Reflections on unresolved psychological problems for a cognitive architecture

Cleotilde (Coty) Gonzalez

Dynamic Decision Making Laboratory
Social and Decision Sciences Department
Carnegie Mellon University
www.cmu.edu/ddmlab

a.k.a. Why we cannot take Newell's vision seriously?

Cleotilde (Coty) Gonzalez

Dynamic Decision Making Laboratory
Social and Decision Sciences Department
Carnegie Mellon University
www.cmu.edu/ddmlab

UNIFIED THEORIES OF COGNITION

Allen Newell, 1990

A single system (mind) produces all aspects of behavior.

ACT-R: A unified theory of cognition

	Declarative Memory	Procedural Memory
Symbolic	Chunks: declarative facts	Productions: If (cond) Then (action)
SubSymbolic	Activation of chunks (likelihood of retrieval)	Conflict Resolution (likelihood of use)

- Symbolic representations of Declarative and Procedural Memories
- Statistical/Mathematical Mechanisms for processing, accessing, retrieving those memories; and learning and adapting behavior

Newell's vision in ACT-R

- We create models in tasks that are more or less complex
- We report accurate fits to human quantitative behavior (e.g., reaction times, error rates) in multiple tasks – not only explain but predict

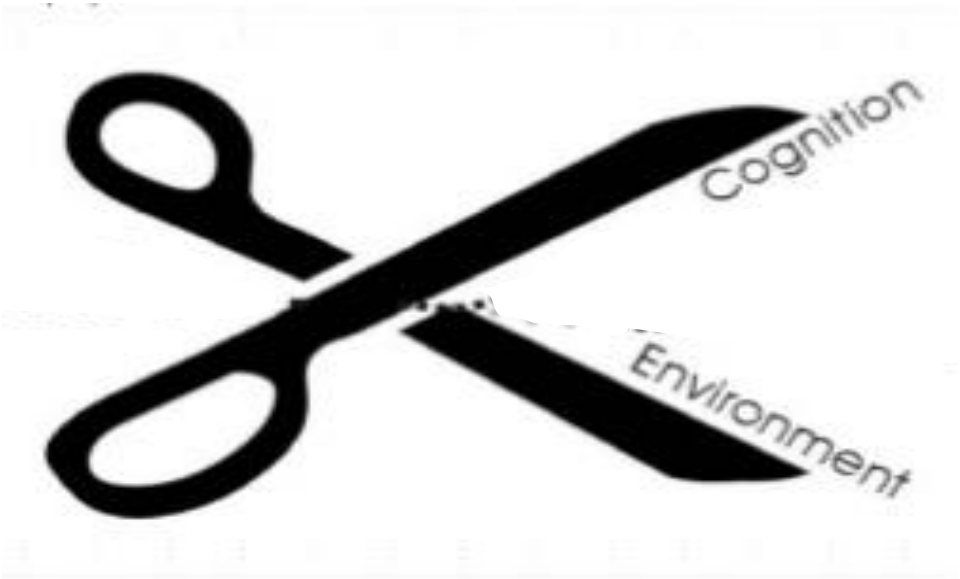
Unified? Integrated?

- Models are most commonly used to explain specific phenomena
- Task-specific cognitive models: Particular examples of human behavior in a particular task

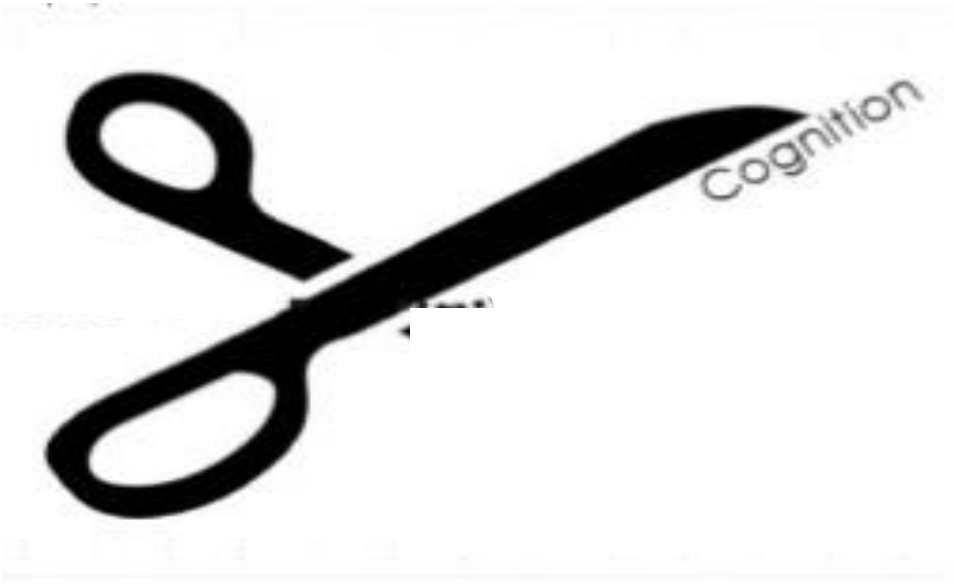
Why we don’t take Newell’s vision seriously??!!

A BEHAVIORAL MODEL OF RATIONAL CHOICE

By HERBERT A. SIMON, 1955



Behavior may be regarded as a characteristic of the decision-maker in a particular environment



Need a *unified theory of environment*

A unified theory of environment

- Task Domain
  - Medicine, Military, Education, Driving…
- Task/Environment Complexity (Campbell, 1988; Wood, 1986)
  - Structural complexity
    - Number of elements: alternatives, attributes, interactions
    - Uncertainty: information level, diversity
    - Constraints: time, load, rate of change
  - Dynamic complexity
    - Relationships between inputs-outputs over time

Demonstration of ACT-R's generality

1. A model able to accomplish same task in different ways

2. A model able to accomplish multiple tasks in the same way

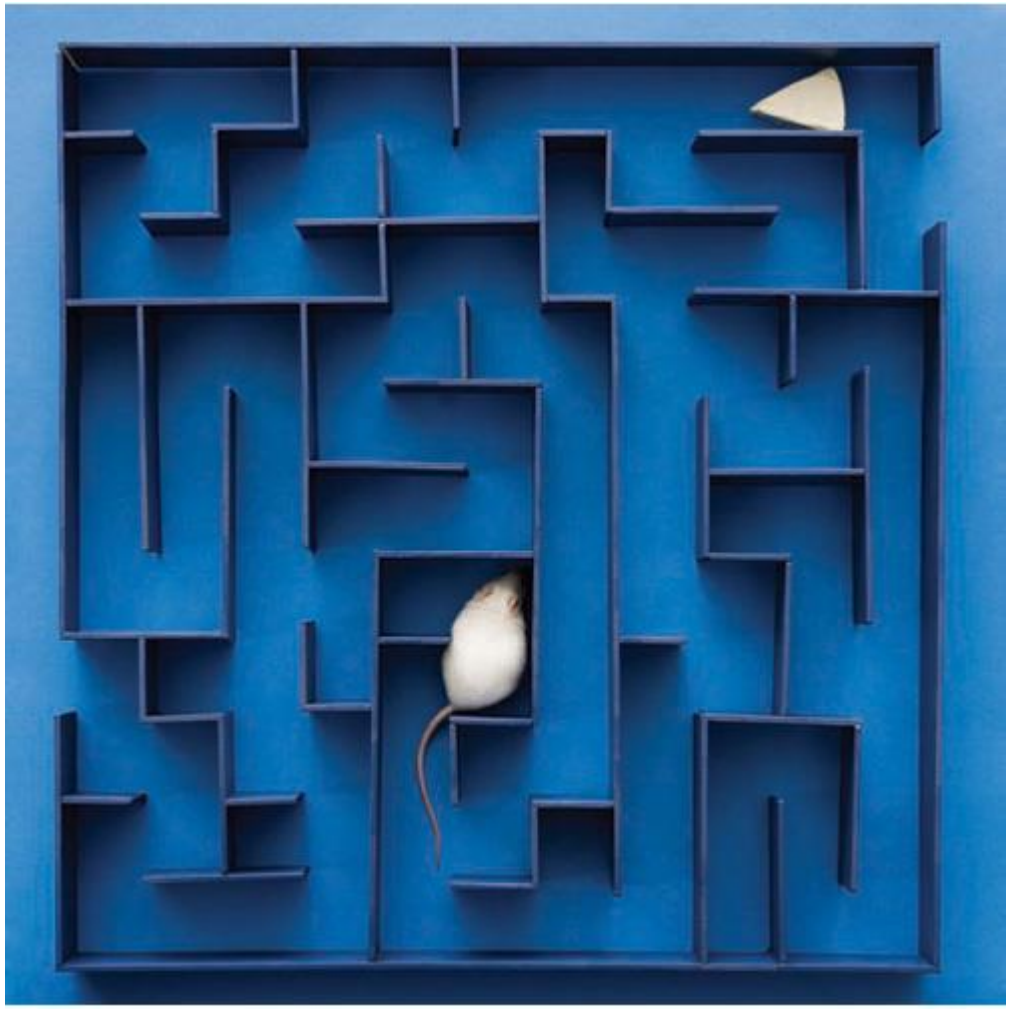
Demonstration of ACT-R's generality

1. A model able to accomplish same task in different ways

2. A model able to accomplish multiple tasks in the same way

Can a model accomplish multiple tasks of different complexity in the same way?

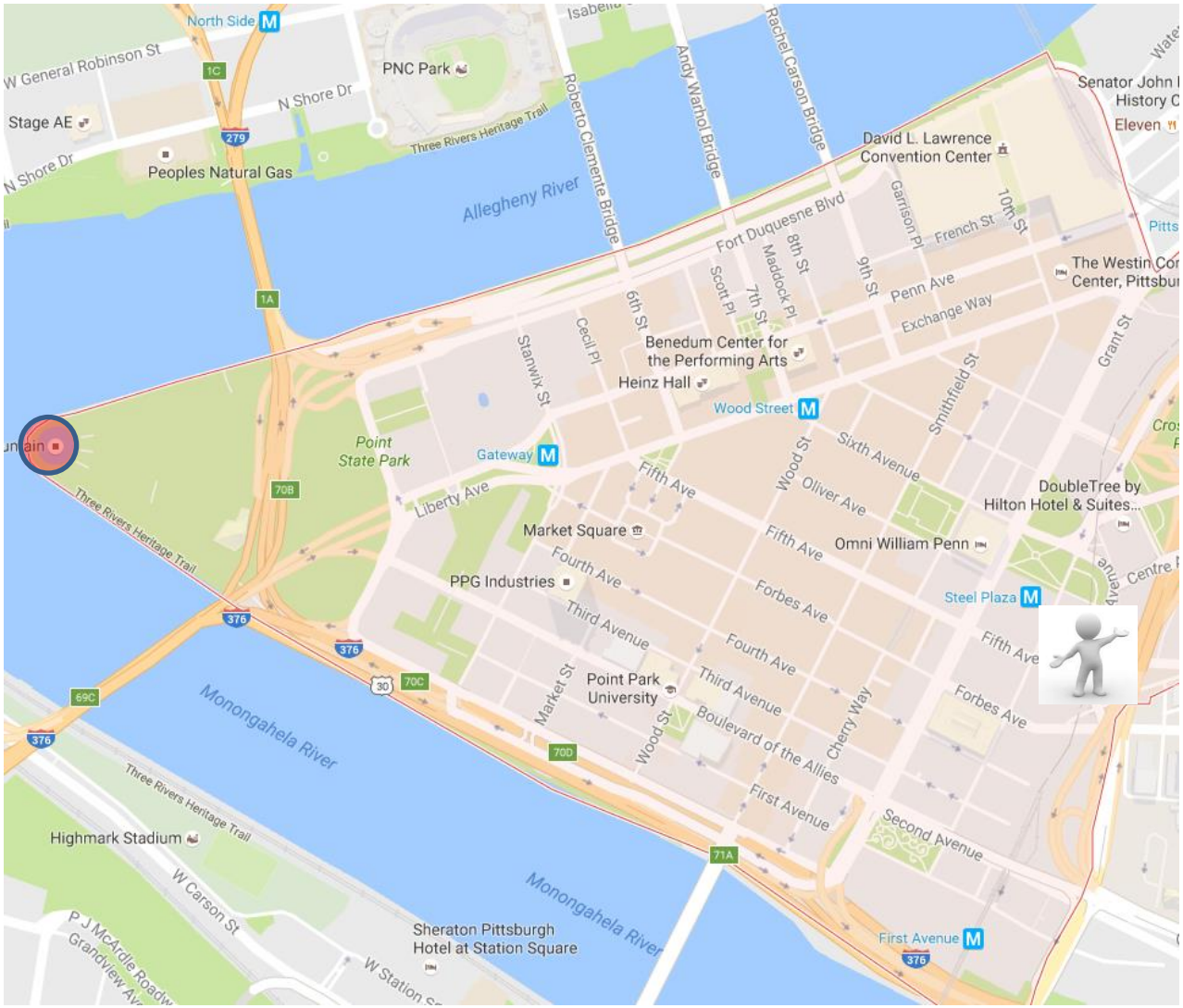
A simple-minded organism in a simple environment



A mouse in a maze

- Goal: to reach the food
- Make a series of decisions (back, forth, right, left, left …) to reach a goal
- Explore the environment and evaluate discrepancies to the goal
- Adapt to changes in the environment and to past decisions made
- Learn from consistency similarities

A complex-minded organism in a complex environment



A human in a city

- Goal: to reach the destination
- Make a series of decisions (back, forth, right, left, left …) to reach a goal
- Explore the environment and evaluate discrepancies to the goal
- Adapt to changes in the environment and to past decisions made
- Learn from consistency similarities

Dynamic Decision Making across levels of complexity

Simple

Least Dynamic

Structural simplicity: binary choice

No changes in the environment: may be probabilistic but probabilities and values don't change over time

Immediate feedback: Action-Outcome closest in time

Value is time independent: Time of the decision is determined by the decision maker, no penalty for waiting

Complex

Most Dynamic

Structural complexity: Multiple options, attributes, and relationships

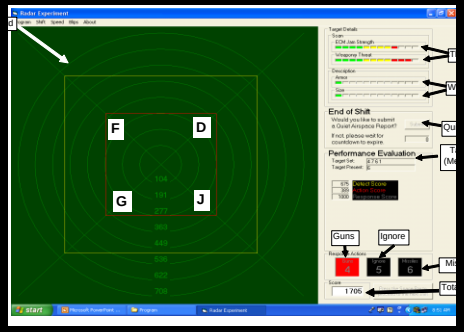
Environment changes Independently and as a consequence of the actions of the decision maker

Delayed feedback and Credit assignment problem (Multiple actions and multiple outcomes separated in time)

Value is time-dependent Value decreases the farther away the decision is from the optimal time

Complex dynamic environments: Microworld research

Gonzalez, Vanyukov & Martin, 2005

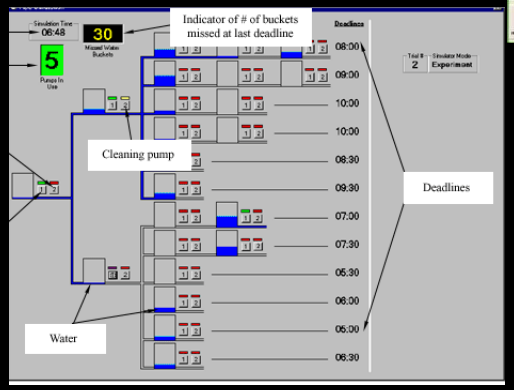


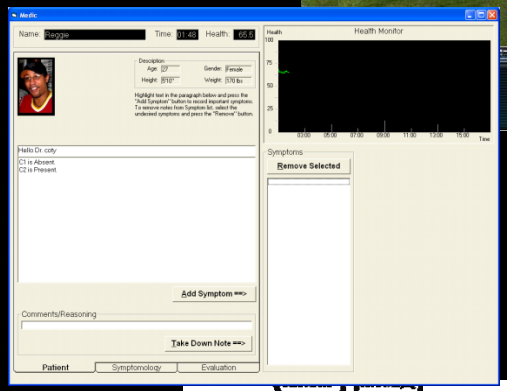
Dynamic Visual Detection

Conflict Resolution

Military Command and Control

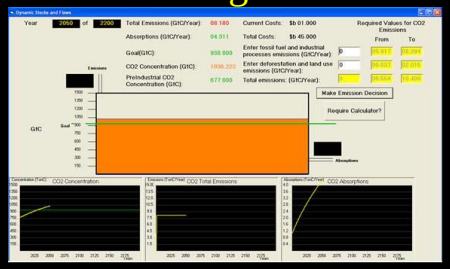
Real-time resource allocation



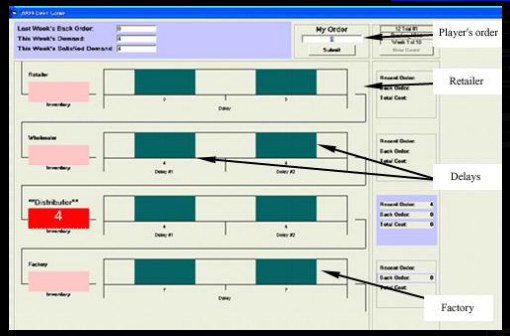


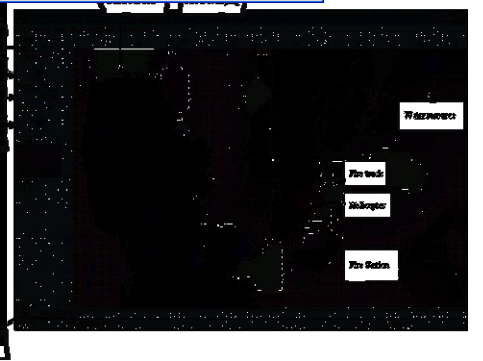
Medical Diagnosis

Climate Change



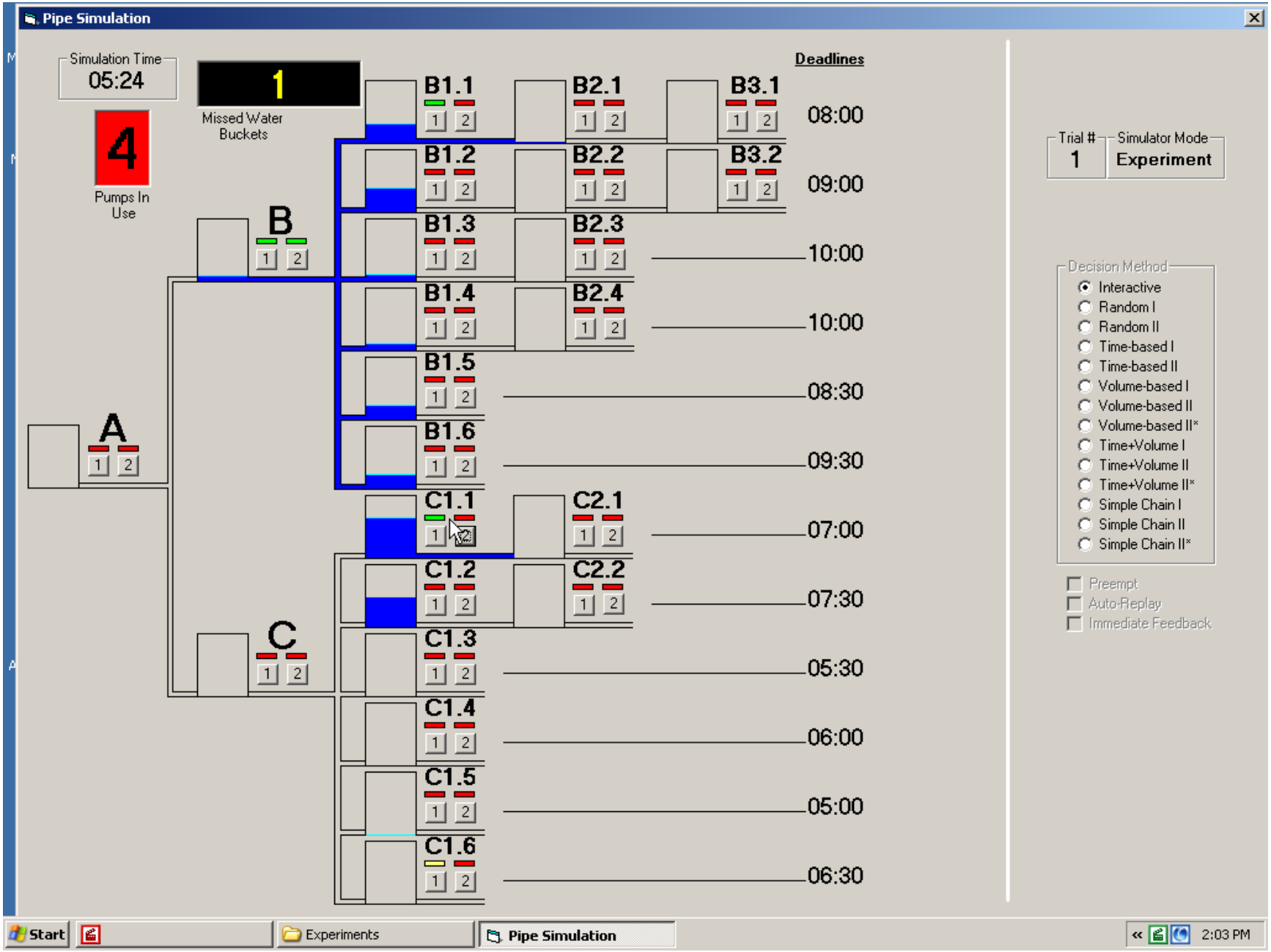
Supply-Chain Management





Fire Fighting

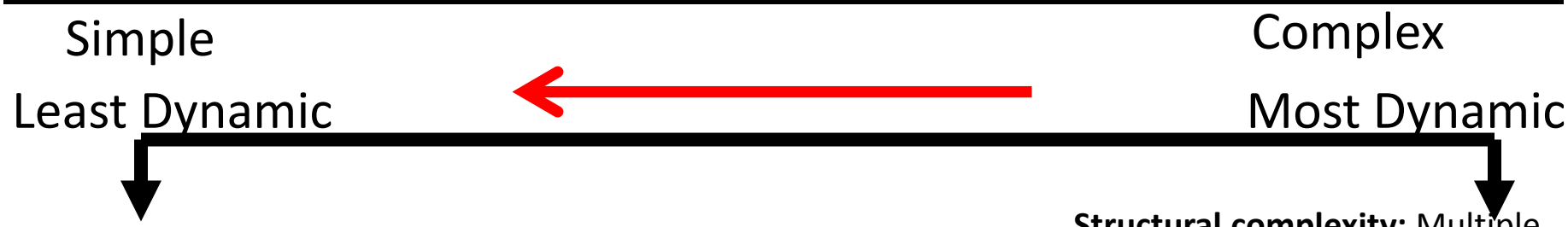
Water Purification Plant (WPP)



All ACT-R's mechanisms for **declarative memory**

Mechanism	Equation	Description
Activation	$A_i = B_i + S_i + P_i - \varepsilon_i$	B_i: Base-level activation reflects the recency and frequency of use of chunk i S_i: Spreading activation reflects the effect that buffer contents have on the retrieval process P_i: Partial matching reflects the degree to which the chunk matches the request ε_i: Noise value includes both a transient and (optional) permanent component (permanent component not used by the integrated model)
Base-level	$B_i = \ln\left(\sum_{j=1}^{n} t_j^{-d}\right) + \beta_i$	n: The number of presentations for chunk i t_j: The time since the jth presentation d: A decay rate (not used by the integrated model) β_i: A constant offset (not used by the integrated model)
Spreading activation	$S_i = \sum_k \sum_j W_{kj} S_{ji}$	k: Weight of buffers summed over are all of the buffers in the model j: Weight of chunks which are in the slots of the chunk in buffer k W_{kj}: Amount of activation from sources j in buffer k S_{ji}: Strength of association from sources j to chunk i
	$S_{ji} = S - \ln(fan_{ji})$	S: The maximum associative strength (set at 4 in the model) fan_{ji}: A measure of how many chunks are associated with chunk j
Partial Matching	$P_i = \sum_k P M_{ki}$	P: Match scale parameter (set at 2) which reflects the weight given to the similarity M_{ki}: Similarity between the value k in the retrieval specification and the value in the corresponding slot of chunk i The default range is from 0 to −1 with 0 being the most similar and −1 being the largest difference
Declarative Retrievals	$P_i = \frac{e^{A_i/s}}{\sum_j e^{A_j/s}}$	P_i: The probability that chunk i will be recalled A_i: Activation strength of chunk i $\sum A_j$: Activation strength of all of eligible chunks j s: Chunk activation noise
Blended Retrievals	$V = \min \sum_i P_i \cdot \left(1 - Sim\left(V, V_i\right)\right)^2$	P_i: Probability from declarative retrieval Sim_{ij}: Similarity between compromise value j and actual value i

A Continuum of Dynamics/Complexity



Structural simplicity: binary choice

No changes in the environment: may be probabilistic but probabilities and values don't change over time

Immediate feedback: Action-Outcome closest in time

Value is time independent: Time of the decision is determined by the decision maker, no penalty for waiting

Structural complexity: Multiple options, attributes, and relationships

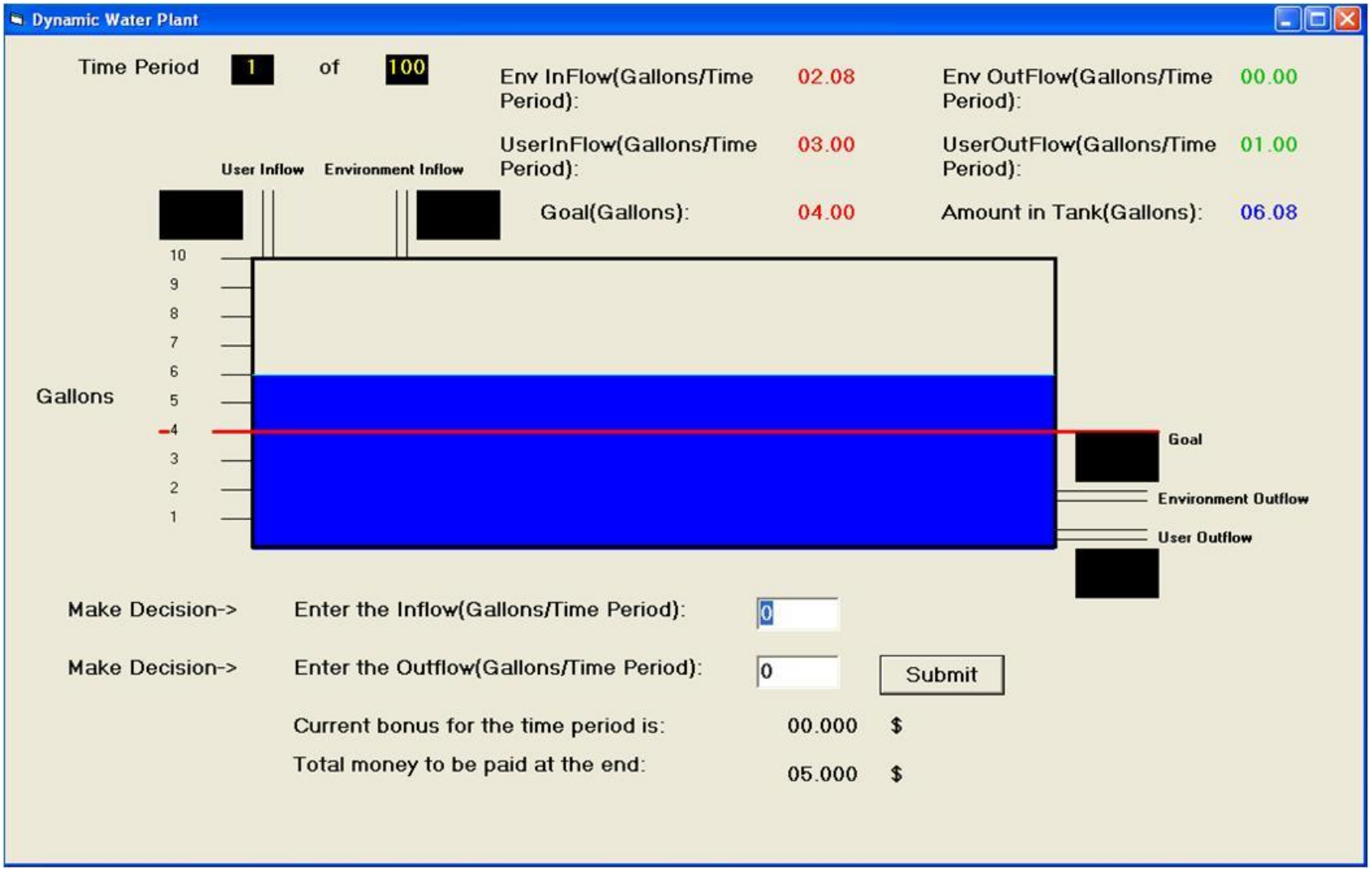
Environment changes Independently and as a consequence of the actions of the decision maker

Delayed feedback and Credit assignment problem (Multiple actions and multiple outcomes separated in time)

Value is time-dependent Value decreases the farther away the decision is from the optimal time

The Beer Game

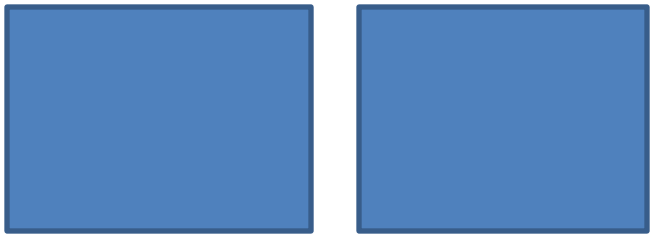
Dynamic Stocks and Flows (DSF)



Choice: Abstract and simple experimental paradigms

Repeated choice Paradigm

Sampling Paradigm



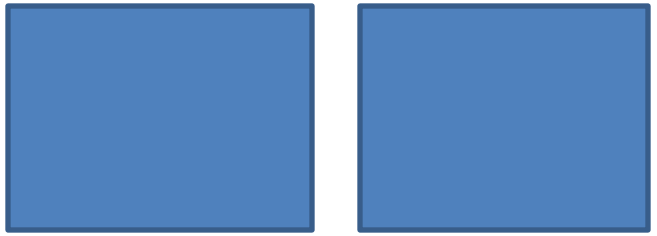
4

4

3

… 0

…..



Sampling

4

4

3

0

Make a choice:

4

Fixed number of trials

Mechanism for **declarative memory:** Base level activation

Cognitive Probability (prob of retrieval from memory)

$$BV = \sum_i P_i \cdot V_i$$

Observed value of outcomes

$$P_i = \frac{e^{A_i/\tau}}{\sum_j e^{A_j/\tau}}$$

Activation

$$A_i = \varepsilon_i + ln \sum_{j=1}^{n} t_j^{-d}$$

Noise

Frequency and *Recency*

Main challenge: Scaling up

Simple

Least Dynamic

?

Complex

Most Dynamic

Structural simplicity: binary choice

No changes in the environment: may be probabilistic but probabilities and values don't change over time

Immediate feedback: Action-Outcome closest in time

Value is time independent: Time of the decision is determined by the decision maker, no penalty for waiting

Structural complexity: Multiple options, attributes, and relationships

Environment changes Independently and as a consequence of the actions of the decision maker

Delayed feedback and Credit assignment problem (Multiple actions and multiple outcomes separated in time)

Value is time-dependent Value decreases the farther away the decision is from the optimal time

UNIFIED THEORIES OF COGNITION

Allen Newell, 1990

Analysis of complex tasks, beyond those involved in simple psychological paradigms

Cybersecurity:

Hyper-complexity - beyond the physical world

Cybersecurity

President Obama unveils **cybersecurity** push in tech indu…
Fortune - 17 hours ago
It may be Friday the 13th, but the White House is hoping a summit o **cybersecurity** and consumer protection will bring a bit of much-needed ...

Cybersecurity Summit Exposes Silicon Valley's Privacy Fears
Wall Street Journal - 16 hours ago

FACT SHEET: Executive Order Promoting Private Sector ...
Highly Cited - Whitehouse.gov (press release) - Feb 13, 2015

Obama's **Cybersecurity** Order Is Meh
Opinion - Gizmodo - 16 hours ago

Obama recruits tech giants for new **cybersecurity** efforts
In-Depth - CBS News - Feb 13, 2015

Obama signs executive order on sharing **cybersecurity** threat ...
Blog - Washington Post (blog) - 16 hours ago

euronews CNET UK AFP Wall Street J... BBC News U.S. News &

Explore in depth (1,156 more articles)

Why Apple's CEO went to Obama's **cybersecurity** summit
Fortune - 18 hours ago
Why did Apple agree to send its CEO to Friday's White House conference on **cybersecurity** — headlined by the President himself — but not ...

Watch Tim Cook Speak At President Obama's Summit On ...
TechCrunch - 20 hours ago

RIT at White House **cybersecurity** summit
Rochester Democrat and Chronicle - 13 hours ago

Apple CEO Tim Cook Speaks at White House **Cybersecurity** Summit
Mac Rumors - 19 hours ago

Watch: Apple CEO Tim Cook talk **cybersecurity** at White House Summit
9 to 5 Mac (blog) - 19 hours ago

Obama arrives in Bay Area for cybersecurity summit

By Aaron Kinney and Josh Richman | *Staff writers*

POSTED: 02/12/2015 07:34:45 PM PST | UPDATED: A DAY AGO

20 COMMENTS

The psychology of security

In the Physical World

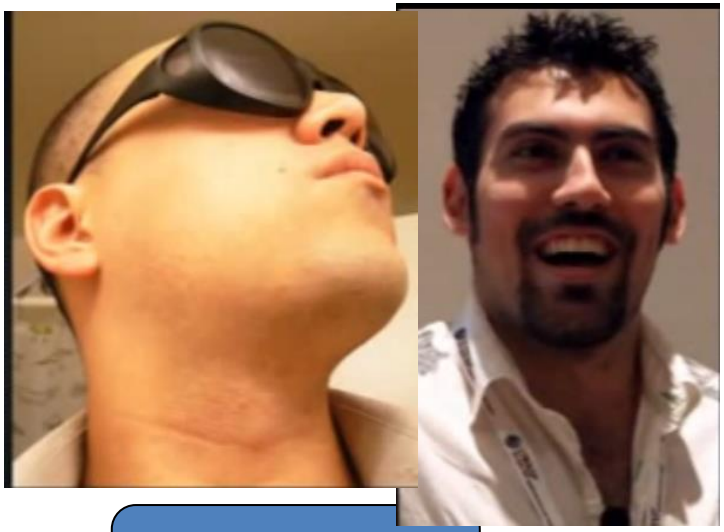
- Laws of physics. Physical weapons (a gun, a knife, a bomb) that we can see, touch, or hear.
- Limited by geography: physical and political borders.
- Change occurs at speeds that we are possible to perceive and process

In the Cyber World

- Use of digital weapons often imperceptible to the human senses.
- An attacker can be in multiple places at the same time, as the same cyber attack can hit multiple targets at once. – no political and regional borders
- A world that is highly dynamic and distributed – hyper-dimensional and variable and imperceptible speeds

Gonzalez, Ben-Asher, Oltramari, Lebiere, 2015

Humans are the weakest link in cyber security



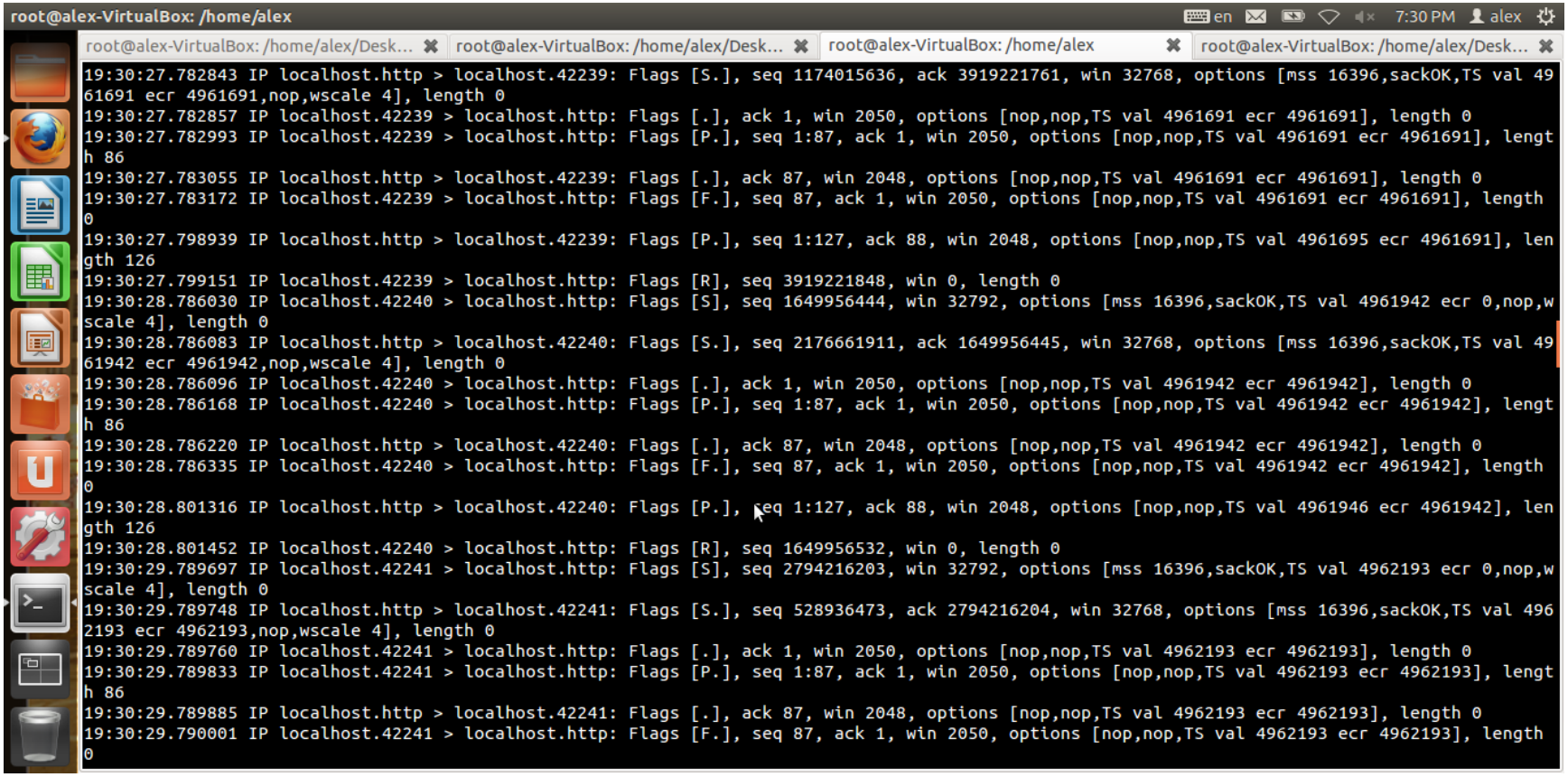
Attackers

Defenders

End Users

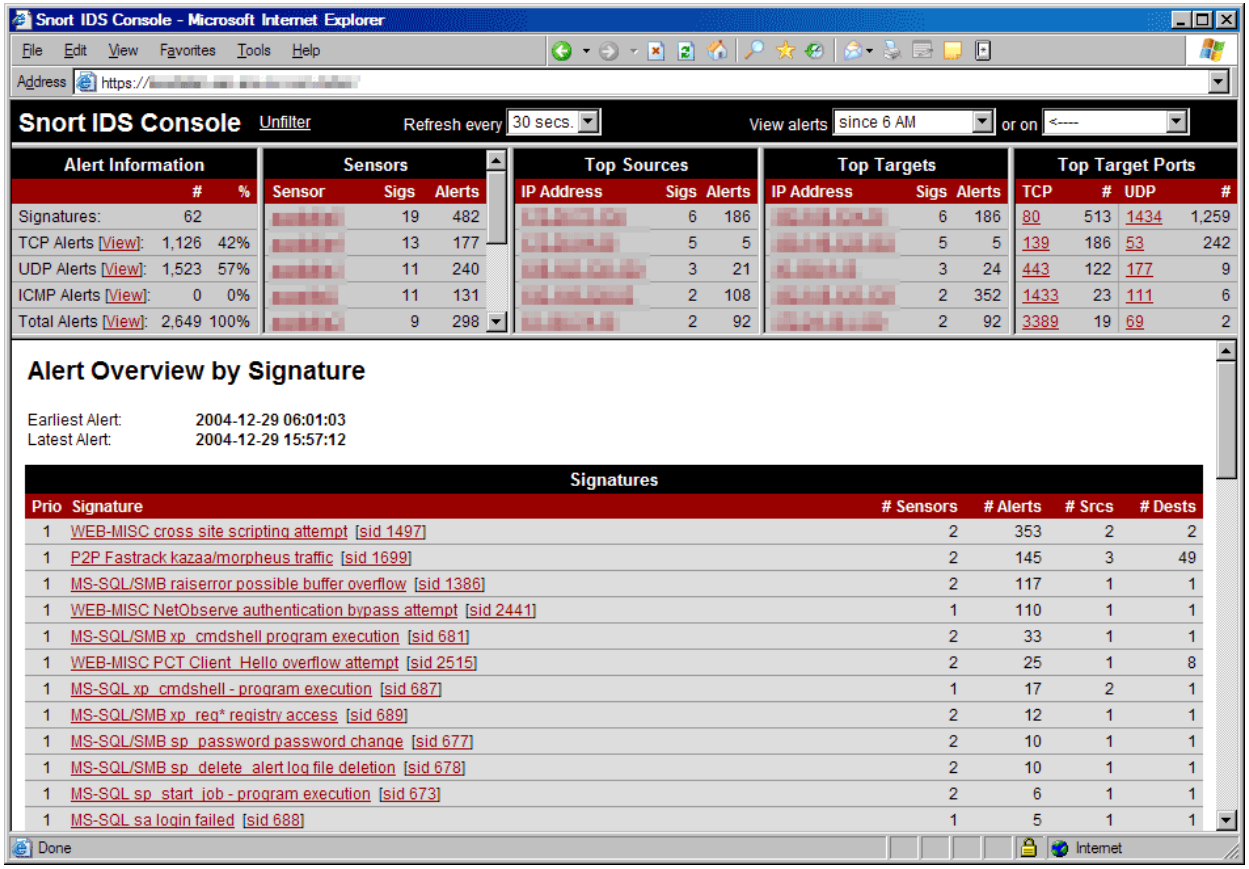
Main Challenges …
What should be part of the architecture?

Unusually large, highly diverse, and complex information



➔Attention, Memory, Mental workload, Pattern Matching

IDS (e.g. Snort): "senses" anomalies in data according to rules



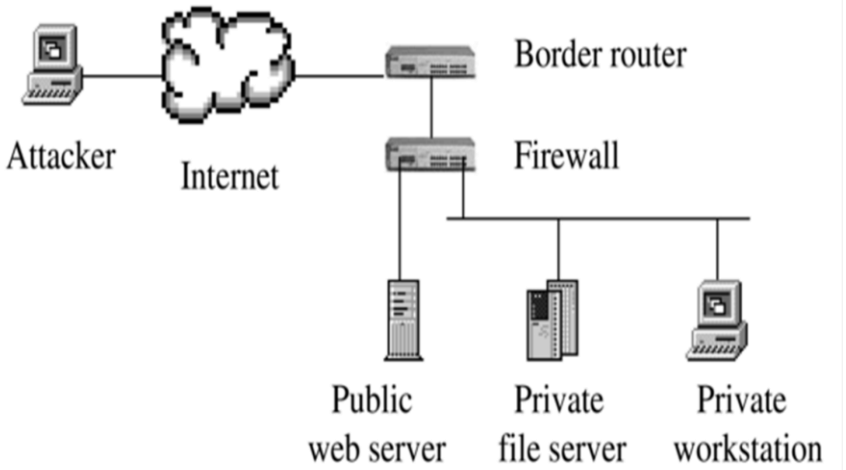
High rate of false signals

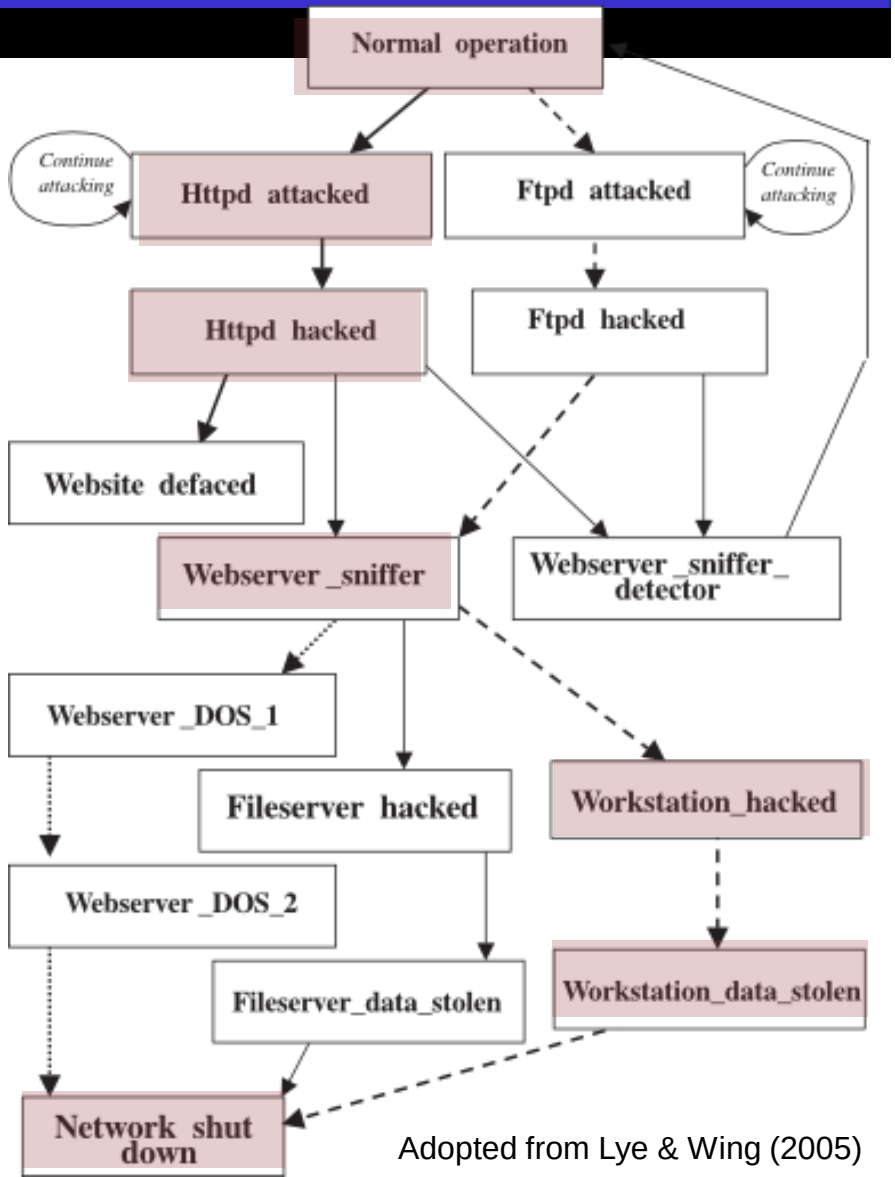
Alerts are relatively low frequency

➔Vigilance (sustained attention)

➔Signal detection: memory of rare events

Interrelations of seemingly unrelated events





➔Sequential information processing; sequence learning

Adopted from Lye & Wing (2005)

Incentives and Dynamic Risk Assessment

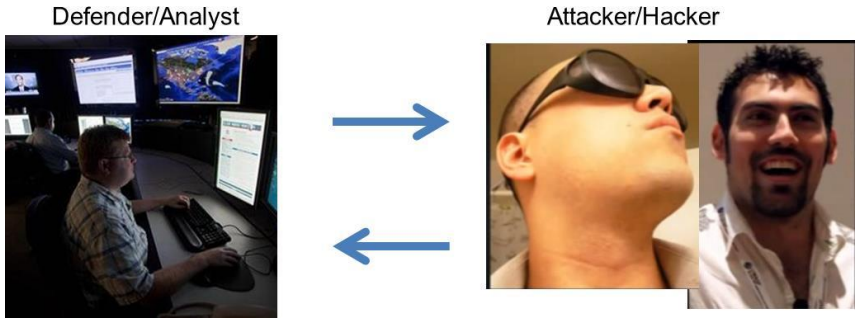
➔ Dynamic evidence accumulation

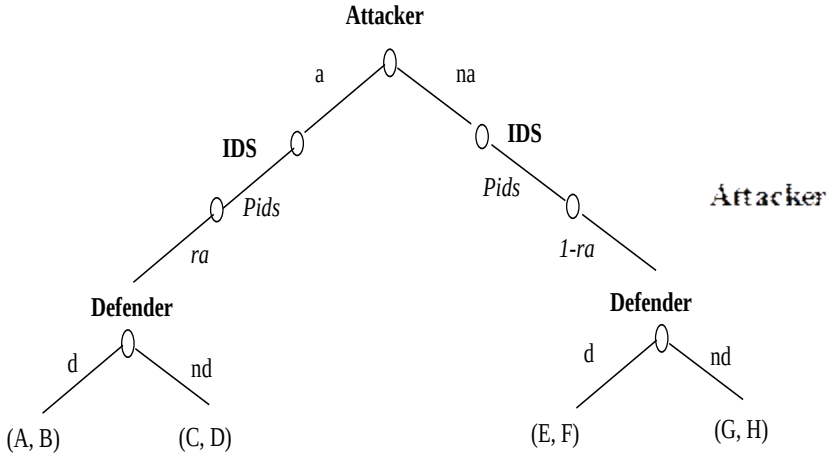
➔ Learning without/with delayed feedback

High uncertainty and rapidly changing environments

- ➔ Recognition, memory and experience-based decision-making
- ➔ Decision biases: Overconfidence, confirmation biases

Adversarial and asymmetric behavior





Attacker

Defender

	Monitor (*d*)	Don't Monitor (*nd*)
Attack (*a*)	-5, 5	10, -15
Not attack (*na*)	0, -5	0, 0

➔Deception, Incentives, Intentions, …. Etc.

➔Strategic behavior - Game Theory

Broadening and scaling up cognitive models to deal with **Social/Aggregated** Mind

Modeling detection with Instance-Based Learning Theory (Dutt, Ahn, Gonzalez, 2011, 2012)

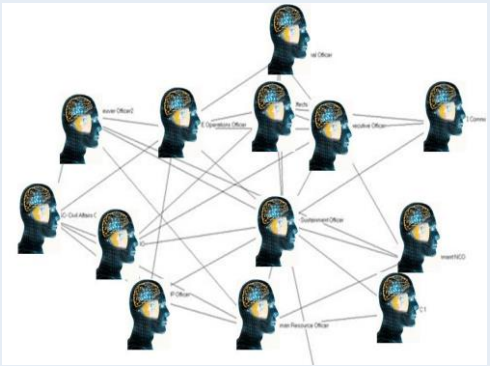
From Individual Decisions from Experience to Behavioral Game Theory: Lessons for Cyber Security (Gonzalez, 2013)

Perspectives from Cognitive Engineering on Cyber Security. (Cooke et al., 2012).

The Cyber Warfare Simulation Environment and Multi-Agent Models (Ben-Asher, Rajivan, Cooke & Gonzalez, in preparation).

Defender

Defender Attacker



Cyber War: multiple attackers Defenders

Individual (Defender). Cognitive theories, Memory and individual behavior

Interdependencies (Defender and Attacker) Behavioral Game Theory

Interdependencies and Group Dynamics (Defender and Attacker within each individual) Behavioral Network Theory; Network science (& topology) Organizational Learning; Political and Social Science

Summary

1. ACT-R has provided robust demonstrations of a unified theory of mind. But, models are mostly used to demonstrate isolated phenomena.
2. Behavior may be regarded as a characteristic of both, cognition and environment
3. To make progress (develop “a single model that acts in a variety of task domains”) we must:
   - Create a unified theory of environments
   - Analyze complex tasks beyond those involved in simple psychological paradigms
4. Cybersecurity is very complex and current theories need to be extended to explain behavior in this environment. But, psychologists are not computer scientists. We must:
   - Search for universal principles
   - Demonstrate that a model is able to accomplish multiple tasks with the same basic underlying psychological principles