfect pint tailored to each drinker's requirements.

Mr Hourahaine said: "Technological innovations will only enhance the customer's experience."

What do People **Really** want from Robots?

1. Autonomy: Ability to do interesting things, *without hand-holding*, in complex environments

2. Robust & Flexible Interaction: Ability to fluidly adapt, to *maintain functionality*, despite continual changes in tasks and environments

Where are those robots?!

Some Real Success Stories

Technical Achievement: Dexterity

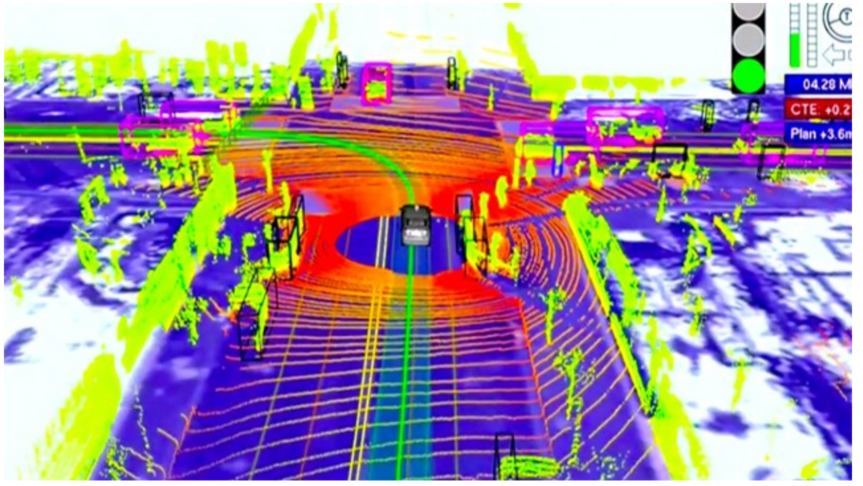
- Over past couple of decades, due to pioneering work by people like Raibert, we have a rich, multi-disciplinary understanding of locomotion, etc.

- Major insight was that control and adaptation are ‘easy’ if we have the ‘right’ structure and morphology

Technical Achievement: Uncertainty

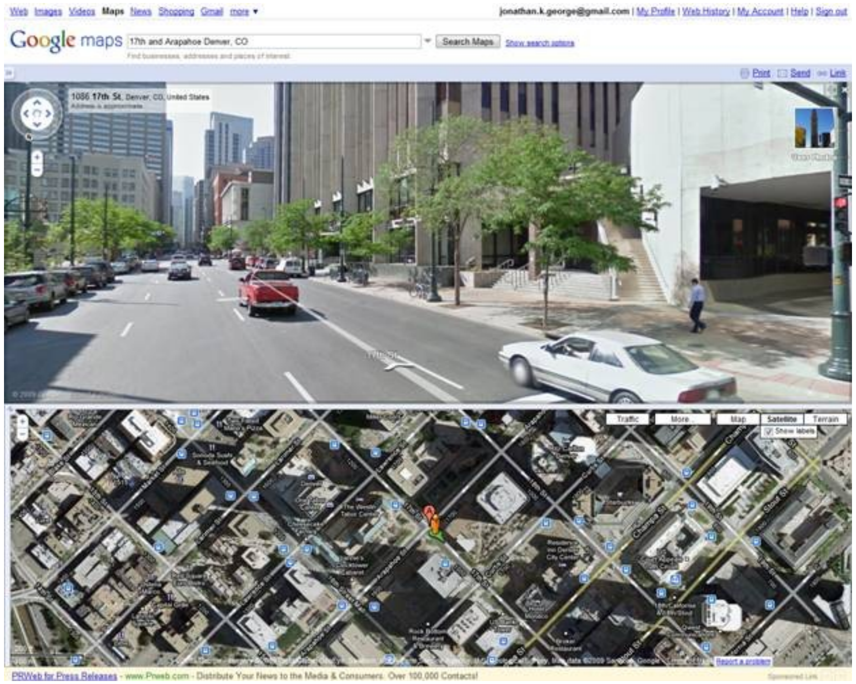
- Major advances came from recognizing that our models are inadequate without a description of uncertainty in sensing and actuation

- Excellent example: Simultaneous Localization and Mapping



Technical Achievement: Systems Integration

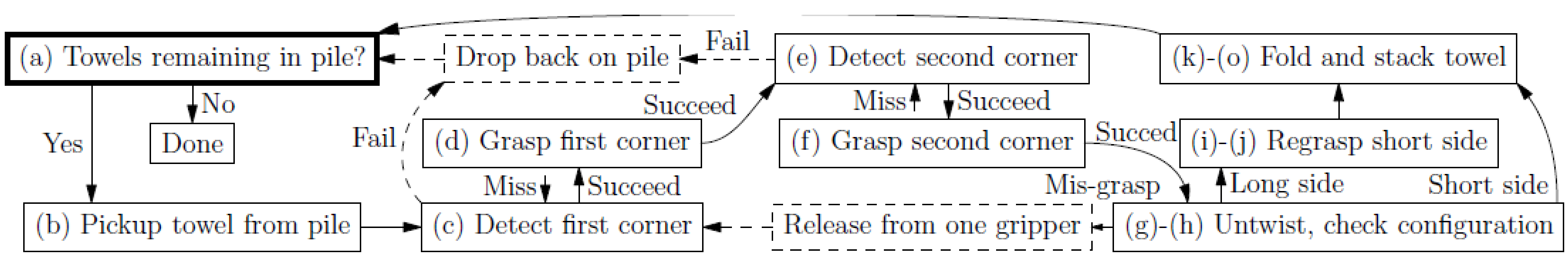
- Consider one example – Google Car
- It utilizes a vast arsenal of computing technology:
  - Network enabled access to Maps and Earth scans
  - Cloud computing to store massive volumes of data
  - Real time systems of high data rate sensors



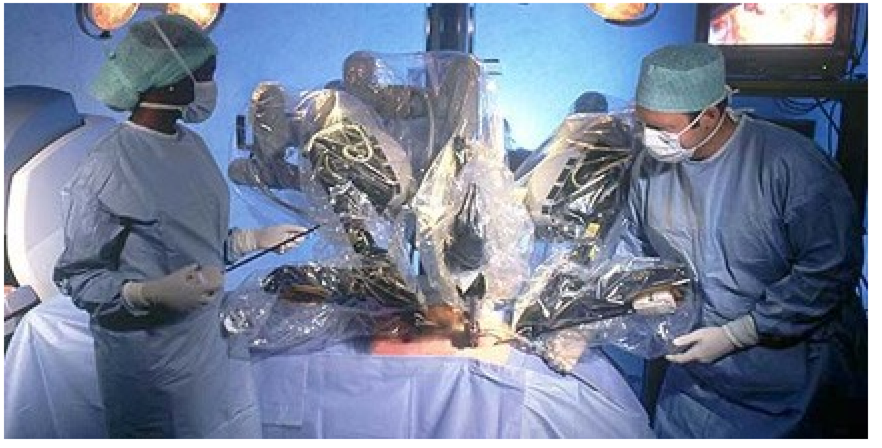
So, Are we Done?

Consider, apart from cost, the following (truth in advertising):

- Every demo with Asimo happens as per Honda approved scripts, in approved environments, by approved personnel
- How is the Google Car actually used?
  - Engineers collect detailed maps elaborating terrain to within mm/cm
  - Google engineer drives ahead just before Google Car is allowed to go
  - Google Car utilizes rich sensing, recently collected scans, etc. to localize itself and follow its stated paths
- Towel folding strategy:



Robotics Success, Type 1:
Focus on dexterity and motor control; avoid 'autonomy'



Q. Is a surgeon using the da Vinci Surgical System operating in "virtual reality"?

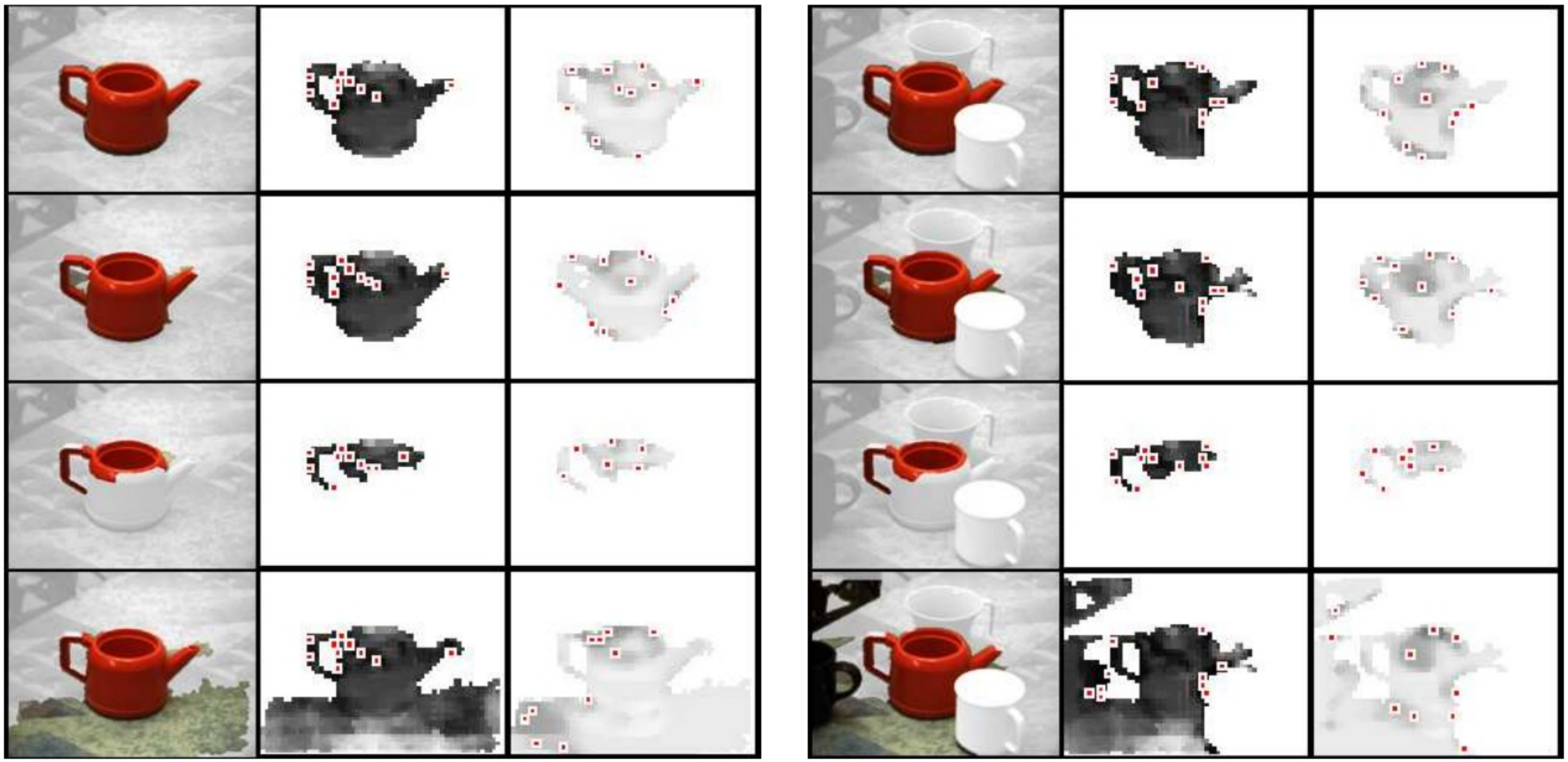
A. No. Although seated at a console a few feet from the patient, the surgeon views an actual image inside the patient's body while operating in real-time. At no time does the surgeon see a virtual image, or program the system to perform any independent maneuvers outside of the surgeon's direct, real-time control.

Q. Is this "robotic surgery?"

A. Robotic surgical devices are designed to perform entirely independent movements after being programmed by a surgeon. The *da Vinci* Surgical System is a computer-enhanced system that introduces a computer interface and 3DHD vision system between the surgeon's eyes, hands and the tips of micro-instruments. The system mimics the surgeon's hand movements in real time. It cannot be programmed, nor can it make decisions on its own to move or perform any type of surgical maneuver. So, while the general term "robotic surgery" is often used to refer to our technology, it is not robotic surgery in the strictest sense of the term.

[Source: FAQs on Intuitive Surgical's web site]

Robotics Success, Type 2: Characterize variability carefully; keep action selection simple



(a) Single Teapot (b) Partly Occluded Teapot

Where is the Boundary Between the Possible and the *Im*possible, today?

High level of variability + rich endogenous dynamics + incompleteness in models of change

"…Sometimes, however, the car has to be more aggressive*. When going through a four-way intersection, for example, it yields to other vehicles based on road rules; but if other cars don't reciprocate, it advances a bit to show to the other drivers its intention. Without programming that kind of behavior, Urmson said, it would be impossible for the robot car to drive in the real world."*

[How Google's Self-Driving Car Works, Discover News, Oct 18, 2011]

What Does his Mean for Real World Autonomy?

- We have often succeeded by cleverly avoiding the unexpected and the unmodelled – (how) can we move towards dealing with it too?

- This is especially key when we couple systems in novel ways – when endogenous dynamics become intricate

An Observation from My Time in Industry

- National Instruments vs. HP/Agilent
- By complete modularization, enable a new model of experimentation, the "Return of Edison"
- What makes this possible:
  - Diverse and heterogeneous components, enabled for interaction
  - Sophisticated tools for composition
- Acknowledge *ad hoc* nature of system design
  - *Zeitgeist* in broader technology space!

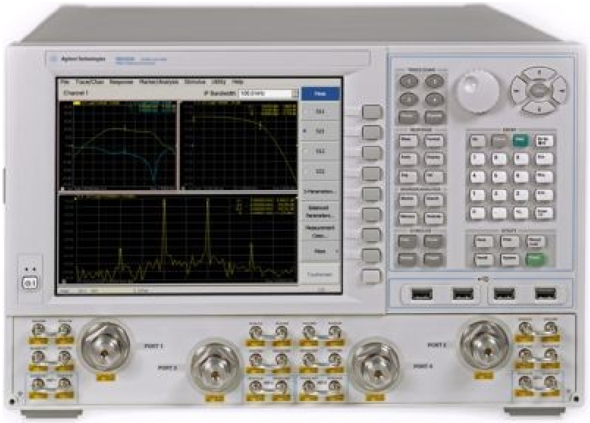


Figure 1

The X Series of multifunction data acquisition devices for PCI Express and PXI Express is comprised of 16 new X Series DAQ devices, which provide enhancements to analog I/O, digital I/O, onboard counters and multidevice synchronization.

An *ad hoc* Approach to Robotics

- Utilize a **diverse, heterogeneous team** of moderately capable, well understood components (Nao, iRobot, Kinect, etc.)
  - Acknowledge that even more complex systems are eventually limited, especially when we bound computation/information/communication
- **Continual adaptation of skills** among individuals (learning, app store like tweaks)
  - No two modules are identical, in the limit, or static; how should we represent knowledge in this setting?
- **Strategic interaction mechanisms** to enable efficient re-configurability in the face of incomplete models (and incompatible fragments), coarse specifications, conflicting preferences and goals, etc.

My *Personal* View of What to Aspire to

We need more of this!

In addition to, and
while we wait for…

Ad Hoc Human-Robot Teams in Virtual Worlds?

Problem: Interactive Decision Making:

Ability for agents to achieve long-lived autonomy

- *in interactions with a* ***continually changing environment*** *(partly, due to the* ***endogenous dynamics*** *of interaction),*
- *with* ***coarse high level goals****,*
- *with* ***incomplete information*** *regarding changes at level of*
  - *components (e.g., drift or failure)*
  - *structural issues (e.g., new participant or environment)*

Decision Making under Uncertainty

The traditional starting point for rational choice is maximization of expected utility

- Define a set of possible acts and states
- Define a preference order over outcomes –utility functions
- Choose acts to maximize utility

Table 1. Savage's example illustrating acts, states, and consequences

Act	State	
	Good	Rotten
Break into bowl	six-egg omelet	no omelet, and five good eggs destroyed
Break into saucer	six-egg omelet, and a saucer to wash	five-egg omelet, and a saucer to wash
Throw away	five-egg omelet, and one good egg destroyed	five-egg omelet

$:\langle S, A, P, R\rangle$

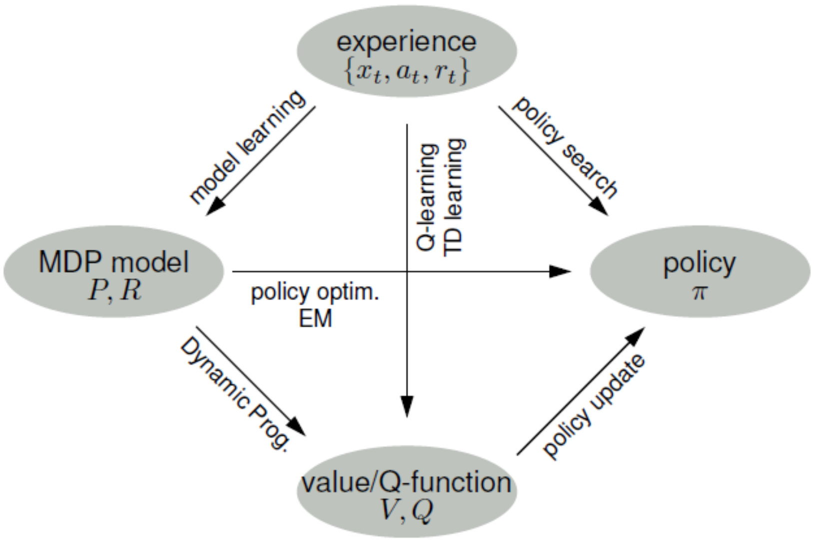
Decision Making Over Time

- Variational optimization – select best ***path*** over time, given transition dynamics
- Stochastic Control or Markov Decision Processes
- MDP: Consider a repeated version of Savage's example

$$p(x' \mid x, a)$$

- Find a policy:

$$\pi(x) = \operatorname*{argmax}_{u} \left[r(x,u) + \int \hat{V}(x')\, p(x' \mid u, x)\, dx' \right]$$



One Interpretation: Re-cast as (Bayesian) inference: infer best assignment to maximize probability of a desired state

Dealing with Partial Observability

- Often, we can't see everything that determines the dynamics of utilities
- Introduce notion of a **belief state** – distributions over possible states of the world
- Instead of directly working with $p(x' \mid x, a)$, introduce and calculate with $b \sim \hat{p}(x)$

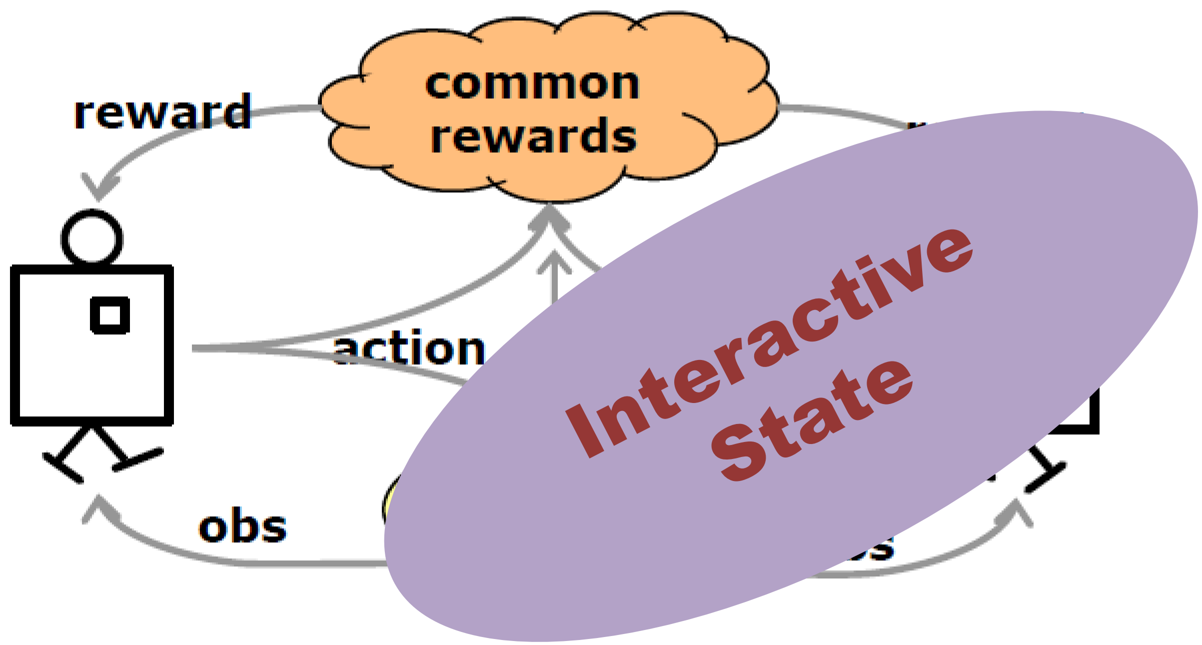
$$V_T(b) \;=\; \gamma \max_u \left[r(b,u) + \int V_{T-1}(b') p(b' \mid u, b)\, db' \right]$$

Interactive Decisions: Dynamics of beliefs

In an interactive world, we hold beliefs over states; ***the states are determined by someone else's actions***

– What are their beliefs; how do they choose actions?

One approach: DEC/I-POMDP



"…instead I looked away. But he understood. Just as I understood that he had understood. Just as he understood that I had understood that he had understood. But all this understanding only went so far as it can go in a few seconds."

[Orhan Pamuk, My Father's Suitcase, Nobel Lecture, 7 Dec 2006]

Models of Rational Interaction: Coordination Games

- Simple classical example is the Stag Hunt (Rosseau)
  - Models cooperation
- Two hunters, each can go after stag or hare but the decision must be made independently

Simple example illustrates:

- Multiple Nash equilibria (payoff/risk dominant)
- Notions of 'trust'
- Coordination failure, problem of equilibrium selection

		Hunter 2: stag	Hunter 2: rabbit
Hunter 1	stag	10,10	0,8
	rabbit	8,0	7,7

Learning for Equilibrium Selection, etc.

- In a repeated game, we want to learn opponent's strategy – family of conditional distributions over future actions. We could learn:

  1. Start with a model of the opponent's strategy.
  2. Compute and play a best [or almost best] response.
  3. Observe the opponent's play and update your model.
  4. Goto step 2.

- When does Bayesian updating yield posteriors consistent with actual strategies?
- Crucial issue – process generating data is nonstationary (especially if everyone is learning and is rational!)

A Result of Kalai and Lehrer [1993]

In a finite stage game played infinitely often by Bayesian rational players, if players' strategies induce a distribution on play that is absolutely continuous w.r.t. their induced beliefs about play, then on almost all play paths every player is a good predictor, and behavior comes asymptotically close to equilibrium

Two key assumptions:

1. Beliefs have some accuracy vis-à-vis strategies
2. Strategies are optimal vis-à-vis beliefs

Result, e.g., by H. Peyton Young

… there exist no general, model-based procedures for multi-agent learning when players are perfectly rational and they have **sufficiently incomplete** knowledge of their opponents' payoff functions.

Crucial assumptions:

- Unknown payoffs are distributed over some interval
- If instead they were known to lie in a finite set, result fails
  - One can **tailor** forecasting rules to take account of the restricted set of payoffs that the opponent could be using
- Agents must optimize exactly. If instead agents almost optimize, the result does not necessarily hold

Subtlety of Real Decision Making Behaviour

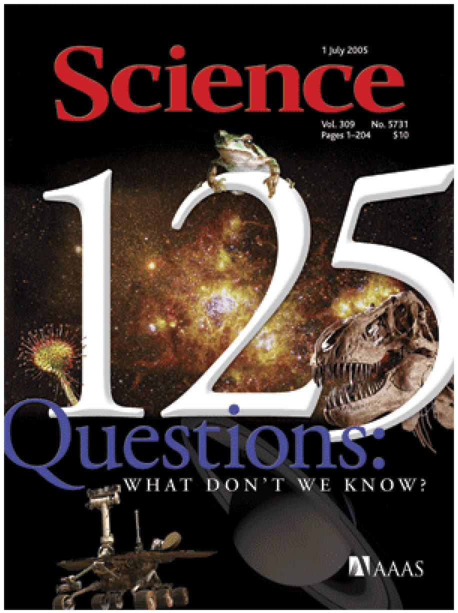
Consider Akerlof's market for lemons:

- Seller knows more than the buyer in a used car market (is the car a 'lemon' or not?)
- Buyer: I don't know if it is a lemon, so I offer an adjusted price
- Seller: If car isn't a lemon, I want a good price
- The only things that will be traded are lemons!

- We don't care about used cars, but the issue of information asymmetry exists in all interactions involving humans & robots

What is the "correct" way to act here?

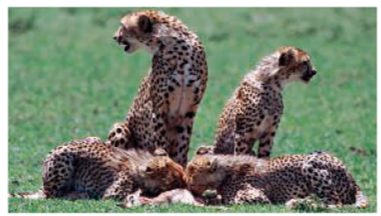
Such Problems have Distinguished Pedigree



Special Section

When Charles Darwin was working out his grand theory on the origin of species, he was perplexed by the fact that animals from ants to people form social groups in which most individuals work for the common good. This seemed to run counter to his proposal that individual fitness was key to surviving over the long term.

By the time he wrote *The Descent of Man*, however, he had come up with a few explanations. He suggested that natural selection could encourage altruistic behavior among kin so as to improve the reproductive potential of the "family." He also introduced the idea of reciprocity: that unrelated but familiar individuals would help each other out if both were altruistic. A century of work with dozens of social species has borne out his ideas to some degree, but the details of how and why



helped humans become Earth's dominant vertebrate: The ability to work together provided our early ancestors more food, better protection, and better childcare, which in turn improved reproductive success.

However, the degree of cooperation varies. "Cheaters" can gain a leg up on the rest of humankind, at least in the short term. But cooperation prevails among many species, suggesting that this behavior is a better survival strategy, over the long run, despite all the strife among ethnic, political, religious, even family groups now rampant within our species.

ilar studies have shown that even when two people meet just once, they tend to be fair to each other. Those actions are hard to explain, as they don't seem to follow the basic tenet that cooperation is really based on self-interest.

The models developed through these games are still imperfect. They do not adequately consider, for example, the effect of emotions on cooperation. Nonetheless, with game theory's increasing sophistication, researchers hope to gain a clearer sense of the rules that govern complex societies.

Together, these efforts are helping social scientists and others build on Darwin's observations about cooperation. As Darwin predicted, reciprocity is a powerful fitness tactic. But it is not a pervasive one.

How Did Cooperative Behavior Evolve?

Evolutionary biologists and animal behavior researchers are searching out the

Modern researchers have discovered that a good memory is a prerequisite: It seems

Tentative answers in game theory – in consideration of equilibrium selection, evolutionary theory of organizations/norms, etc., **but how to implement**?

The slice of the Question I Want to Address, in the Long Term:

When, why and how might ***computationally/ informationally bounded agents*** *be able to solve these apparently intractable problems efficiently?*

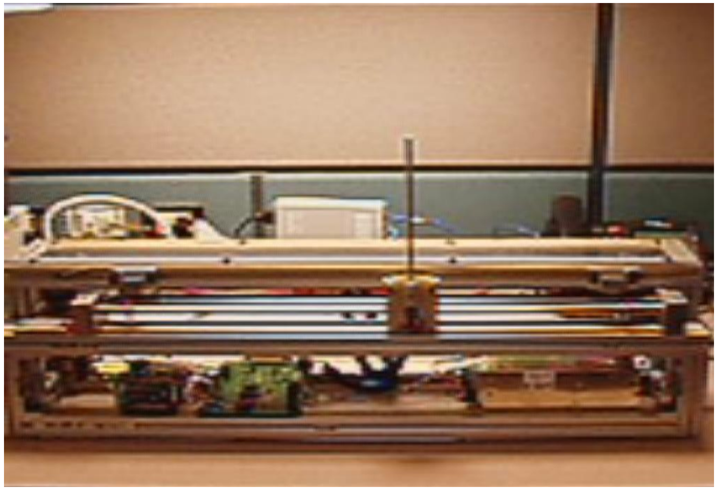
*Might there be structural principles that guide our search for appropriately '****simple' strategies****?*

Three Worked Examples,

to illustrate the notion of

"simple strategy"

Warm-up Example #1: Control of Cart-Pole



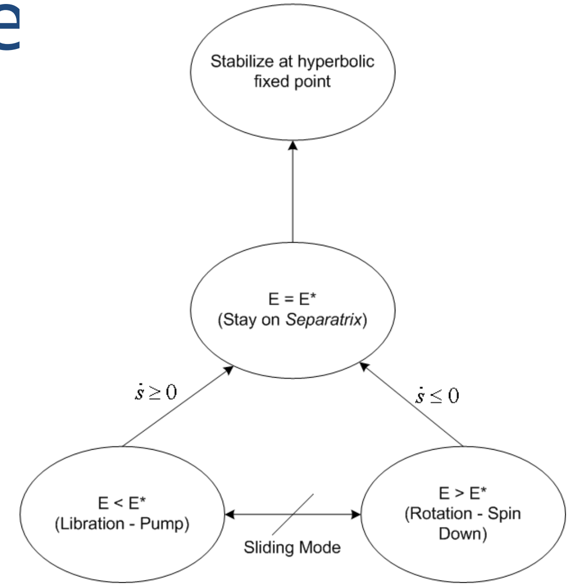
- Two subsystems – pendulum, cart on finite track; only one actuator – cart
- We seek *global asymptotic stability* of 4-dim system
- Very popular example in robot learning

My main concern:

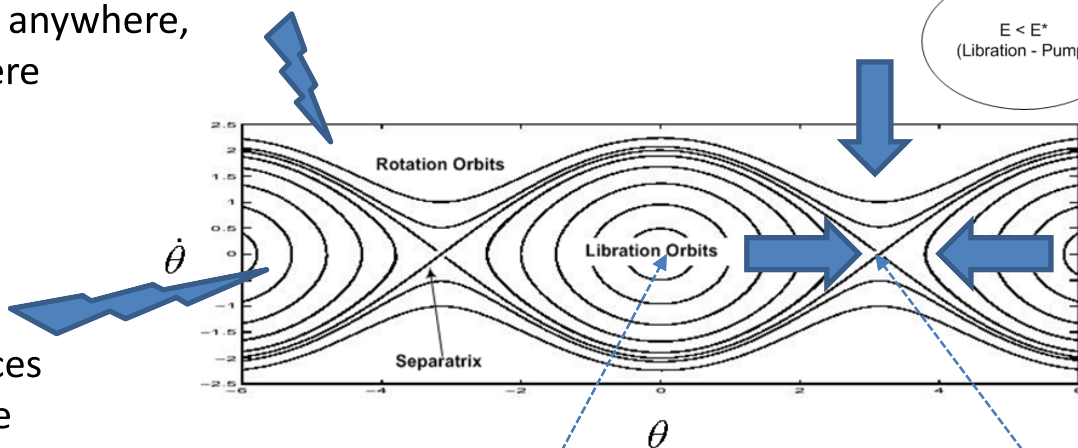
Is there a principled explanation for the simplest intuitive strategy…?

S. Ramamoorthy, B.J. Kuipers, Qualitative heterogeneous control, **HSCC ’02 & ’03**

Reactive Behaviour - Global Structure



Adversary could push system anywhere, e.g., here



Can describe global strategy as a qualitative transition graph

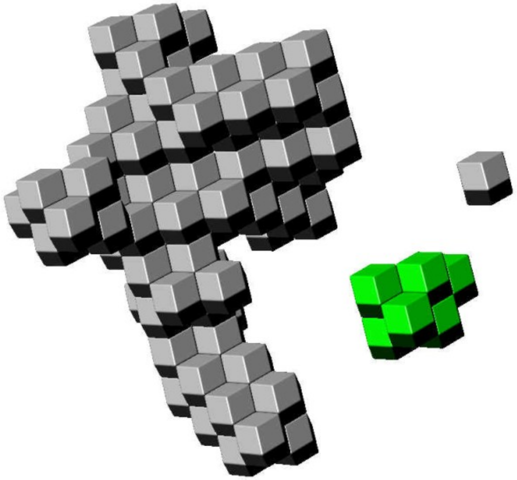
Larger disturbances could truly change quantitative details, e.g., any number of rotations around origin

The uncontrolled system converges to this point

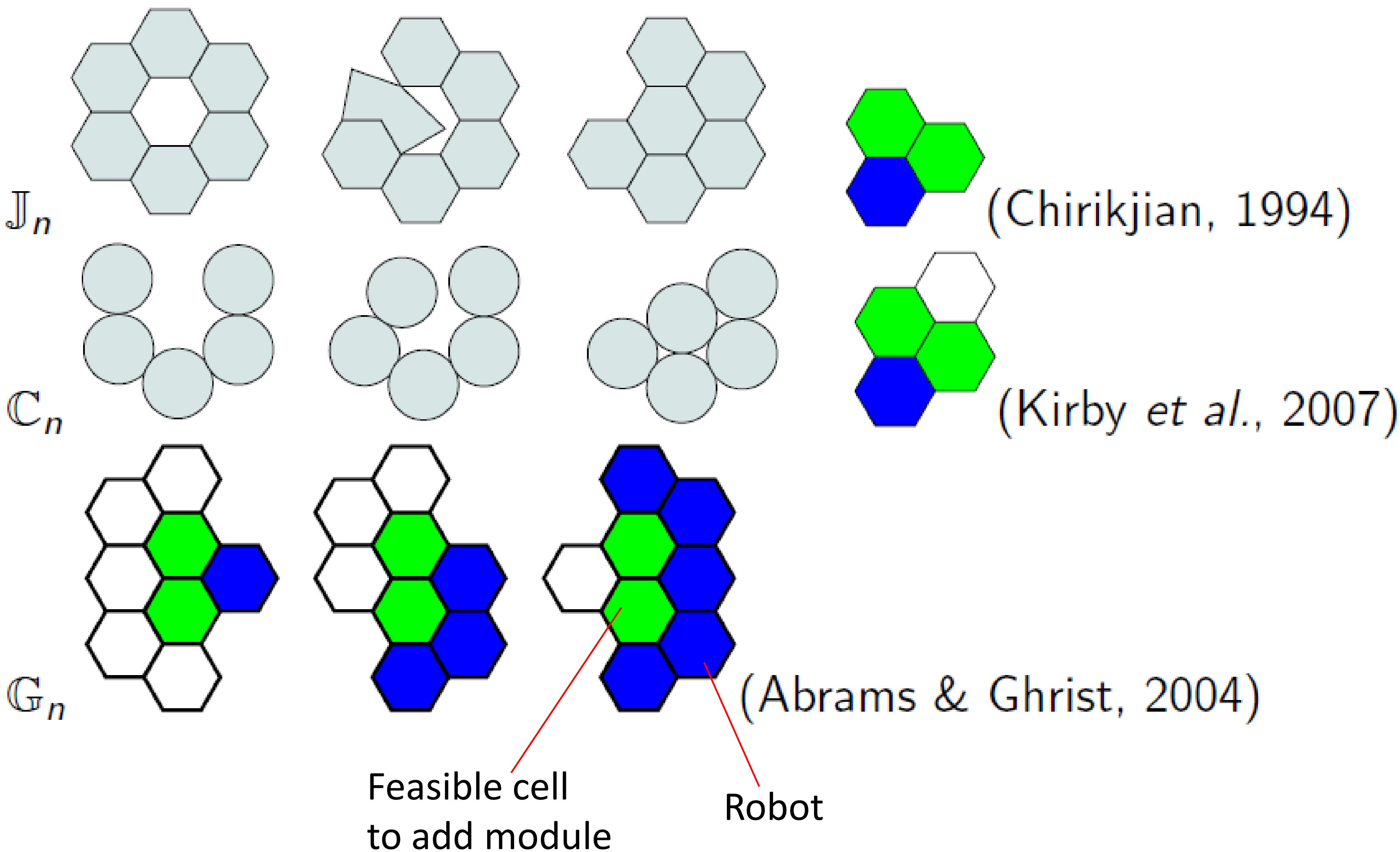
We want to reach and stay here

Example #2:
Shape Reconfiguration in Modular Robotics

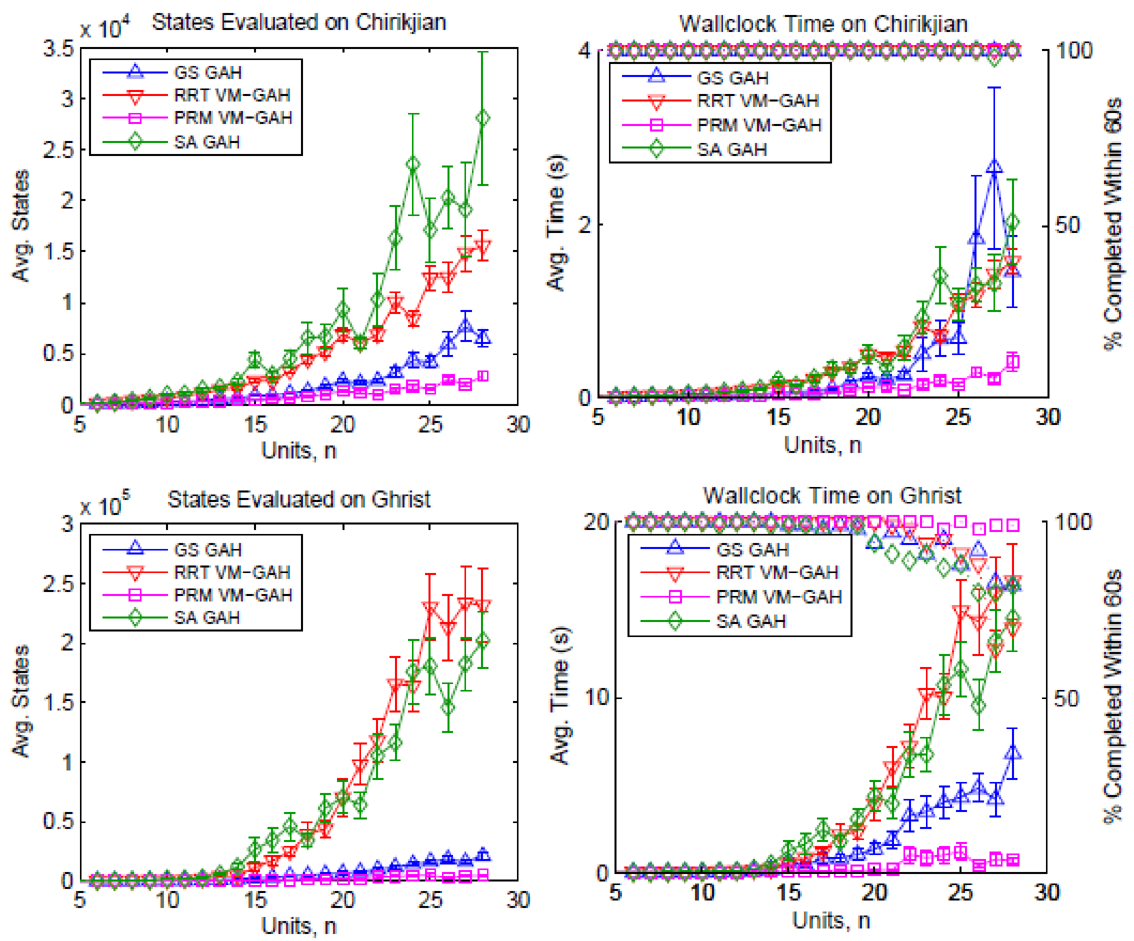
- Robot is defined by morphology & connectivity (shape problems)
- Many practical realizations – e.g., Xerox PARC, Intel, CMU, MIT, Cornell



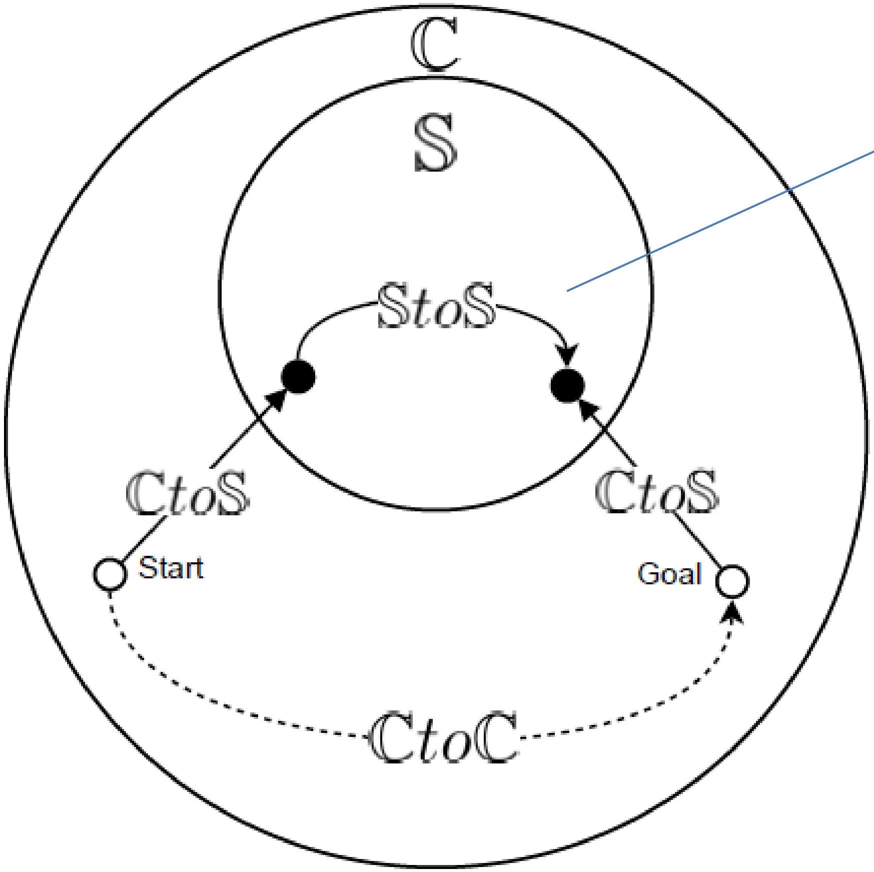
Hexagonal Metamorphic Robot Models - Catalog defines Local Constraints



Inefficiency of Sampling Based Motion Planning



An Alternative



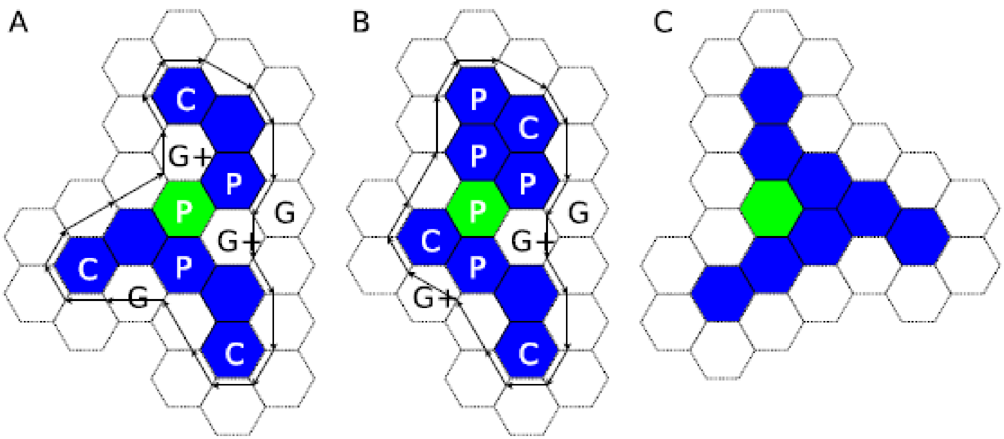
Subspace with nice properties

T. Larkworthy, S. Ramamoorthy, **A characterization of the reconfiguration space of self-reconfiguring robotic systems**, *Robotica* (2011)

A Greedy Planner

Definition

P = Placed
C = Contract
G = Growth
G+ = Unviolating G



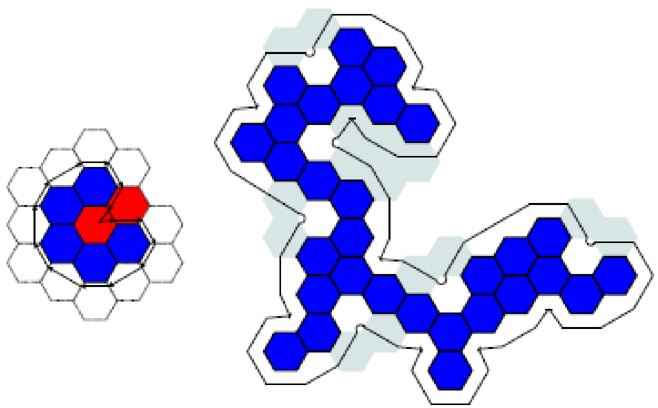
$\mathbb{S}to\mathbb{S}$

while(curr != goal)
  if(!*improve*()) **error**
  updateLabels()

improve

try *move*(C to G+)
try *move*(C to G)

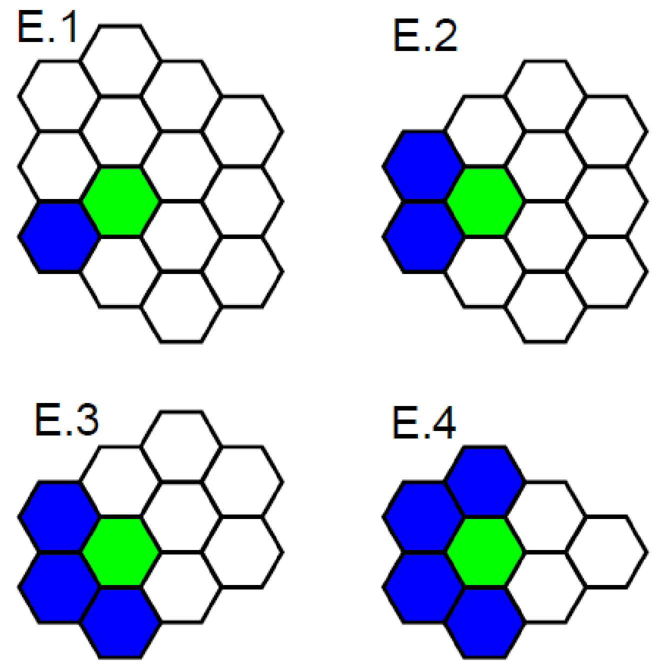
The Surface Model: Local Constraints



$\mathbb{S} \leq \mathbb{G}$

Not allowed to contain kink or dual path violations on a Hamiltonian Path around the perimeter

Generators:



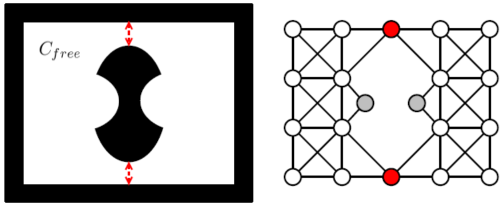
S to S Planning is, empirically, linear time – Why?

Algebraic Connectivity of C-Space

- Consider unweighted graph G = (V, E) generated as follows:
  - Generate random configuration by uniformly selecting action starting from random initial start points
    - By action, we mean a full “move” not one atomic module movement
    - Effectively, an initial point is allowed to ‘diffuse’ in c-space
- By defining the cost of a cut in terms of number of edges, define the Cheeger constant as,

$$h_G = \min_S \frac{cut(S)}{min\{vol(S), vol(V-S)\}}$$

- Computing this for various points in c-space characterizes it



Approximating Algebraic Connectivity

- Directly computing Cheeger constant for large module numbers (e.g., 20000) may be infeasible
- Instead, define a Laplacian, *L,* with elements,

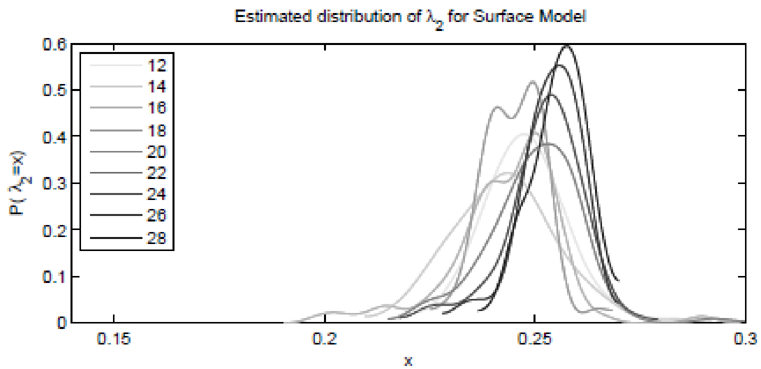
$$l_{i,j} \quad := \begin{cases} 1 & if\, i = j\, and\, deg(v_i) \neq 0 \\ -\frac{1}{\sqrt{deg(v_i)deg(v_j)}} & if\, i \neq j\, and\, v_i\, is\, adjacent\, to\, v_j \\ 0 & otherwise \end{cases}$$

- Its eigenvalue bounds the Cheeger constant, $\sqrt{2\lambda_2} > h_G \geq \frac{\lambda_2}{2}$

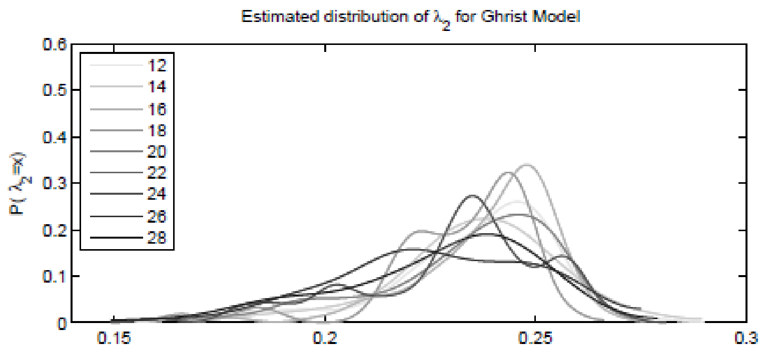
Experiment with Algebraic Connectivity

- On a randomly chosen set of configurations, expand neighbors to fixed depth and estimate Cheeger constant (via eigenvalues of Laplacian):

$$h_G = \min_S \frac{cut(S)}{min\{vol(S), vol(V-S)\}}$$

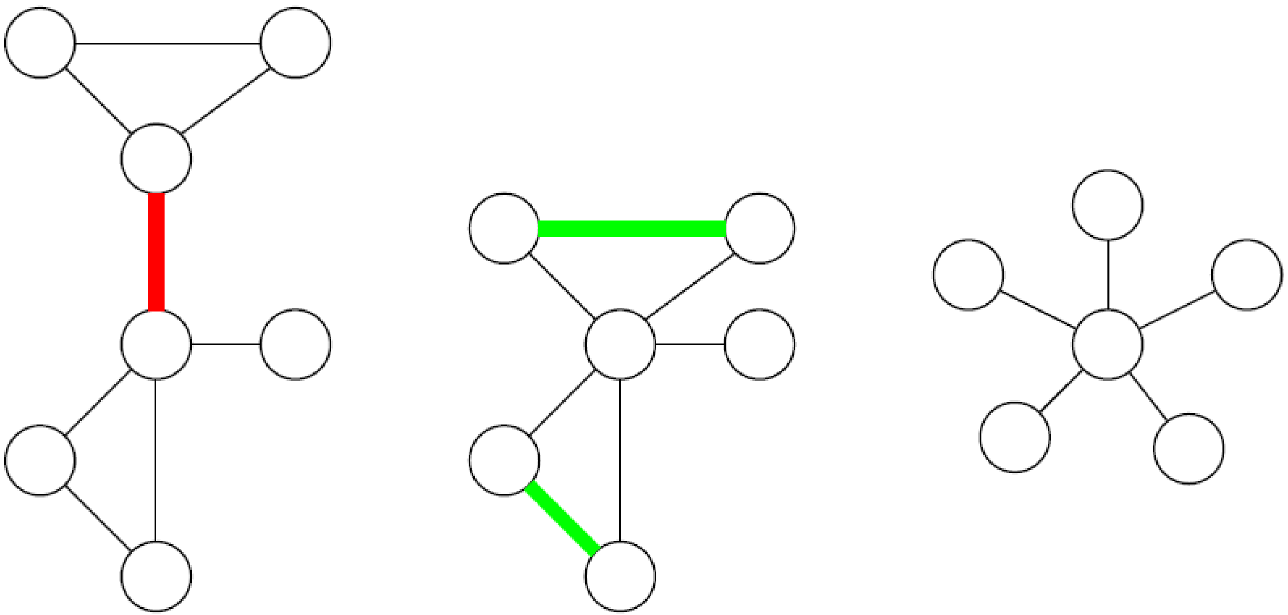


For comparison:



This is suggestive, but there is an even better reason.

Graph Minors

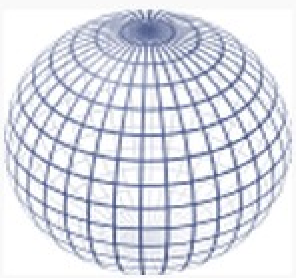


Definition

H is a minor of G, written $H \leq G$, if there exists a sequence of edge contractions and edge/vertex deletions that transform G into H

Implications of Graph Minor Property

- If H < G then H’s graph genus is less than or equal to G
- Many properties preserved under minor (tree width)
- Cheeger constant of H is less than G
- Gives an indication of which graph is more complicated (if either)



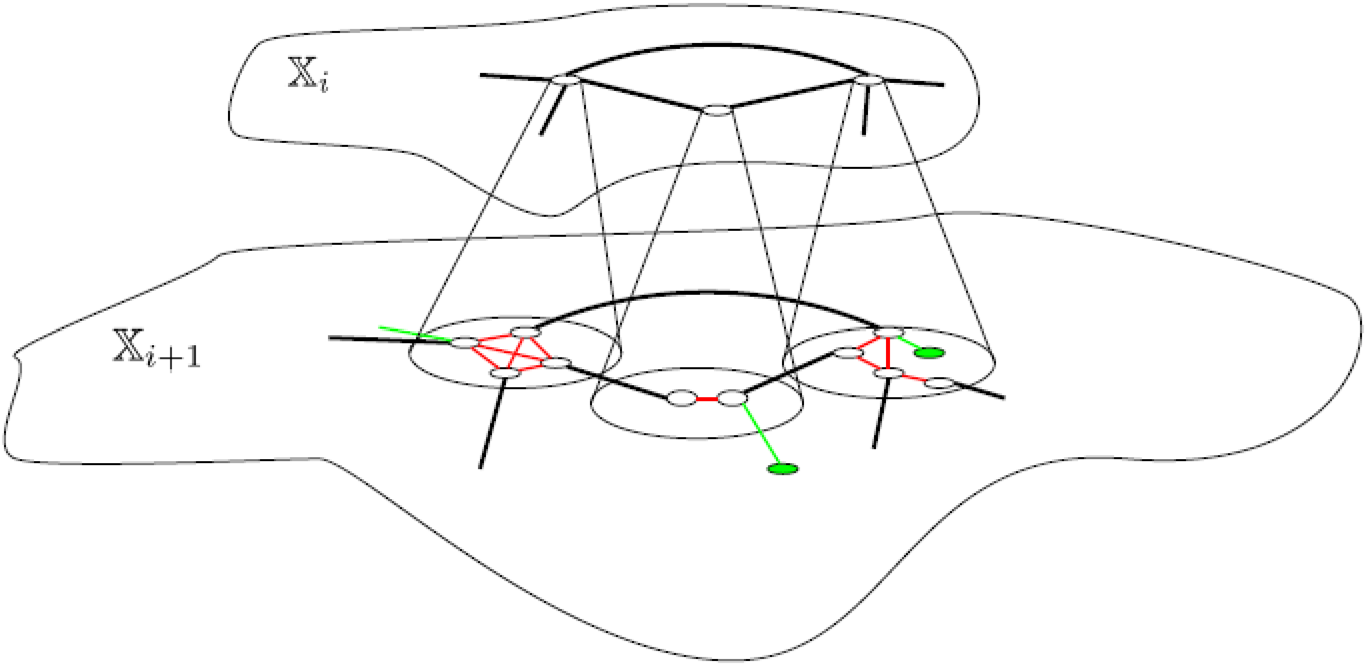
genus 0

genus 1

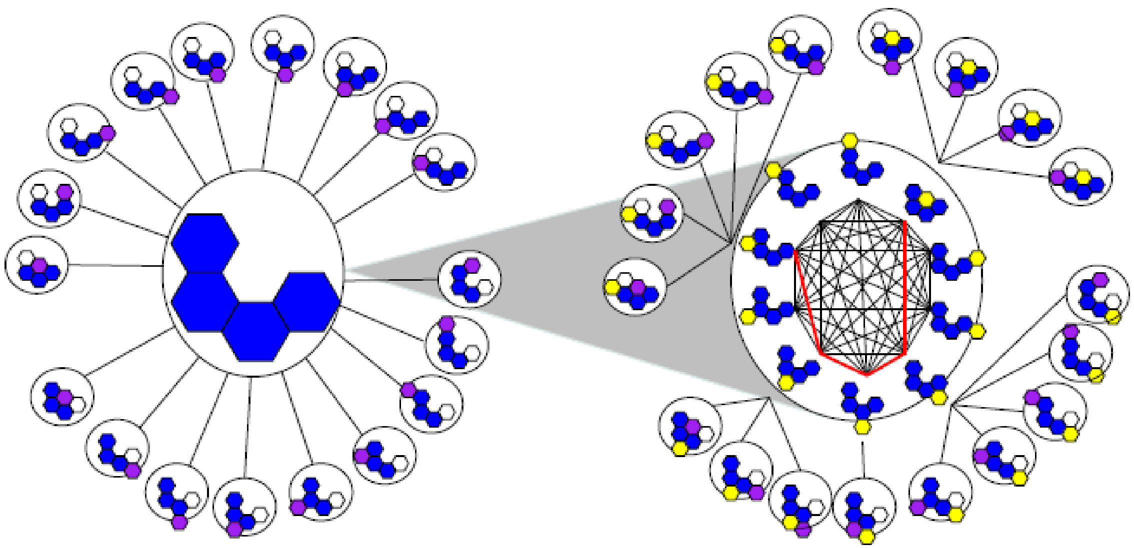
genus 2

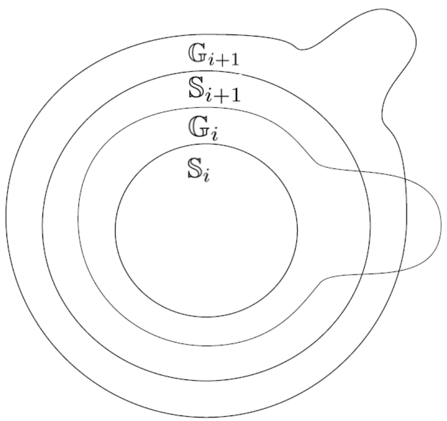
genus 3

What Happens as Modules are Added?



Local Structure and Graph Minors: When is an Action Set ‘Difficult’?





A path plan in S_n exists in S_{n+1}: suggests recursive solution strategy
- Not so for Ghrist’s catalog

Example #3: Low-Complexity Coordination Strategies

How to coordinate to an equilibrium even though the agents are not rational, and do not learn about the system as a whole or even about each other?

H.P. Young’s Interactive Trial and Error Learning model:

- Adjust behavior only in response to own realized payoff
- No knowledge of the overall structure of game
- Cannot observe actions or payoffs of most other players
- Occasionally tremble and make mistakes

Examples: driving in a big city, packet routing in networks

Essence of the ITE Algorithm

- Agent has a 'mood': {*content, discontent, hopeful, watchful*}
- Player state = {mood, benchmark action, benchmark payoff}

What do the profiles mean?

- Content: Experiment with small probability; If experiment results in a higher payoff, adopt the new action & payoff as benchmarks
- If payoff *increases without experimenting,* become hopeful but don't change benchmark action right away
- If payoff stays up become content again with new higher payoff as benchmark
- If payoff *decreases below benchmark without experimenting,* become *watchful but don't* change benchmark action right away

…

…

- If payoff stays below benchmark become discontent
- If payoff goes back above benchmark become hopeful
- Discontent: Flail around; try a new action at random and with probability $0 < p < 1$ stay discontent

  With probability $1 - p$ *spontaneously* become content with the current action and payoff as new benchmarks

Result

Theorem

If all players use interactive trial and error learning and the experimentation rate ε > 0 is sufficiently small, then for almost all n-person games on finite action space *that possess at least one* pure Nash equilibrium*, a Nash equilibrium is played at least 1 - ε of the time.*

Why (in a nutshell)?

- Every recurrence class (potential minimizer) contains at least one all-content state in which the action benchmarks constitute a pure Nash equilibrium
- Stochastically stable states are all of this form

How to Build on Such Examples?

- In each case, I have shown you a ‘hack’ backed by Theorem(s).
- Can these then be obtained by special kinds of
  - Unsupervised/developmental/life-long learning?
  - Mathematical discovery?!

For this to be possible at all, we need:

1. More empirical examples of principles extracted from such problems: understand the **phenomenon of decision making**
2. A ‘description language’ for capturing all this in a unified way

Why Focus on Specific Problems? Maybe We Need Better General Algorithms?

Consider set of **all** structurally distinct strictly ordinal 2x2 (no)-conflict games [Rapoport & Guyer '66]; five major categories of multi-agent learning algorithms

Exhaustive evaluation in ad hoc team setting yields:

Agent	Conv.	Fexp.	Welfare	Fairness	NE	PO	WO	FO
JAL	1	3.9866	7.9720	15.9063	1	0.9920	0.9920	0.9920
CJAL	1	3.9831	7.9663	15.8874	1	0.9897	0.9897	0.9897
WOLF-PHC	0.9996	3.9449	7.8908	15.6426	1	0.9638	0.9638	0.9638
RegMat	0.9990	3.9107	7.8170	15.3906	0.9954	0.9457	0.9457	0.9457
NashQ	0.9987	3.9840	7.9733	15.9144	0.9954	0.9939	0.9939	0.9939

Table 1: Results for no-conflict games.

Agent	Conv.	Fexp.	Welfare	Fairness	NE	PO	WO	FO
JAL	0.8901	3.0140	6.0592	8.9997	0.8982	0.7781	0.7021	0.6164
CJAL	0.9456	3.0326	6.0978	9.0900	0.8470	0.8050	0.7184	0.6250
WOLF-PHC	0.9430	3.0392	6.0620	9.0517	0.9047	0.7636	0.6992	0.6142
RegMat	0.8673	3.0313	6.0368	8.9610	0.8946	0.7662	0.7000	0.6109
NashQ	0.9990	3.0446	6.0667	9.0755	0.8722	0.7767	0.6946	0.6097

Table 2: Results for conflict games.

[S. Albrecht, S.R., subm. AAMAS 2012]

Two Domains of Interest

- Robot football
  - RoboCup: world champion robots vs. humans, by 2050
  - Many different leagues: Standard platform, Simulation
  - Connections to human football, sports sciences, etc.?

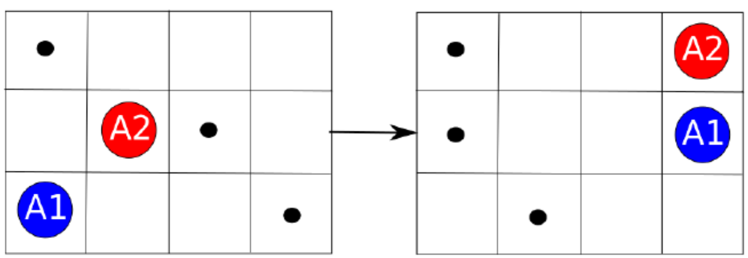
- Agents in Electronic Markets
  - Penn-Lehman Automated Trading (PLAT) Project
    - Also other related domains, e.g., Trading Agent Competition
  - My vision of a new 'Market Analysis and Design' Competition
  - Connections to economics, especially *behavioral* finance/econ.?

Current Research Theme #1:
Control of Interactive Decisions

How does one shape a partially controllable interaction to achieve outcomes defined jointly by oneself, other agents and an external world ?

[A. Valtazanos, A. Robinson]

Interaction with Limited Authority



Initial State Target State

- A1: Controlling agent
- A2: Adversarial agent (to be controlled)
- External objects

- Lead the 'world' from an initial to a final state
- The state itself has many components:
  - Controllable
  - Uncontrollable
  - Jointly Manipulable

Interaction Control: Model

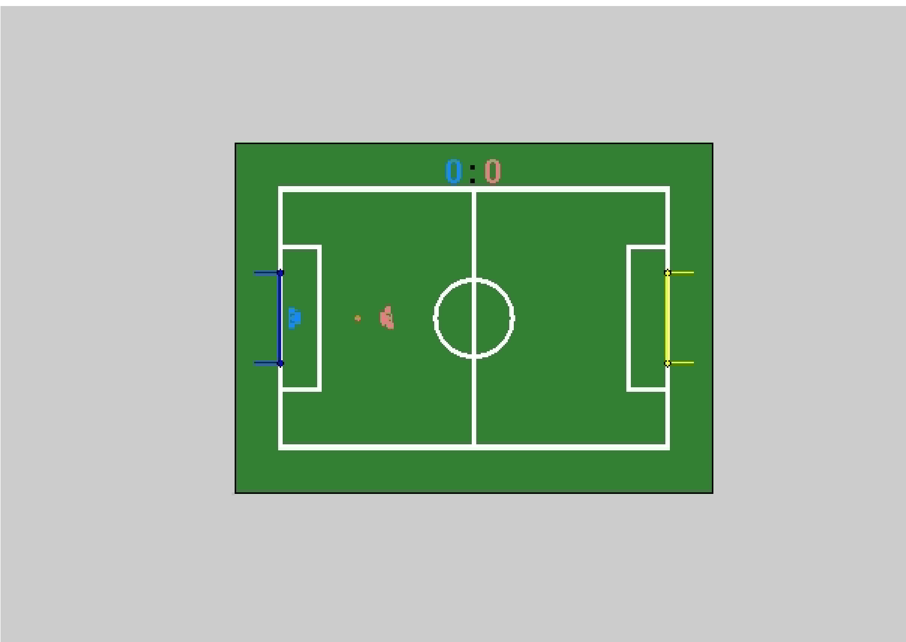
- The state of the art in this area is to maximize a reward in the I-POMDP model
  - Use particle filter over beliefs (nest to 1 or 2 levels, if needed)
- This often works but we aren’t really ‘controlling’ the other player explicitly – is that possible?

Our approach, Hierarchical Interaction Control Process:

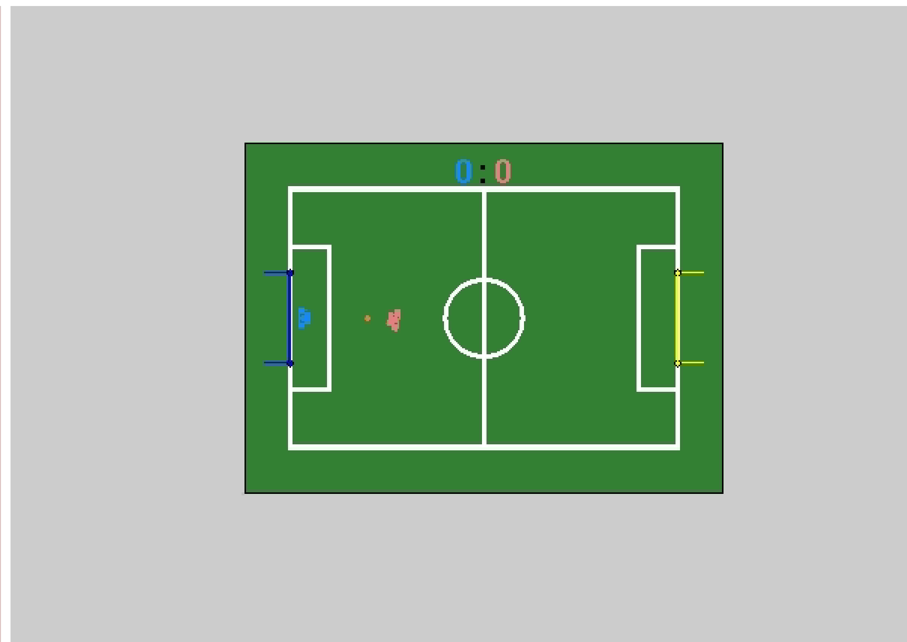
- Identify interaction predicates (e.g., I can see the ball, opponent can’t) that can be locally sensed and acted upon
- Learn local policies to enforce these predicates to some level
- Define global policy as a sequence of such local games

Experiment – Strategic Interaction

POMDP Striker



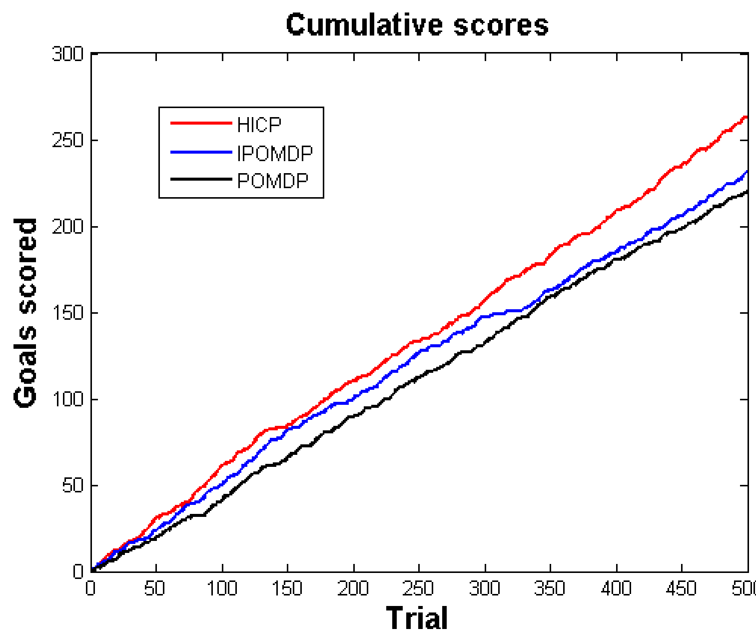
HICP Striker



Preliminary Experiment: HICP vs. (I)POMDP

Table below shows statistics of sequences of consecutive successes (S) & consecutive failures (F) of striker

- Gives an indication of how reliably limits of the heuristic goal keeper are exploited



Kick type	max(S)	max(F)	μ(S)	μ(F)	σ^2(S)	σ^2(F)
HICP	9	7	2.275	1.8902	3.2146	1.9014
IPOMDP	5	11	1.775	2.2118	0.9361	3.3356
POMDP	4	9	1.5517	2.2326	0.7153	2.2041

Where to next, on this front?

- Ability to control coordination and beliefs when the other person is genuinely being strategic
  - Currently, we have adaptation but not recursion of beliefs!

- Ability to incorporate information constraints explicitly, e.g., privacy – I want robot to track and interact with my child without revealing too much to a hacker if compromised!

Current Research Theme #2:

Life-long Decision Making

Not knowing what games are coming up, in continually changing multi-agent environments, how should we represent knowledge and learn strategies?

[M. Hawasly, B. Rosman, A. Robinson]

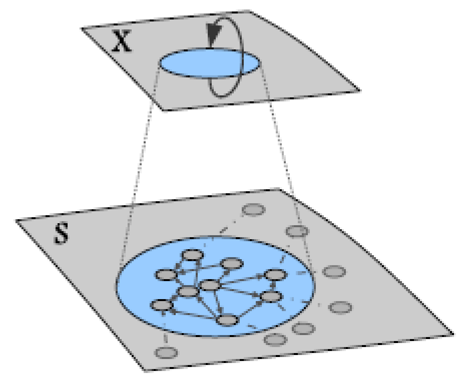
Life is like a 'box' of situations,
you never know which one you're gonna get!

What is the Problem?

- Over a lifetime, we can learn **many** skills (dribble, kick, etc.) while playing in numerous training games
  - Each of these can be individually learnt, e.g., using RL
- Eventually, you'll be put in test situations where you decide:
  - **What game am I supposed to be playing now?**
  - How best to reuse versions of my learnt knowledge?
- You must do this subject to incomplete knowledge of the state of play – you haven't seen this opponent before, you are not yet sure what the relevant attributes are, etc.
- Our approach: Keep a structured representation of 'skills' and play incomplete information games, e.g., aspiration learning

Model of Capabilities

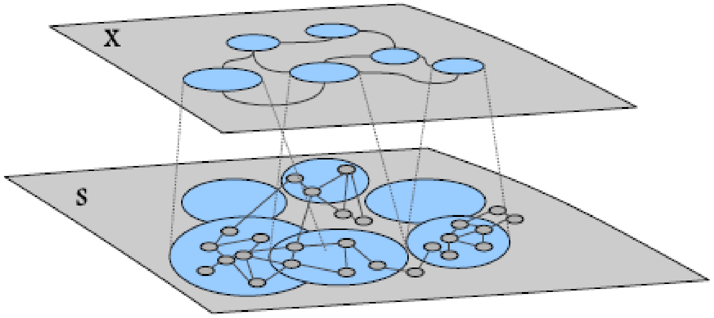
- We shift focus to what we can do and what we may lose due to opponent
- Define capabilities as policies that can preserve a local predicate w.r.t. a domain and conditions



$$C = \langle D_c, \pi_c, p_c, \rho_c \rangle$$
$$\forall x \in D_c,\ \exists(a_1, a_2, \ldots, a_n) = \pi_c(x) \in A^n_{\pi_c} : P(p_c(\pi_c \circ x)) \geq \rho_c$$

Transition Capability: Switchers

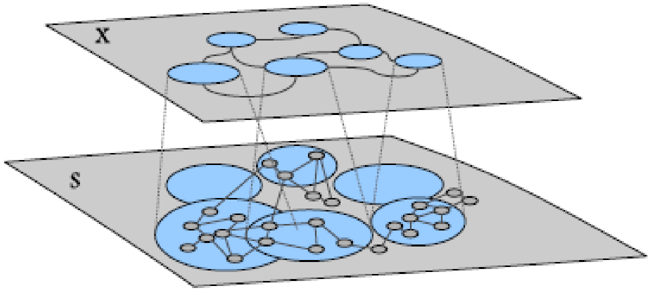
- Decompose problem into predicates that can be maintained and transitions between such sub-problems



$$u = \langle D_{c_1}, \pi_u, D_{c_2}, \rho_u \rangle$$
$$\forall x \in D_{c_1},\ \exists a = \pi_u(x) \in A_{\pi_u} : P(p_{c_2}(a \circ x)) \geq \rho_u$$

Policy Space in terms of Capabilities

- May think of this as related to options framework
- A better analogy, capabilities = ‘local games’
  - Policies can be strategies for a multi-agent game



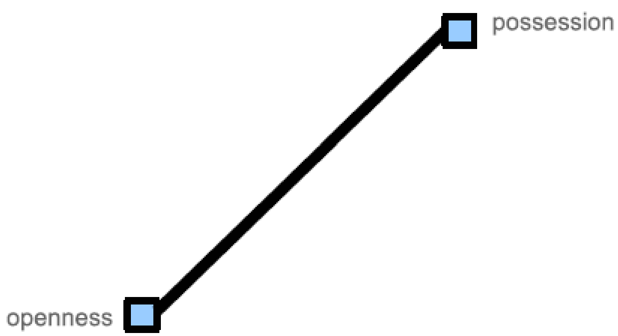
Space of Predicates

every predicate of a capability is a vertex (*0-simplex*)

$\langle D_c, \pi_c, p_c, \rho_c \rangle$

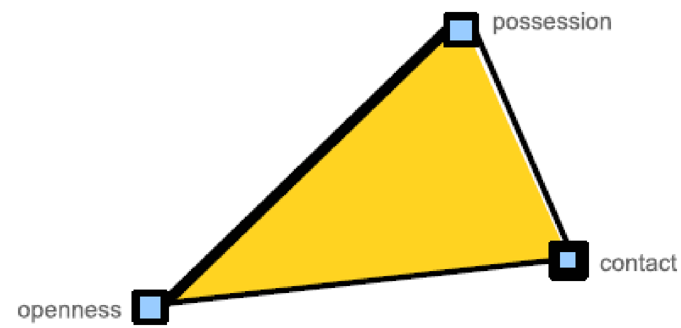
any possible combination of predicates is a simplex

- capturing local constraints



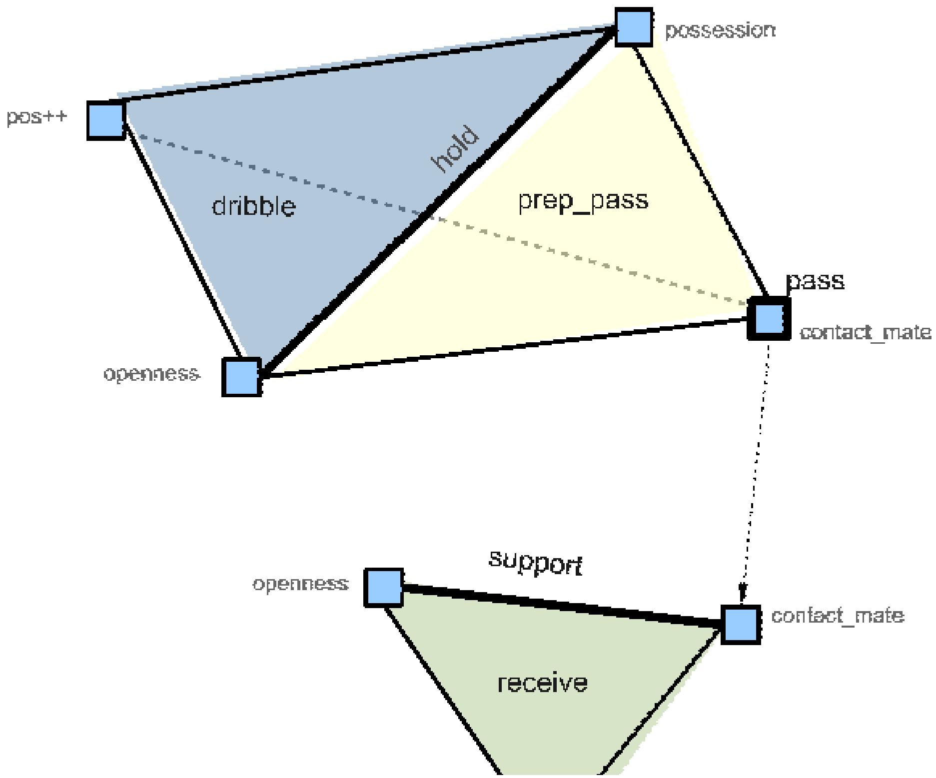
1-simplex

$\langle D_{c1} \cup D_{c2}, \Pi_{c1c2}, p_{c1}, p_{c2}, \rho_{c1}, \rho_{c2} \rangle$



2-simplex

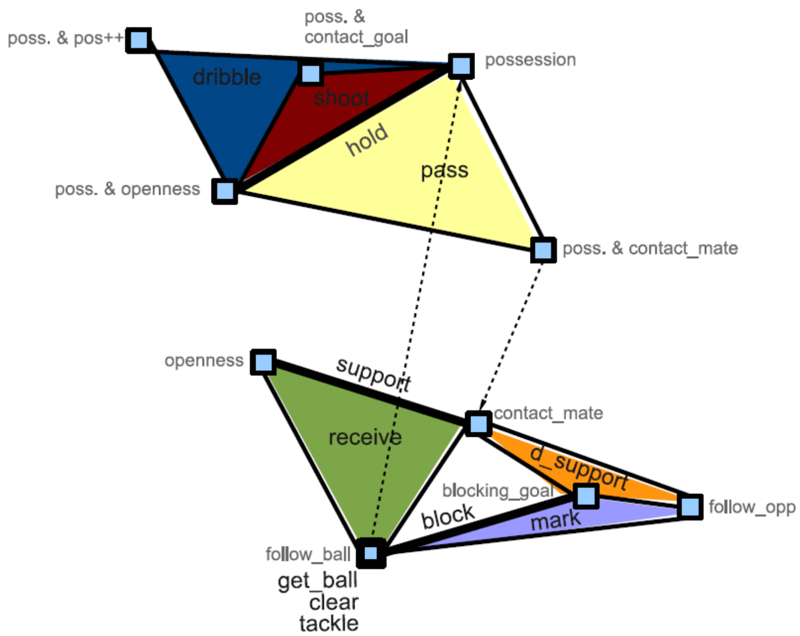
The Predicate Complex: Simplicial Complex formed by Capabilities



Safety: having multiple alternatives at all times

$=$ living on higher dimensional simplices

Topology of Complex Informs What to Learn/Plan for



Implications:

- If we can maintain a suitable topology, randomized algorithms can be made efficient

- Knowing which simplices are 'active', we can structure the use of aspiration learning and other low-complexity learning strategies

Dynamics over Predicate Complex

- *living on a simplex* refers to preserving the simplex predicates

$$\sigma \longrightarrow \sigma; \qquad \rho_\sigma \to \rho_\sigma^+$$

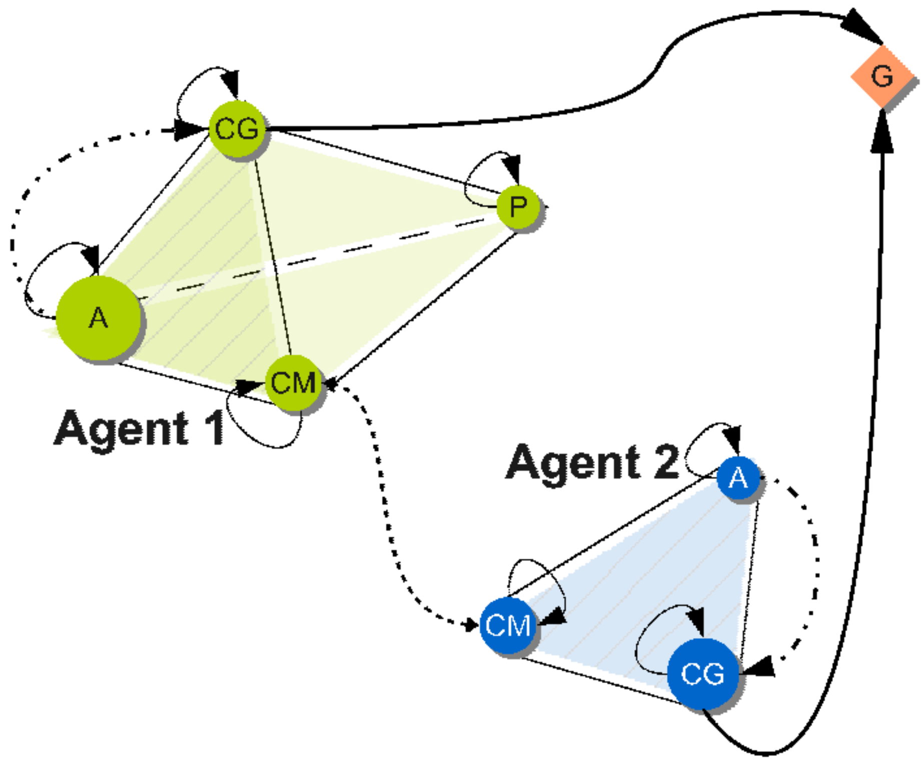
- switchers give a stochastic map

$$\Sigma \times \mathbf{U} \to \Sigma$$

- adversary works to toggle off active vertices

$$\sigma = \{v_1, v_2, v_3\} \rightsquigarrow [\{v_1, v_2\}, \{v_1, v_3\}, \{v_2, v_3\}]$$

Dynamics over Predicate Complex



task: find a sequence of simplices and actions, online, so that the goal simplex becomes active under the dynamics:

$$\sigma_0 \xrightarrow{u_0} \sigma_1 \xrightarrow{u_1} \sigma_2 \longrightarrow \ldots \longrightarrow \sigma_g$$

2D Robot Soccer

Question:

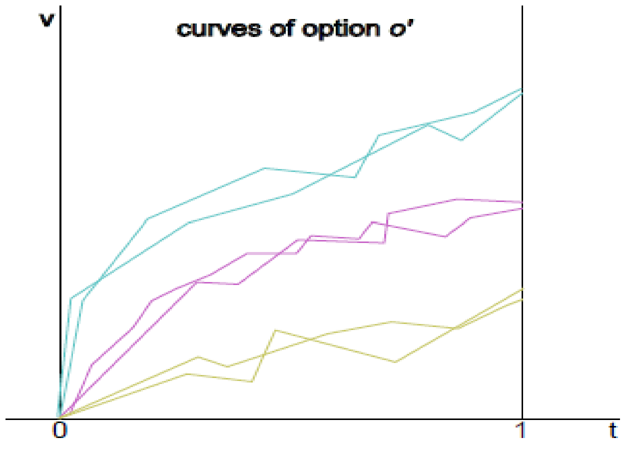
Implicit in our construction is the notion of a 'tactic' associated to 'situations'.

Is there structure in the 'space of strategies'?
Is there a canonical description of a situation?

[H. Mahmud]

Life as a Sequence of Markov Games

- Markov game = <agents, states, dynamics, reward>
- World is a sequence of unknown game contexts: Θ_i
- Within each context, a set of local strategies or options are feasible to use – e.g., policy to dribble around a defender
- For all realizations of contexts and Markov games, we can collect information (over a lifetime) and store that as a performance history of an option
- Renormalize:



Structure in the Space of Options

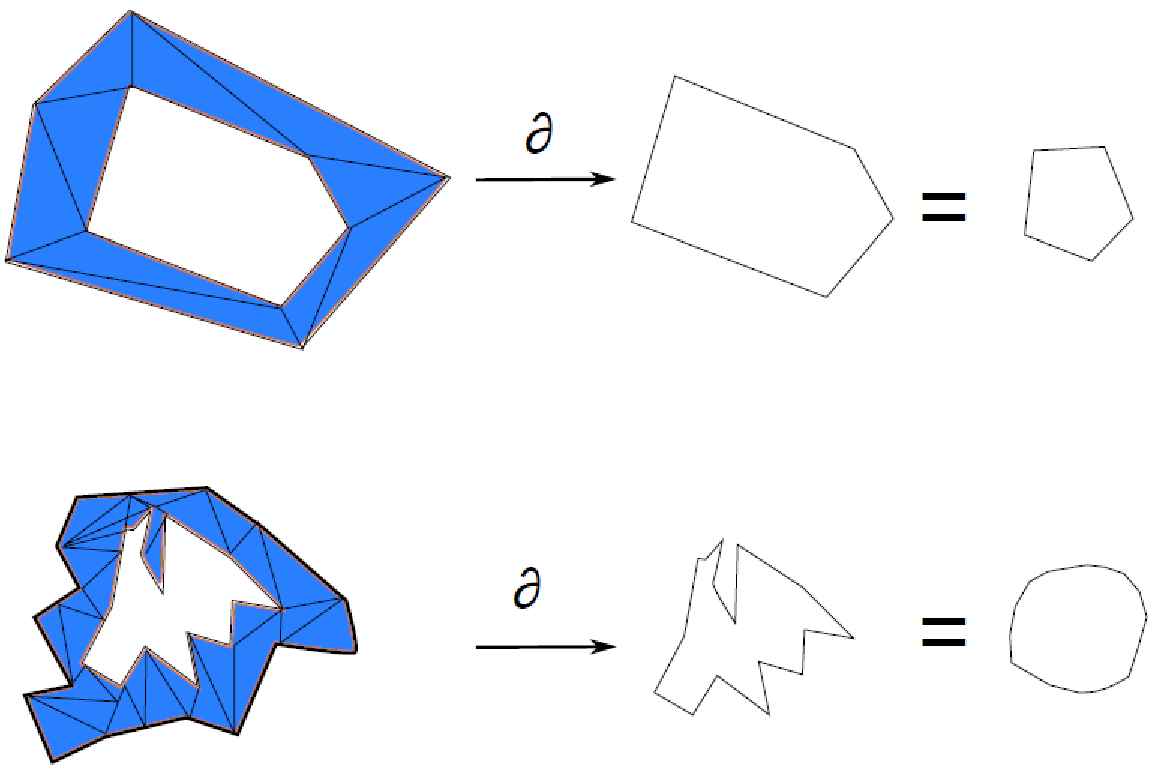
- We can compare two options in terms of (diffeo)morphisms between the sets of reward curves, BUT…

… we can reduce graph isomorphism to computing this distance

- Instead, we look to computational topology for help
- Define reduction between sequences of options and from that define a simplicial complex in the space of options
  - Key is to define a boundary: no more simplification possible
- Apply reduction until we have a minimal description
- Do all of the above – offline, in your dreams – so that online, make use of **simpler model** to transfer knowledge and plan

What is Computational Topology Doing?

- **Key Idea:** The boundary operator ∂
  - **Given:** A complexity level k view of a topological object.
  - **Computes:** A complexity level $k-1$ view of the topological object.



Research Theme #3:

Collective Decision Making

What, when and how can a collective of simple, bounded agents compute global properties?

[with R. Santhanam & A. Salamon; A. Novik]

Observation: Stylized Facts about Asset Prices

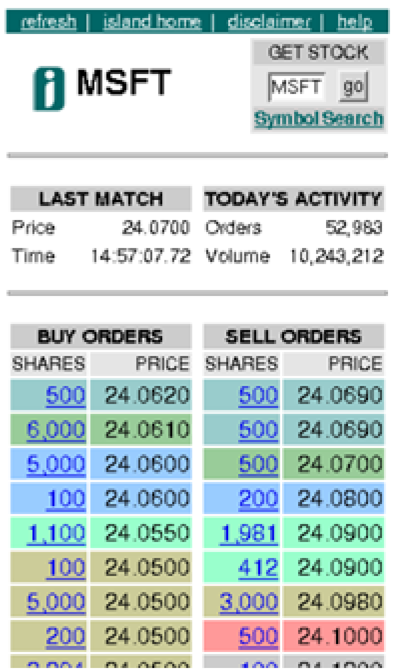
There seems to be *no generic* statistical model that is able to capture all stylized facts, such as the following:

- **Heavy tails**: Returns distribution may have pareto-like tails (persists even after correcting for volatility)
- **Volatility clustering**: Different measures of volatility display positive autocorrelation over many days
- **Gain/loss asymmetry**: Large drawdowns, few upward jumps
- **Asymmetry in time scales**: Coarse grained measures of volatility predict fine-scale volatility better than vice versa
- **Trading volume** is correlated with all measures of volatility
- **Intermittency**: Bursts in volatility time series

Further Complications

Where does price come from? Microstructure

Price set by the trade mechanism – as such, there is no single price (depends on trading direction, speed, agent attributes and other effects)



refresh | island home | disclaimer | help

MSFT

GET STOCK

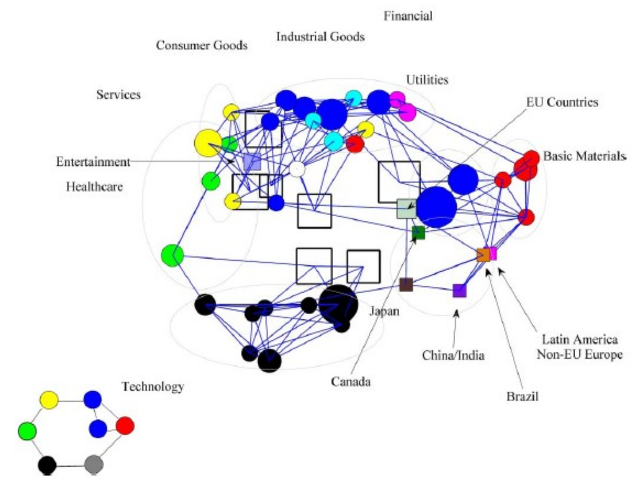
MSFT go

Symbol Search

LAST MATCH		TODAY'S ACTIVITY	
Price	24.0700	Orders	52,983
Time	14:57:07.72	Volume	10,243,212

BUY ORDERS		SELL ORDERS	
SHARES	PRICE	SHARES	PRICE
500	24.0620	500	24.0690
6,000	24.0610	500	24.0690
5,000	24.0600	500	24.0700
100	24.0600	200	24.0800
1,100	24.0550	1,981	24.0900
100	24.0500	412	24.0900
5,000	24.0500	3,000	24.0980
200	24.0500	500	24.1000

Network effects – endogenous dynamics of interaction matters!



G. Leibon et al., Topological structures in the equities market network, *PNAS* 105(52) 2008

Some Questions

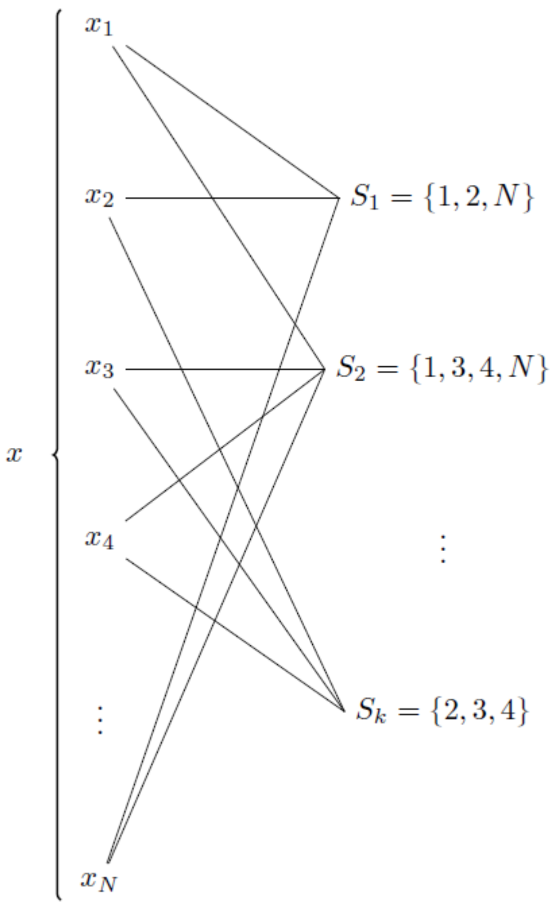
- Econometricians ask how better to model/predict the process and financial engineers ask how better to optimize and ‘price’

- My interests:
  - For bounded agents working with **incomplete** and **asymmetric** information about the full system; are there low-complexity decentralized strategies for characterizing the qualitative structure/state of the dynamics?
  - Can this work in settings with informational constraints (e.g., privacy or disclosure restrictions)?

Macroscopes

- **Global** functions that need to be computed by **local** bounded agents
- How hard is this? Communication complexity

- New issues to consider:
  - How are inputs allocated to different players?
  - What does any player know about allocation structure or others' information?



Example Properties of Macroscopes

Consider the simple change detection problem:

- Every k-player single-blind Constancy macroscope on N D-ary inputs can be solved with cost r(log D) + k, where r is number of connected components of intersection graph of allotment structure. This bound is optimal to within a constant factor.

- Every k-player double-blind Constancy macroscope on N D-ary inputs can be solved with cost kd log(D+1). Moreover, there are k-player double-blind Constancy macroscopes which require cost kd log(D).

Where to next, on this front?

- Extend these results to include more sophisticated communication protocols, meta-information (privacy) and function types (e.g., ranking)

- Can these results be extended to address endogenous dynamic properties?!

- Synthesize macroscopes with well understand behaviour, in:
  - Networked markets
  - Decentralized robots with partial views

Concluding Remarks: Summary

1. Strategically sophisticated inter-dependent decision making is a major open area, of fundamental scientific interest, with immense application potential and in need of study within CS

2. In particular, we need a good description language and implementable algorithms for encoding and learning tasks over a **lifetime** of different interactions

3. We need efficient mechanisms for composing decentralized, heterogeneous, individually unreliable resources into *ad hoc* teams that can take on a large family of tasks

Concluding Remarks: Open Questions

1. Typically, future is **endogenously** created and interactive dynamics are non-trivial, so much of the data you need is unavailable ahead of time. What are useful models of learning in this setting?

2. There is a disconnect between general models such as POMDP/POSG, typically intractable, and powerful but specialized – often behavioral – tricks that seem to make humans efficient under bounded rationality. (How) can we reconcile these?

3. The world is complex – which is why formal models are elaborate, requiring depth. How can/should an *ad hoc* collective support this depth despite continual change in components and contexts?