Adversarial Decision-Making

Brian J. Stankiewicz

University of Texas, Austin
Department Of Psychology &
Center for Perceptual Systems &
Consortium for Cognition and Computation

February 7, 2006

Collaborators

- University of Texas, Austin
  - Matthew deBrecht
  - Kyler Eastman
  - JP Rodman
  - Chris Goodson
  - Anthony Cassandra
- University of Minnesota
  - Gordon E. Legge
  - Erik Schlicht
  - Paul Schrater
- SUNY Plattsburgh
  - J. Stephan Mansfield
- Army Research Lab
  - Sam Middlebrooks

University XXI / Army Research Labs

National Institute of Health

Air Force Office of Scientific Research

Overview

1. Description of sequential decision making with uncertainty.
2. Description of Optimal Decision Maker
   - *Partially Observable Markov Decision Process*
3. Adversarial Sequential Decision Making Task
   - Variant of "Capture the Flag"
   - Empirical studies comparing human performance to optimal performance in Adversarial Decision Making Task.
4. Future Directions and Ideas
   - How to model and understand "Policy Shifts"

Sequential Decision Making with Uncertainty

- Many decision making tasks involve a sequence of decisions in which actions have both immediate and long-term effects.
- Certain amount of uncertainty about the true state.
- True state is not directly observable but must be inferred from actions and observations.

SDMU: Examples

- Medical diagnosis and intervention
- Business investment and development
- Politics
- Military Decision Making
- Career Development

Questions

- How efficiently do humans solve sequential decision making with uncertainty tasks?
- If subjects are inefficient, can we isolate the *Cognitive Bottleneck*?
  - Memory
  - Computation
  - Strategy

SDMU: Problem Space

1. Interested in defining problems such that 'rational' answers can be computed.
2. Allows us a 'benchmark' by which to compare humans
3. Partially Observable Markov Decision Process

Standard MDP Notation

- S: Set of states in the domain
  - Set of possible ailments that a patient can have.
  - E.g., Cancer, cold, flu, etc.
- A: set of actions an agent can perform
  - E.g., Measure blood pressure, prescribe antibiotics, etc.
- O: $S \times A \rightarrow O$ set of observations generated
  - "Normal": Blood pressure.
- T: $S \times A \rightarrow S'$ (transition function)
  - E.g., Probability of becoming "Healthy" given antibiotics.
- R: $S \times A \rightarrow \Re$ Environment/Action Reward
  - $67.00 to measure blood pressure

Putterman 1994

Belief Updating

$$p(s'|b,o,a) = \frac{p(o|s',b,a)p(s'|b,a))}{p(o|b,a)} \tag{1}$$

- Update current Belief given the previous action (a) and current observation (o) and the belief vector (b).
- E.g., "What is the likelihood that the patient has cancer given that his/her blood pressure is normal?"
- Belief is updated for all possible states.

Computing Expected Value

$$V(b) = \max_{a \in A} \left[\rho(b, a) + \sum_{b' \in B} \tau(b, a, b') V(b') \right] \tag{2}$$

- $\rho(b, a)$: Immediate reward for doing action a given the current belief b.
- $\tau(b, a, b')$: Probability of transition to new belief (b') from current belief (b) given actions a.
- $V(b')$: The expected value in the new belief state b'.
- Optimal observer chooses the action that maximizes the expected reward.

Tiger Problem

1. Tiger Problem
   - Simple example of Sequential Decision Making under Uncertainty task.
   - Illustration to provide intuitive understanding of **POMDP** architecture.

Tiger Problem: States

- Two doors:
  - Behind one door is Tiger
  - Behind other door is "pot of gold"

Tiger Problem: Actions

- Three Actions:
  1. Listen
  2. Open Left-Door
  3. Open Right-Door

Tiger Problem: Observations

- Two Observations:
  1. Hear Tiger Left ($Hear_{Left}$)
  2. Hear Tiger Right ($Hear_{Right}$)

Observation Structure

$$p(Hear_{Left} | Tiger_{Left}, Listen) = 0.85$$

$$p(Hear_{Right} | Tiger_{Right}, Listen) = 0.85$$

$$p(Hear_{Right} | Tiger_{Left}, Listen) = 0.15$$

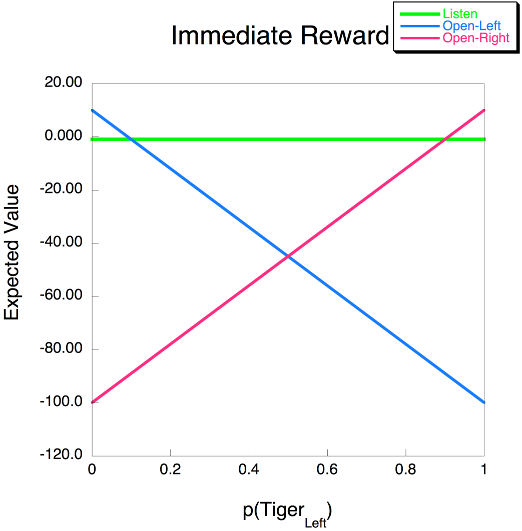
$$p(Hear_{Left} | Tiger_{Right}, Listen) = 0.15$$

Tiger Problem: Rewards

Table: Reward Structure for Tiger Problem

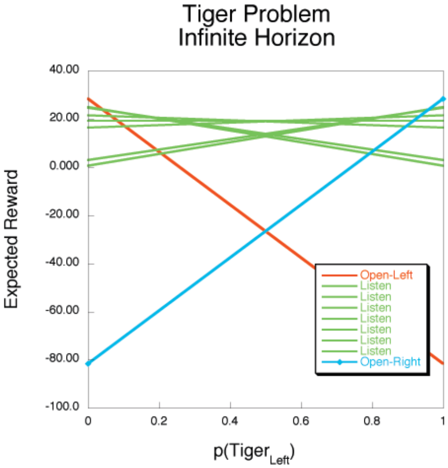
	Tiger=Left	Tiger=Right
Listen	-1	-1
Open-Left	-100	10
Open-Right	10	-100

Tiger Problem: Immediate Reward



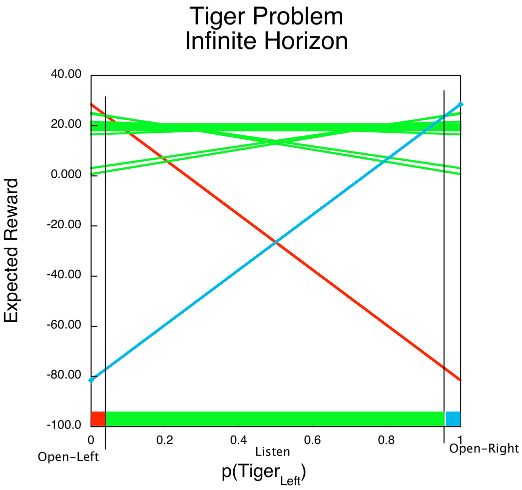
- Immediate Rewards.

Tiger Problem: Expected Reward



- Expected reward functions for multiple future actions with an infinite horizon.

Tiger Problem: Policy



- From expected reward, generate the optimal *Policy* (π).
- The policy chooses the action (a) that maximizes the expected reward for the current belief.

Tiger Problem: Policy

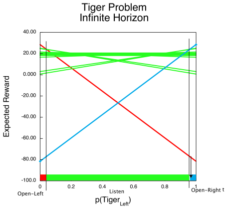


Table: Belief Updating for Tiger Problem

Act. Num	Action	Observation	$p(Tiger_{Left})$
0	—-	—-	0.5
1	Listen	$Hear_{Left}$	0.85
2	Listen	$Hear_{Left}$	0.9698
3	Open-Right	*Reward*	0.5

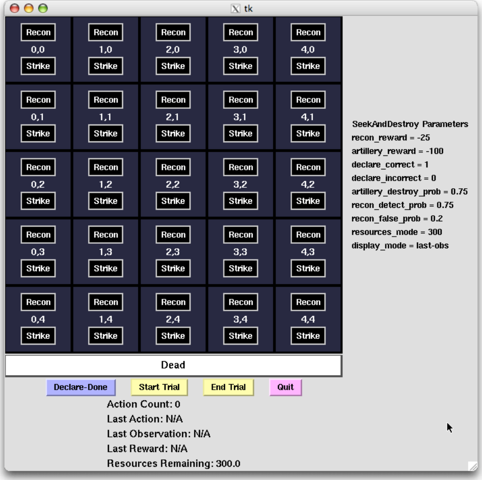
POMDP: Computing Expected Value

1. Using a POMDP we can generate the optimal policy graph for a **Sequential Decision Making Under Uncertainty Task**.
   - Policy graph provides us with the optimal action given a belief about the true state.
2. Using a POMDP we can compute the **Expected Reward** given the initial belief state and optimal action selection.
   - Using the optimal expected reward structure we can compare human performance to the optimal performance.
   - By comparing human behavior to the optimal Expected Reward we can get a measure of **efficiency**.

Empirical studies

1. Capture The Flag
   - Enemy is attempting to capture your 'flag'.
   - Locate and "destroy" enemy before flag is captured.
   - When enemy is destroyed 'Declare' Mission Accomplished.
   - Maximize reward.

Capture The Flag: Task



- 5x5 arena
- Single, enemy
- *Reconaissance* to any of the 25 locations
- Artillery to any of the 25 locations
- Enemy starts in upper-two rows.
- **Goal**: Locate & Destroy the enemy before reaching flag.

Capture The Flag: Task

- Observations:
  - 'Correct Identification': $p(\text{``}Positive''|Enemy) = 0.75$
  - 'False Alarm': $p(\text{``}Positive''|NoEnemy) = 0.20$
- Actions:
  - 'Likelihood of Destroying Enemy': $p(Destroyed|Enemy =< x, y >, Strike =< x, y >) = 0.75$
  - 'Probability that the Enemy will Move': $p(EnemyMove) = 0.2$
- Rewards:
  - $Reward(\text{``}DeclareFinished''|Destroyed) = 1000$
  - $Reward(\text{``}DeclareFinished''|NotDestroyed) = -2500$
  - $Reward(Artillery) = -100$
  - $Reward(Reconnaissance) = -25$

Capture The Flag: Questions

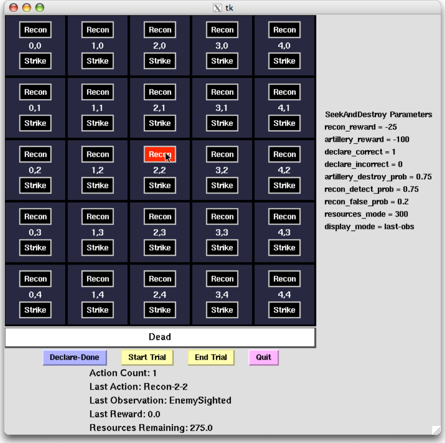
- Test the following possible cognitive limitations:
  1. **Memory Limitation**?
  2. **Belief updating**?
  3. **Suboptimal Decision Strategy/Policy**?

Capture The Flag: Design

- Three conditions:
  1. Only last observation (Baseline)
  2. All observations (Memory)
  3. Belief Vector (Belief Updating)

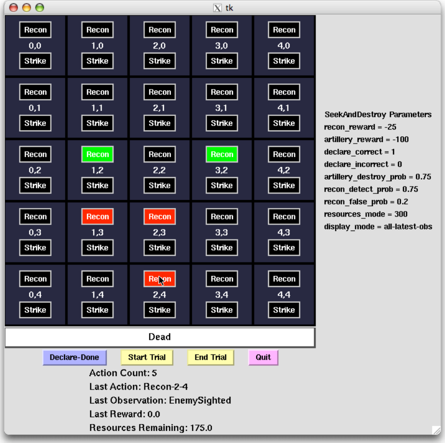
Capture The Flag: Conditions

Last Observation



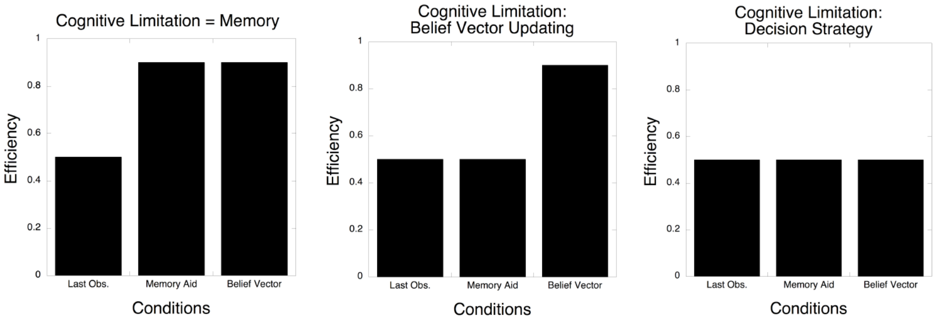
Capture The Flag: Conditions

All Observation



Capture The Flag: Conditions

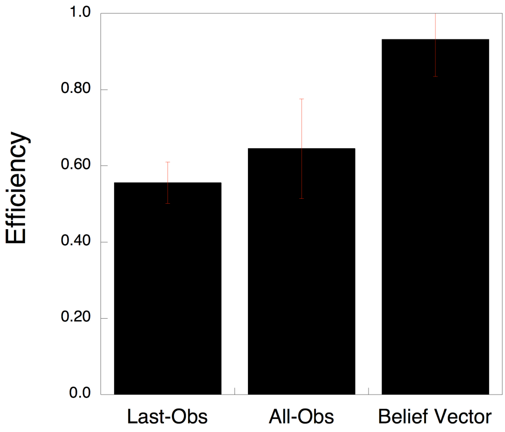
Capture The Flag: Predictions



Capture The Flag: Methods

- 6 subjects (4 Male)
- 60 Trials / Condition
- Trials were run in blocks of 15 trials
- Blocks were run in random order
- Within Subjects Design

Capture The Flag: Results



Capture The Flag: Summary

- No significant improvement in performance when memory aid is given (Last-Obs vs. All-Latest-Obs).
- Significant improvement when belief-state was provided.
- Suggests human inefficiency is in belief updating.
- Consistent with previous findings.
  - E.g., Spatial Navigation (Stankiewicz, Legge, Mansfield & Schlicht (in press) JEP:HPP).

Policy Identification

- Current problem: Adversary has a single policy.
- Possible that the Adversary has multiple policies ($\vec{\pi}$).
- Each policy (π_i) generates specific behaviors for the adversary.
- Given observations (o) decision maker can begin to estimate which policy is the adversary's current policy.
- $p(\pi|a, o, b)$

Policy Transitions

- Given that the adversary has multiple policies, how is one chosen?
- Perhaps randomly on each epoch/encounter.
- Perhaps transitions ($T(\pi, E, \pi')$) between policies based on previous epochs/encounters.
- As a decision maker, I may want to *shift* my opponent to a specific policy that benefits me.
- **Question**: Will we find similar findings in this "hierarchical" problem?

Summary & Conclusions

- Developed Optimal Decision Making Model for Capture The Flag Task.
- Studied human sequential decision making performance on the same task.
- Investigated the cognitive limitations associated with *Sequential Decision Making with Uncertainty*.
- Found that a major limitation to optimal decision making is generating and maintaining an accurate belief vector.
- This was true for both Spatial Navigation and for Capture the Flag Tasks

Thank you

Thank You

Capture The Flag: Optimal Policy