Learning Space Exploration Agents: Opportunities and Challenges

Steve Chien

Artificial Intelligence Group
Jet Propulsion Laboratory
California Institute of Technology

chien@aig.jpl.nasa.gov

http://www-aig.jpl.nasa.gov

Portions of this work were in collaboration with Barbara Engelhardt, NASA/JPL and UC-Berkeley.

This work was performed at the Jet Propulsion Laboratory, California Institute of and Space Technology, under contract with the National Aeronautics Administration

Outline of Talk

- Motivating NASA Missions
- Why Safe Learning
- Work at JPL in Adaptive Problem Solving

Motivation

- NASA wants to explore hostile, unpredictable environments
- Ambitious Science Goals require survival for extended durations (up to decades)
- In order to meet these mission requirements adaptation is *required!*

A Motivating Example

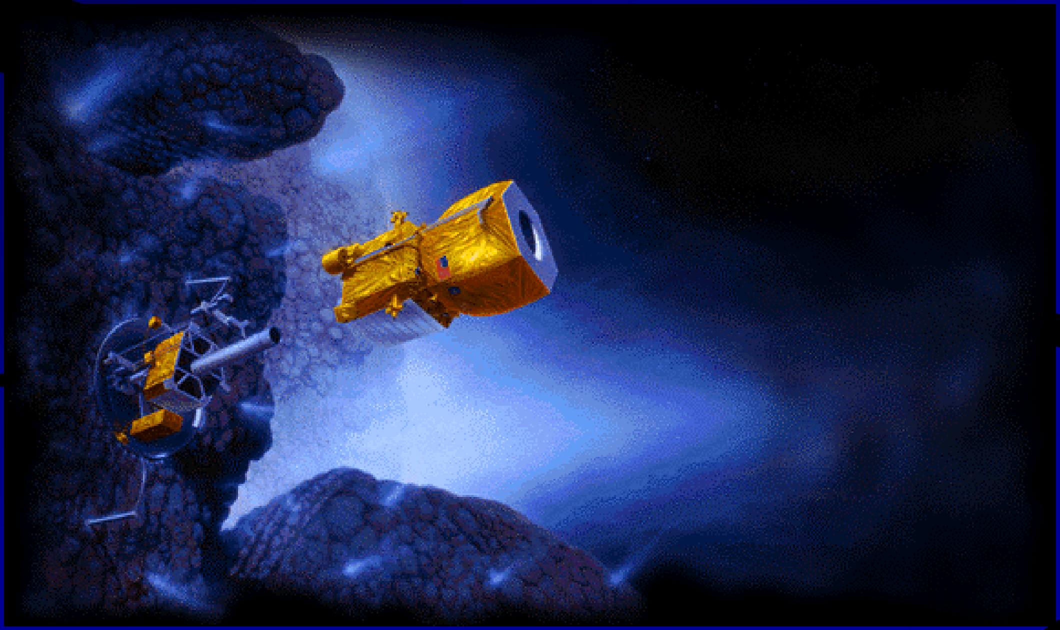
Unknowns and Impact on Operations

- Thickness and composition of ice-cap
  - energy expended to penetrate surface
  - data volume and type collected
  - ability to communicate while below cap (reliability, rate)
  - effectiveness of melting strategies (fast v. slow)
- Properties of underground ocean
  - energy and time cost to move/explore
  - effectiveness of sensors (reliability, range, discriminability)
  - ability to communicate
  - predictability of above

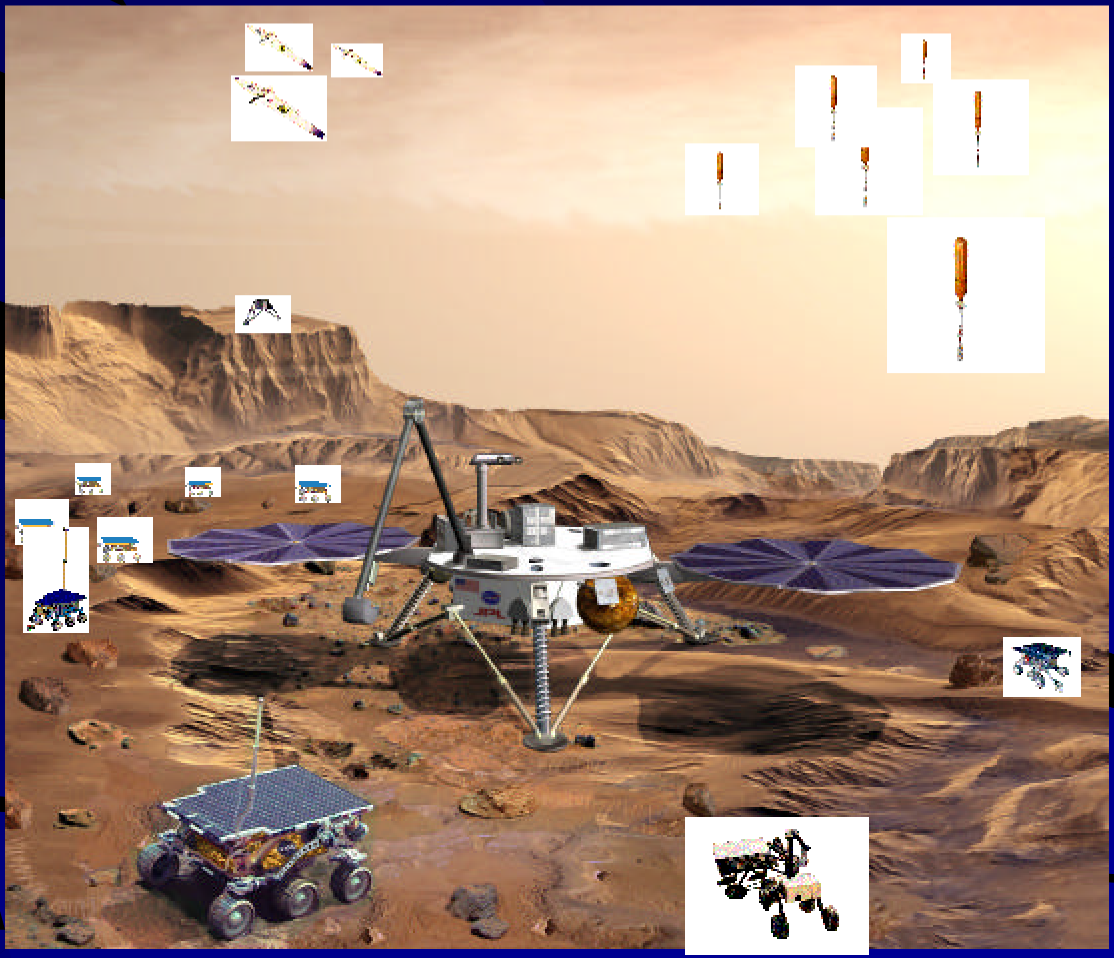
Comet Lander

Examples of Unknowns and Impact on Planning

- Hardness of surface
  - time to drill to specified depth
  - power consumption of drilling activities
- Outgassing properties of comet under solar illumination
  - affects lighting for pictures
  - may affect communications links



Mars Robotic Outpost



- Long-term environmental changes (general warming trend)
- Medium-term environmental changes (seasons)
- Shorter-term environmental changes (storms)
- Hardware degradation
- Communications performance
- Mobility
- Sensor effectiveness

…

Adaptive, self-organizing Exploration Agents

- conduct extended (decades long) environmental and geological Martian survey

Learning is Key

- To adapt performance to unknown environment
  - For survivability
  - For efficient operations
    - unknown effectiveness of operations
- To adapt to changing environment
  - Climate, seasons
  - Shorter term variations (storms, day/night)
- To adapt to hardware degradation
  - Even more important for swarms
    - role assignment

Safe Learning is Key

- Flight Project Community is very risk averse
  - Some reward for accomplishing more science
  - Huge penalty for loss of mission
- Missions represent enormous investment
  - Smaller missions $300M range
  - Larger missions $1B range
  - All represent ~10 year investment of institution (conception to completion of operations)
- Decision to use technology is based on
  - Gain from technology use
  - Risk from technology use (mission loss, cost, schedule)!

Important Classes of Learning

- Off-line
  - Train on large datasets off-line to optimize eventual on-line performance
    - Requires data (simulator)
    - How to get training set, what if training set not realistic
    - Validate performance on dataset, realism of dataset
- On-line
  - Learn "on the fly"
    - Can adapt to unpredicted variations
    - Can learning keep up with variation rate (learning ~ variation rate mismatch)
    - Validate learning algorithms reliability!
- Hybrids possible
  - On-line adapt to strategies learned off-line
    - Validate off-line algorithm performance <u>and</u> on-line selection

Summary - Motivation

- NASA has critical need for learning systems
- New classes of missions are enabled by effective, safe learning
- Timescale of these missions is such that the technology must be mature in ~2005 timeframe
  - Tremendous opportunity for these technologies and for NASA

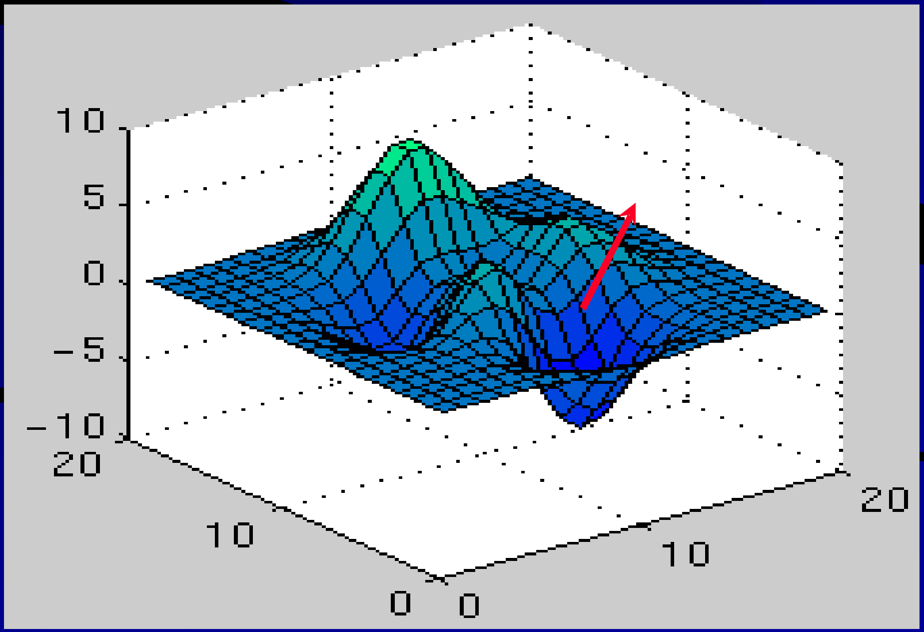
An Example of Learning at JPL: Adaptive Problem Solving as Stochastic Optimization

Adaptation for Autonomy

- Cannot construct optimal control strategy for autonomous spacecraft before mission
  - Knowledge about the spacecraft environment is required
  - Domain shift or unknown situations may occur during mission (e.g., changing environment, spacecraft degradation, failures, ...)
- ***Adaptive problem-solving*** enables self-modification of the control strategy based upon environmental feedback
- Two parts of adaptation in stochastic environment:
  - generation of candidate control strategies
  - evaluation of control strategies

Stochastic Optimization

- Gradients can provide valuable information to guide search in strategy optimization space



Search space is space of control strategies for problem-solver
Each point in the space is a specific search strategy
The expected utility of a point can only be estimated stochastically because we can only score a strategy on a specific problem and expected utility is average score over an unknown problem distribution

Stochastic Optimization

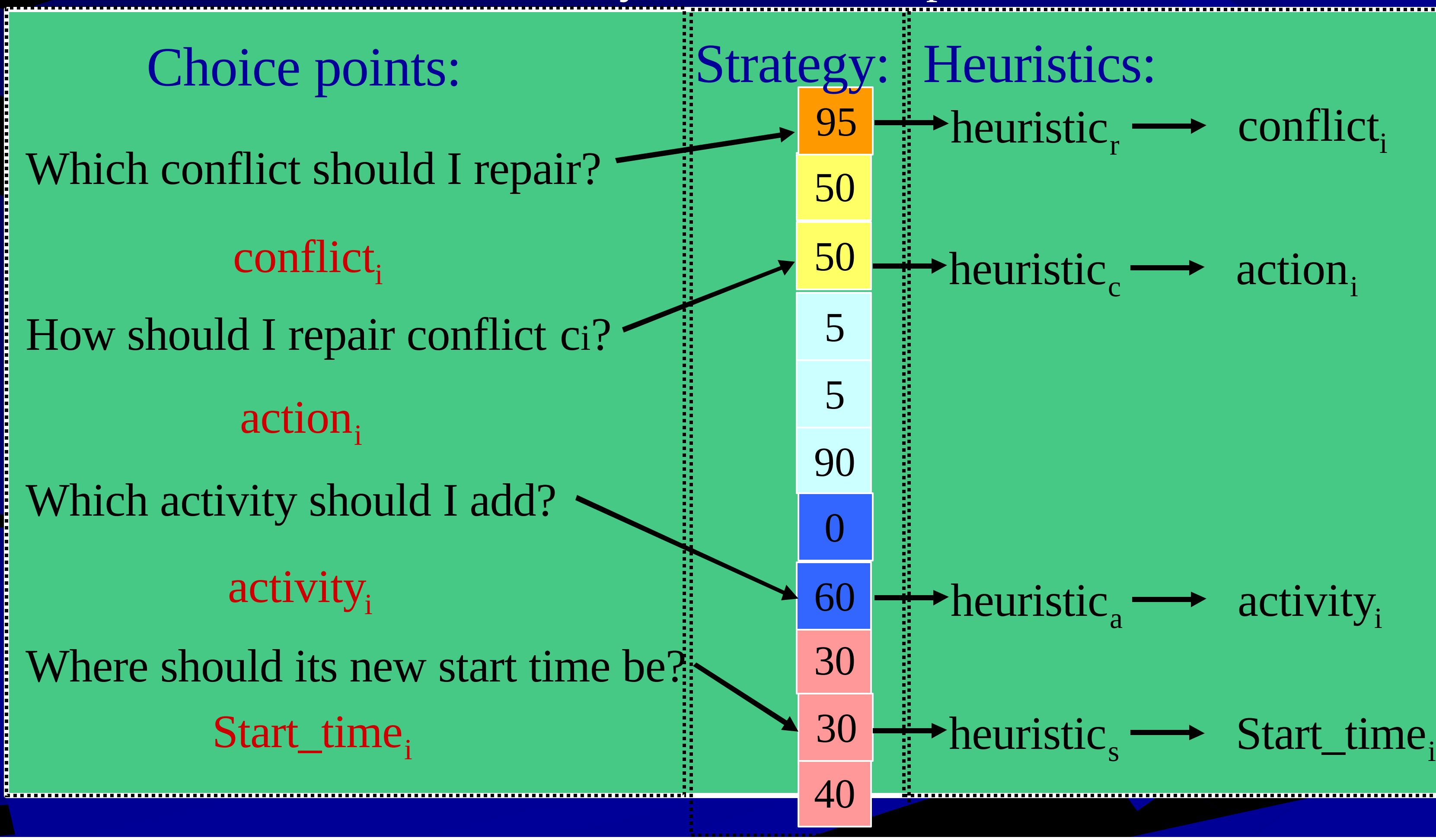
- Knowing surface of search space could help decide search algorithms
- Can local search techniques work?
  - How well do they perform?
  - What are the characteristics (smoothness, local maxima, ...) of the surface defining the search space?
  - How do automated approaches compare to human expert best solutions?

Generic Planning System

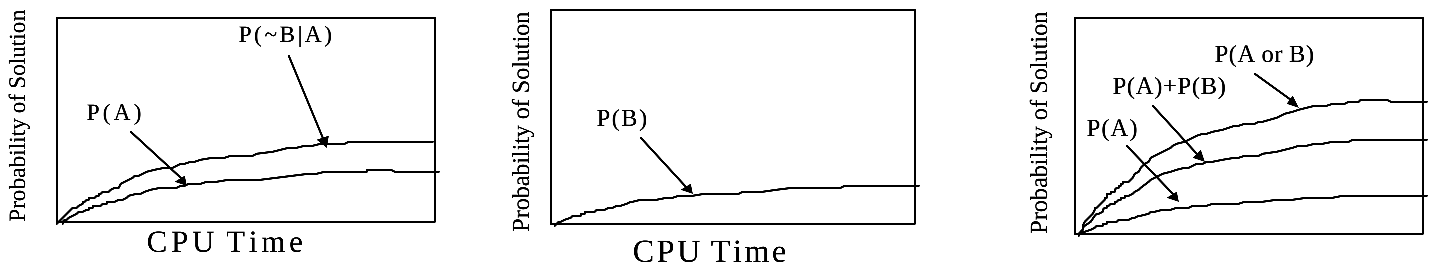
- Adaptive problem solving is applied to the generic planning system ASPEN.
  - Automatically generates a sequence of activities to accomplish input goals.
  - Attacks individual conflicts (related to resources, states, or activity parameters) using iterative repair.
  - Control strategy determined through a set of heuristics which chooses the modifications taken to repair conflicts at certain points in the search (“choice points”)

Strategy Vector

For each iteration of repair, ASPEN makes choices about what repairs to perform. The strategy vector is a set of weights which determines which heuristic to choose stochastically at each choice point.

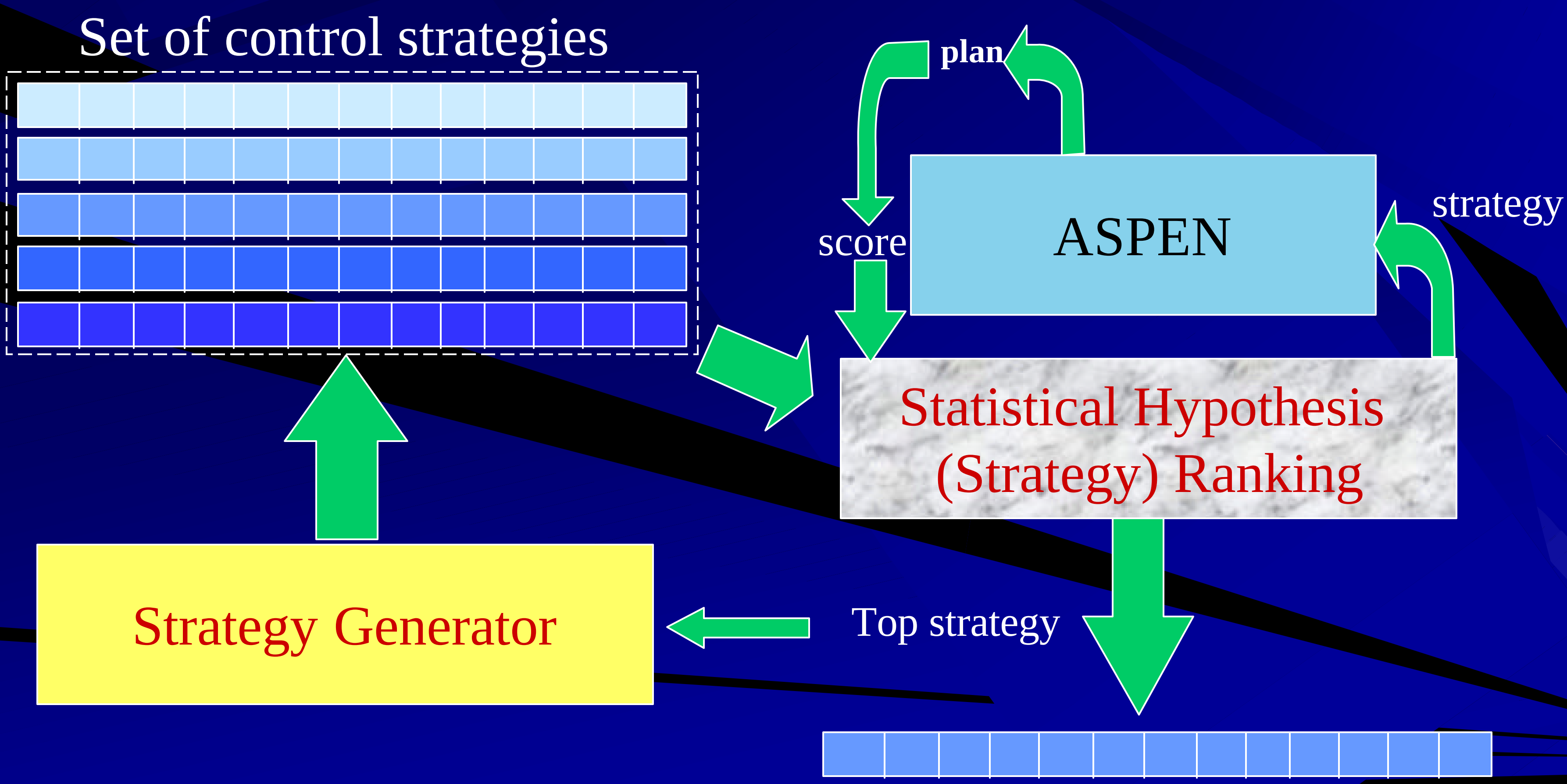


Portfolio Synergy

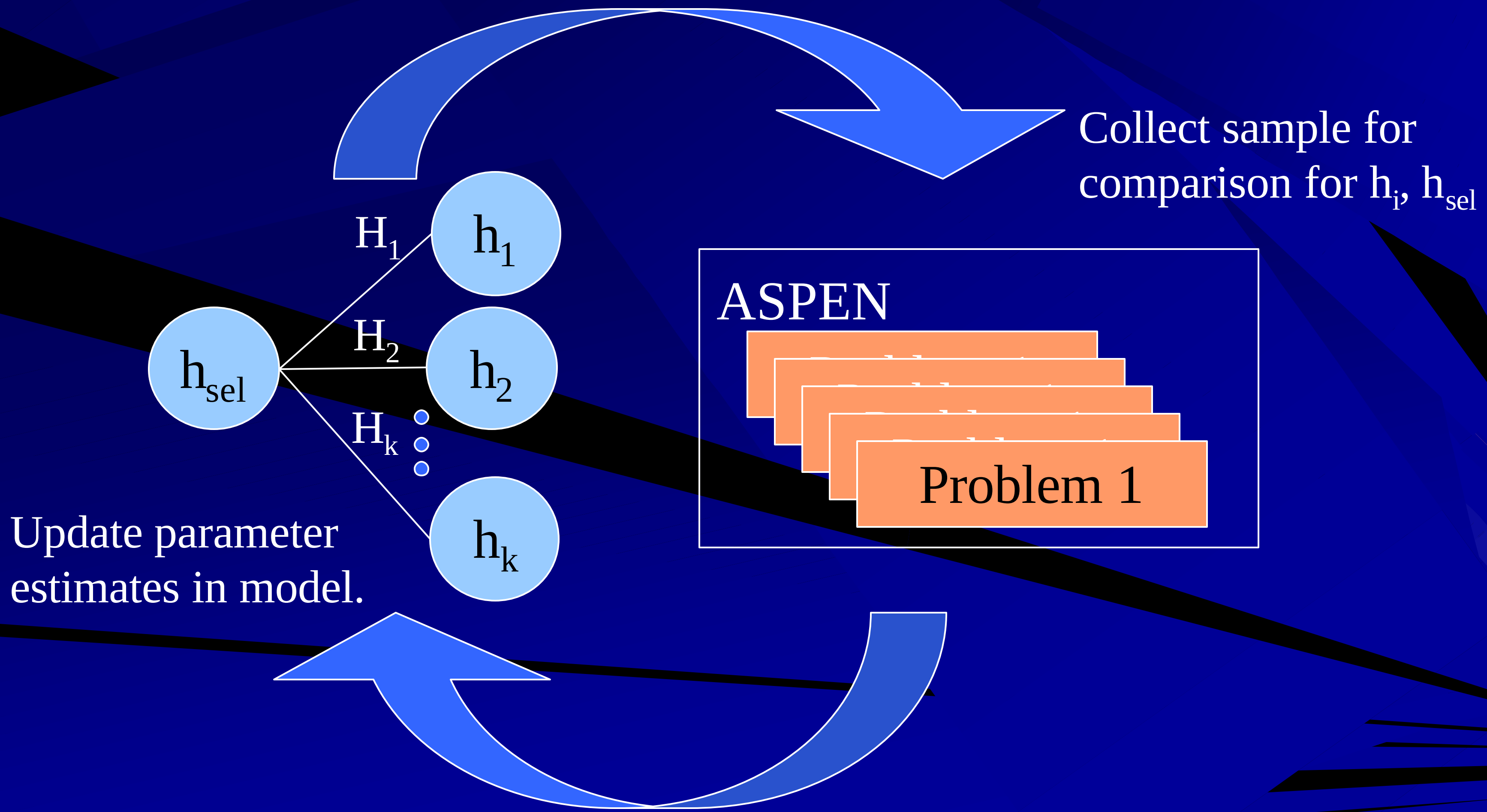


Complementary portfolio algorithms increase Portfolio robustness.

Adaptive Problem Solving



Selection Procedure



Decision Criterion

Probably Approximately Correct (PAC) Requirement:
Hypothesis estimated to be the best must be within some user-specified constant ε distance from the true best hypothesis with probability 1 - δ.

To bound the overall error, we must bound the sum of the errors for the k-1 comparisons:

$$\Pr\left[\bigcup_{i=1}^{n}(\hat{U}(h_i) - \hat{U}(h_{\mathrm{sel}}) > e)\right] < d$$

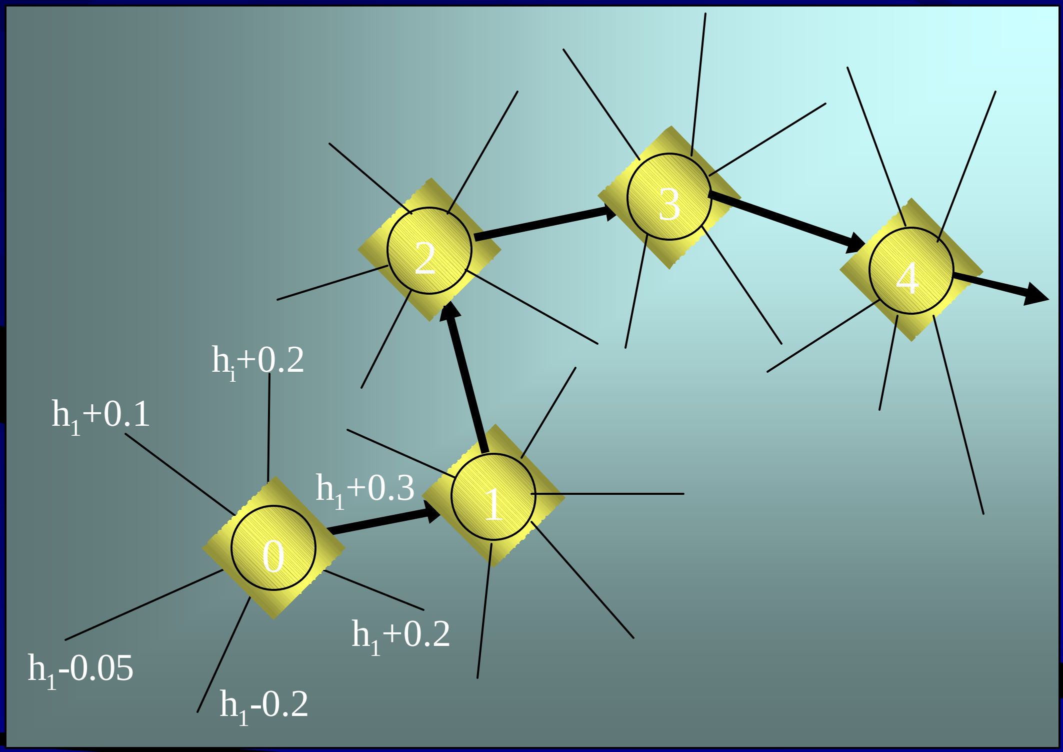
Given the normality assumption, the probability of incorrect selection for a pair-wise comparison, α_i, is:

$$a_i = \Phi\left(-(H_{sel} - H_i)\frac{\sqrt{n}}{\sqrt{s^2_{sel,i}}}\right)$$

We can use this estimate to determine the number of samples needed to achieve a specified error bound:

$$n_{i,sel} = \frac{s^2_{sel,i}}{(H_{sel} - H_i)^2}[\Phi^{-1}(a_i)]^2$$

Search Procedure



At each step, hypothesis h_i is the starting point for step function f_j, which is used to generate the next set of hypotheses to evaluate.

Search Algorithms

- Local Beam Search:
  - Select top *b* hypotheses using PAC with confidence *c*
  - Generate next set of hypotheses in the *neighborhood* of these hypotheses
- Genetic Algorithm:
  - Stochastically choose parent hypotheses, ranked using PAC requirement, based on ranking
  - Using crossover, mutation, and reproduction with given probability, generate offspring from the parents
- Random Search:
  - Hypothesis is a random point from the search space

Earth Orbiter-1

- Science Activities
  - imaging surface targets using advance multi-spectral imaging device.
  - Calibration
- Engineering activities
  - locking solar array drive (SAD)
  - Uplink data
  - Downlink data
  - pointing imaging device
  - maneuvering/rolling spacecraft
  - warm-up/turn on thruster

- Resources
  - Solar array
  - Aperture cover
  - Processor
  - Memory
  - Wideband recorder processor
  - Battery
  - Heater
  - Propellant

Deep Space 4 (CNSR)

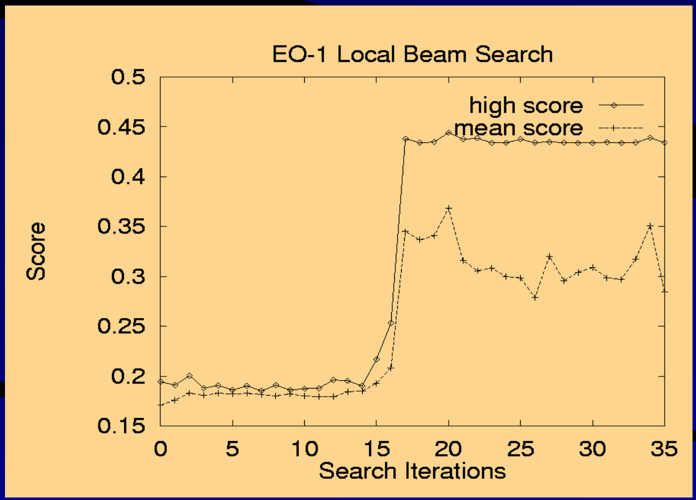
- Sample Activities
  - Move the drill to the hole
  - Drill the hole (mining)
  - Move sample to oven and deposit
  - Use oven (bake sample, take data)
  - Let oven cool down before re-use
- Imaging activities
- Engineering activities
  - uplink from lander to spacecraft
  - compress data in buffer

- Resources
  - Comm system
  - Data Buffer
  - Battery charge level
  - Power Level
  - Drill
  - Camera (CIVA)
  - Oven (2) state *
  - Drill Location
  - Camera state *
  - Comm state *

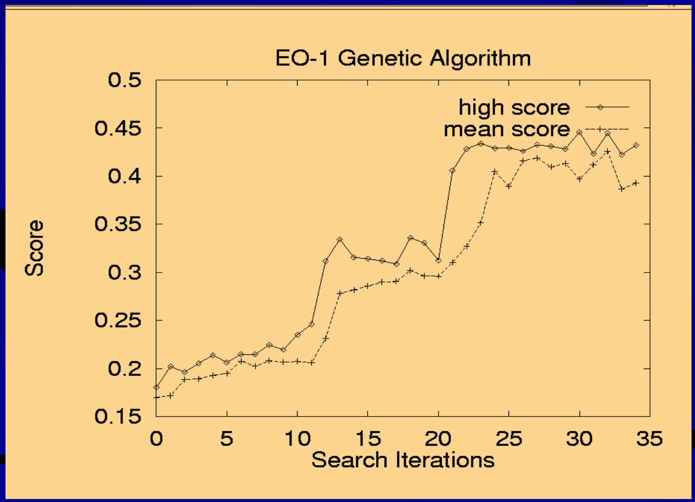
* can be failed

Results

- Evaluated on three spacecraft models: Earth Observer 1 (EO-1), Space Technologies Four Landed Operations (ST-4), Rocky-7 Mars Rover Domain.

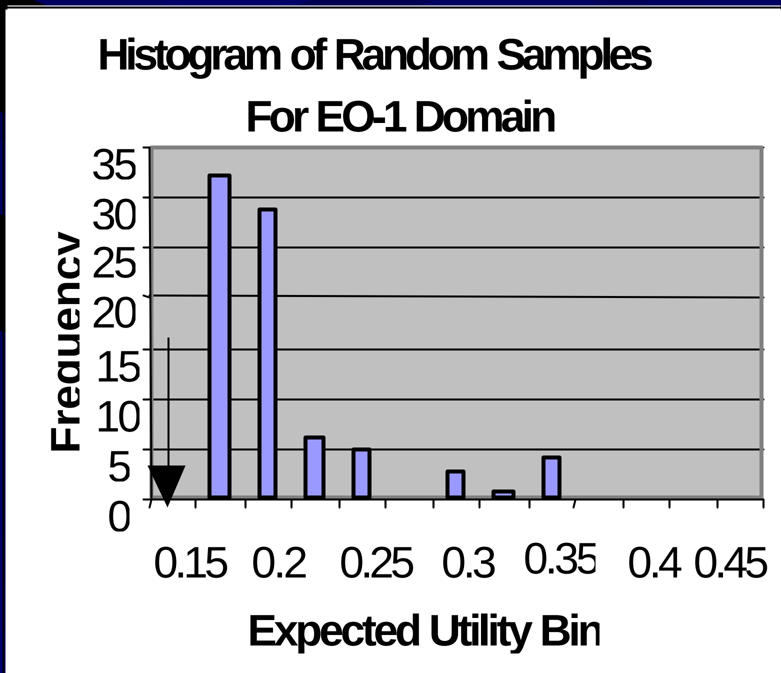


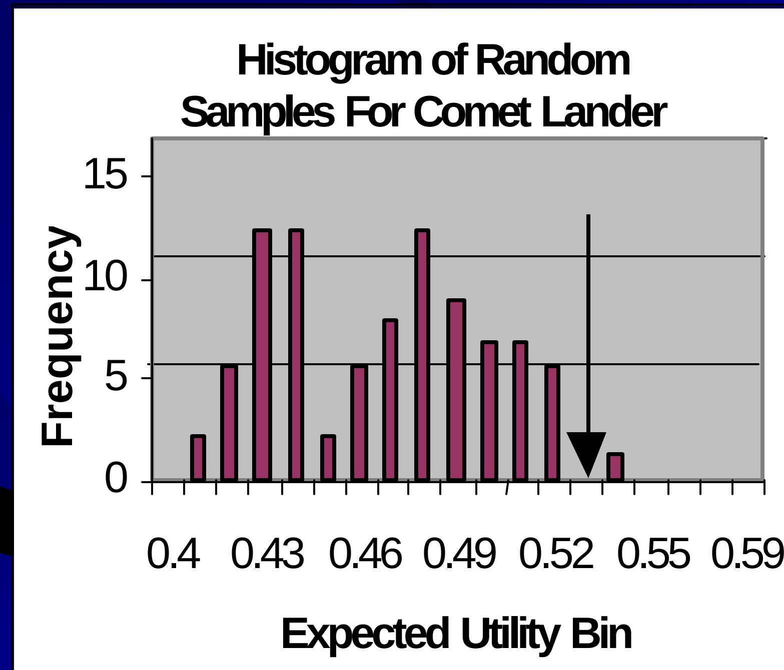
128% improvement in high score from original hypothesis



147% improvement in high score from original hypothesis

Machine vs. Human Expert





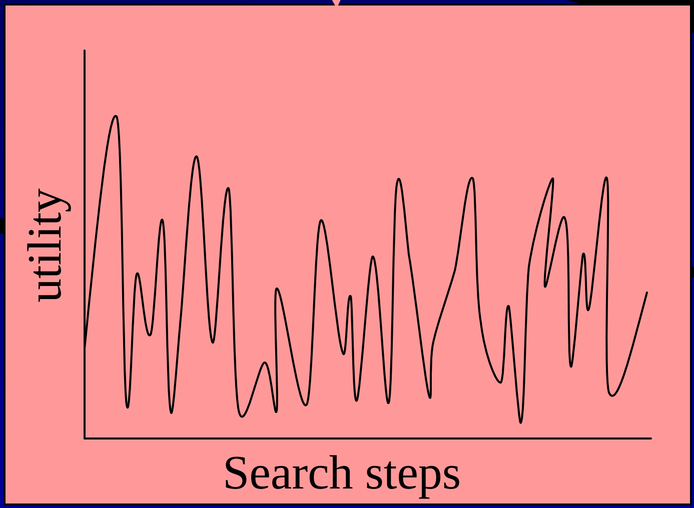
Arrows represent human expert strategy

Notice that in both cases, random sampling found higher scoring hypotheses than the expert with 100 samples

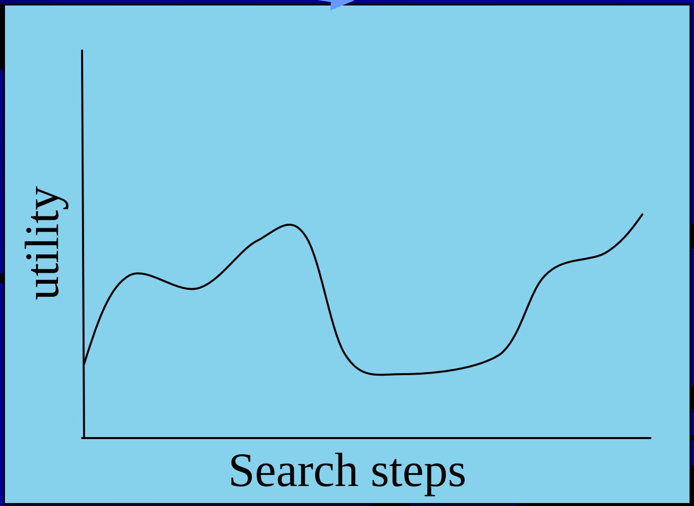
Smoothness Property

Domain	Random Search		Local Beam Search		Genetic Search	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Comet Lander	0.0435	0.0293	0.0086	0.0066	0.0134	0.0093
EO-1	0.0442	0.0466	0.0114	0.0331	0.0145	0.0244

The *mean* shows the average distance between two steps of the search, which is a measure of the smoothness of the step function.



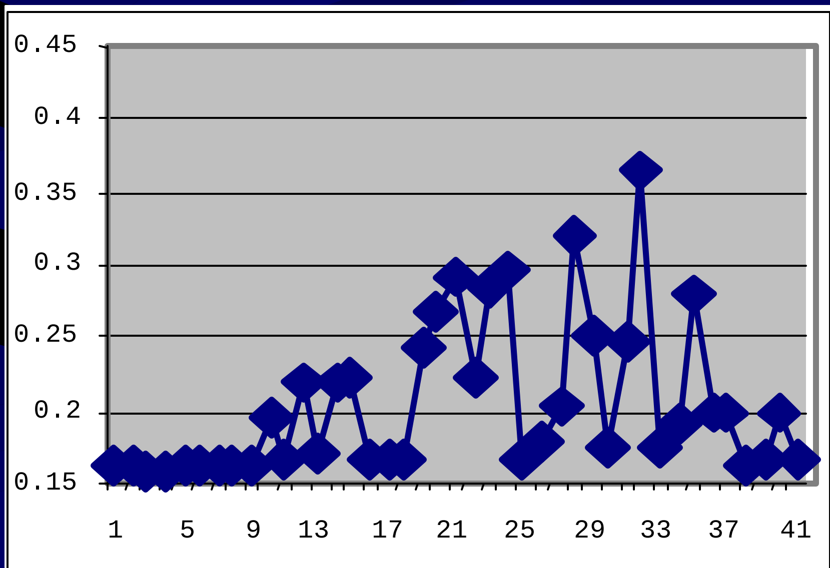
Rougher, less continuous



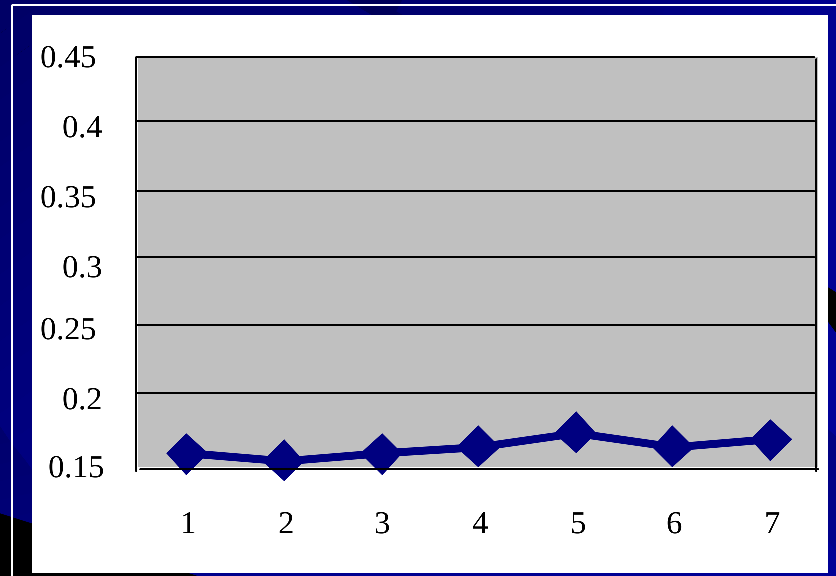
Smoother, more continuous

Valley Hypothesis

Beam search walk for EO-1



Genetic search walk for EO-1



Valley hypothesis: the human expert hypothesis is in a large valley of local minima.

Multiple walks starting at the human expert hypothesis show that it takes many steps to escape the local minimum surrounding the expert strategy

Variable Learning Rates

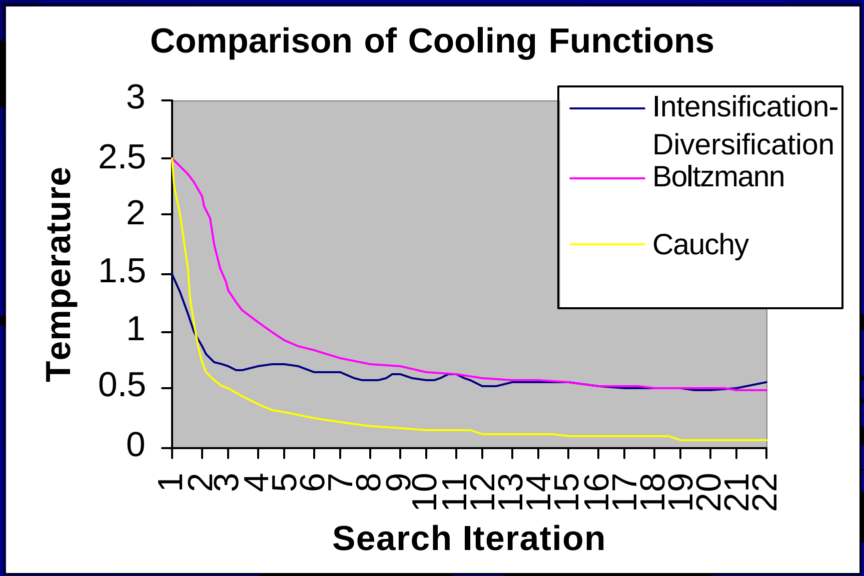
- The purpose of dynamic learning rates is to search the space broadly at first to find a good local basin.
- When a good local basin has been found, restrict the search to find the local optimum there.

Learning Rates Application

- Probability of accepting a suboptimal step (step confidence)
  - For PAC cool selection confidence
- Step size (exploration)

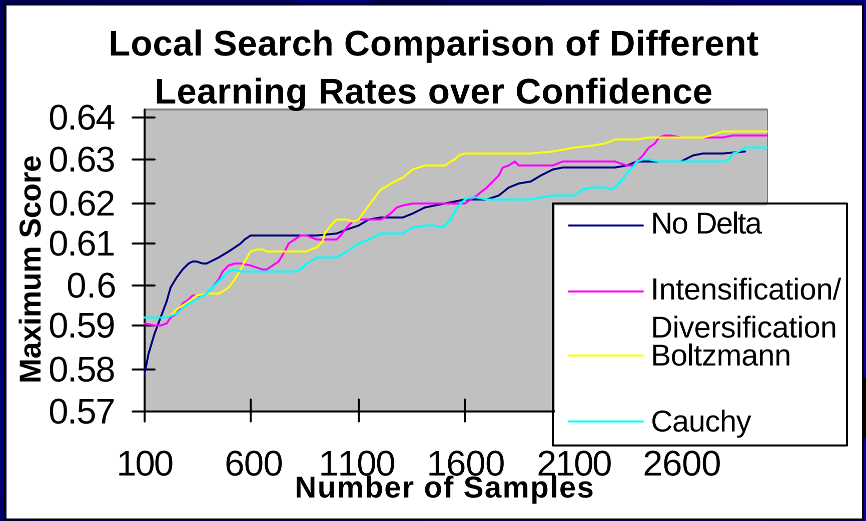
Cooling Rates

- Intensification/Diversification
  – Based on gain from initial step (Tabu)
- Boltzman Annealing
  – Cooling based on temperature analogue
- Cauchy
  – Steeper cooling than Boltzman

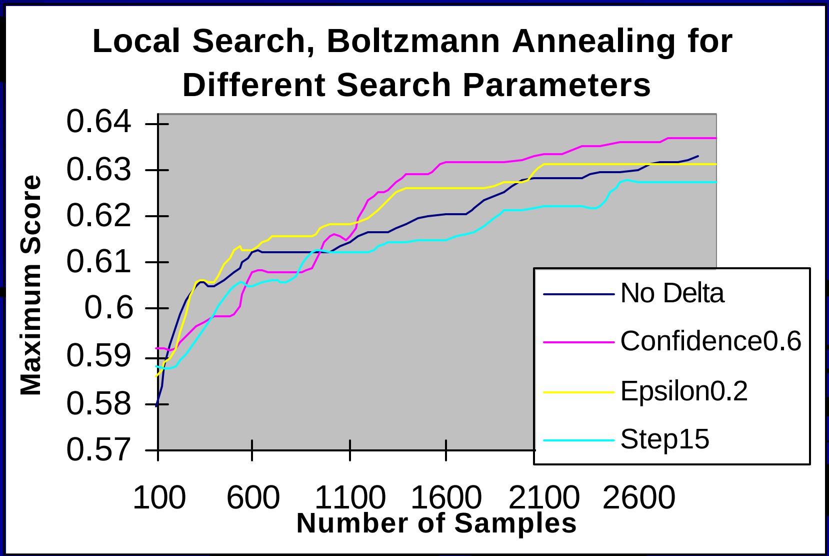


Local Search

- Confidence

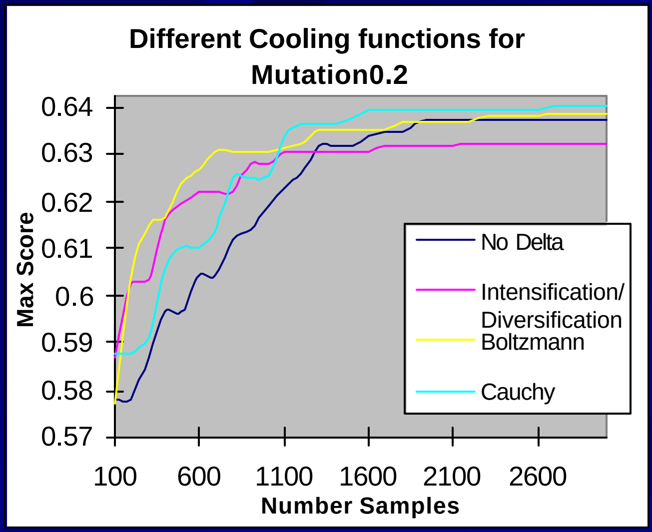


- Boltzman over different parameters

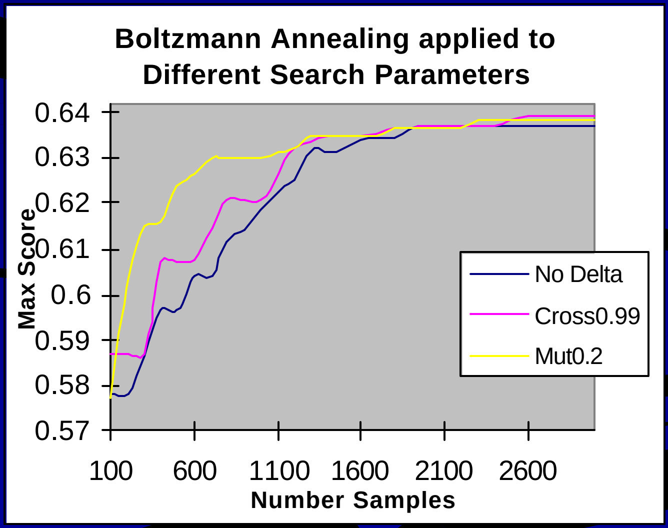


Genetic Search

- Different Cooling for mutation rate

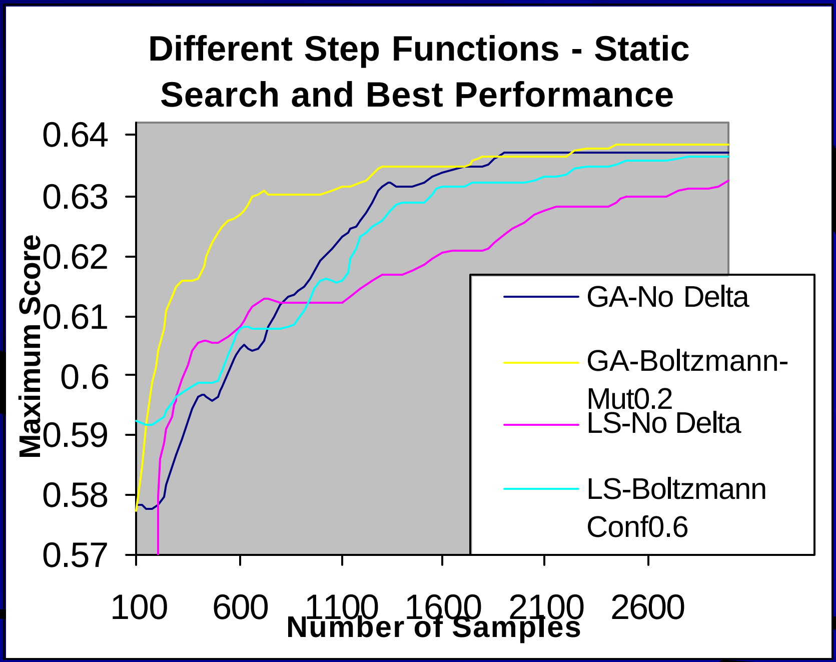


- Boltzman for varying parameters



Learning Rates

- Best performing algorithms comparison
  - Genetic Algorithms perform best
  - Enable learning to occur in roughly half the samples with Boltzman on mutation rate



Current Work

- Continue to perform landscape analysis
  - measure, characterize epistasis of search space, covariance for dimensions
  - number of local maxima, average distance between local maxima
  - correlation distance function
- Include gradient searches
- Use meta-level learning to inform strategy search techniques.

Current Work 2

- Implement and test alternate decision criteria that do not assume normal distribution of hypothesis utilities:
  - Chebyshev's inequality
  - Chernoff bounds
  - Bernstein's inequality

Conclusions

- Safe Learning key to enabling new classes of space exploration missions
- Presented specific work in using:
  - local search for stochastic optimization
  - Applied to choosing heuristics for planner
- Results:
  - Improved on human expert solutions in two domains
  - Showed smoothness of search space using two different step functions
  - Illustrated "valley hypothesis," human expert hypothesis actually lie in large local minimum
  - Preliminary results that varying learning rate can improve performance