Cognitive Science
& Technology

NOTE: All content is this presentation has previously been submitted for R&A and approved for unlimited release

SAND2005-7849P

Operations System Safety Philosophy Incorporating Cognitive Systems Technologies

December 6, 2005

Dr. Chris Forsythe

Distinguished Member Technical Staff

Cognitive and Exploratory Systems and Simulation

Sandia National Laboratories

Sandia National Laboratories

Our research and development program has focused on development of technologies enabling "cognitive systems"

- **Embed within machines** highly realistic and individualized **computer models of cognitive processes vital to human communication, cooperation and collaboration.**
- **Provide software that** acquires accurate models of an individual's knowledge **of a domain or task by observing their day-to-day system interactions.**
- Create systems that interact with users in a knowing cognitive manner:

  (1) know what you know, what you don't know, what you do, how you do it,

  (2) can place current events in the context of past experiences and

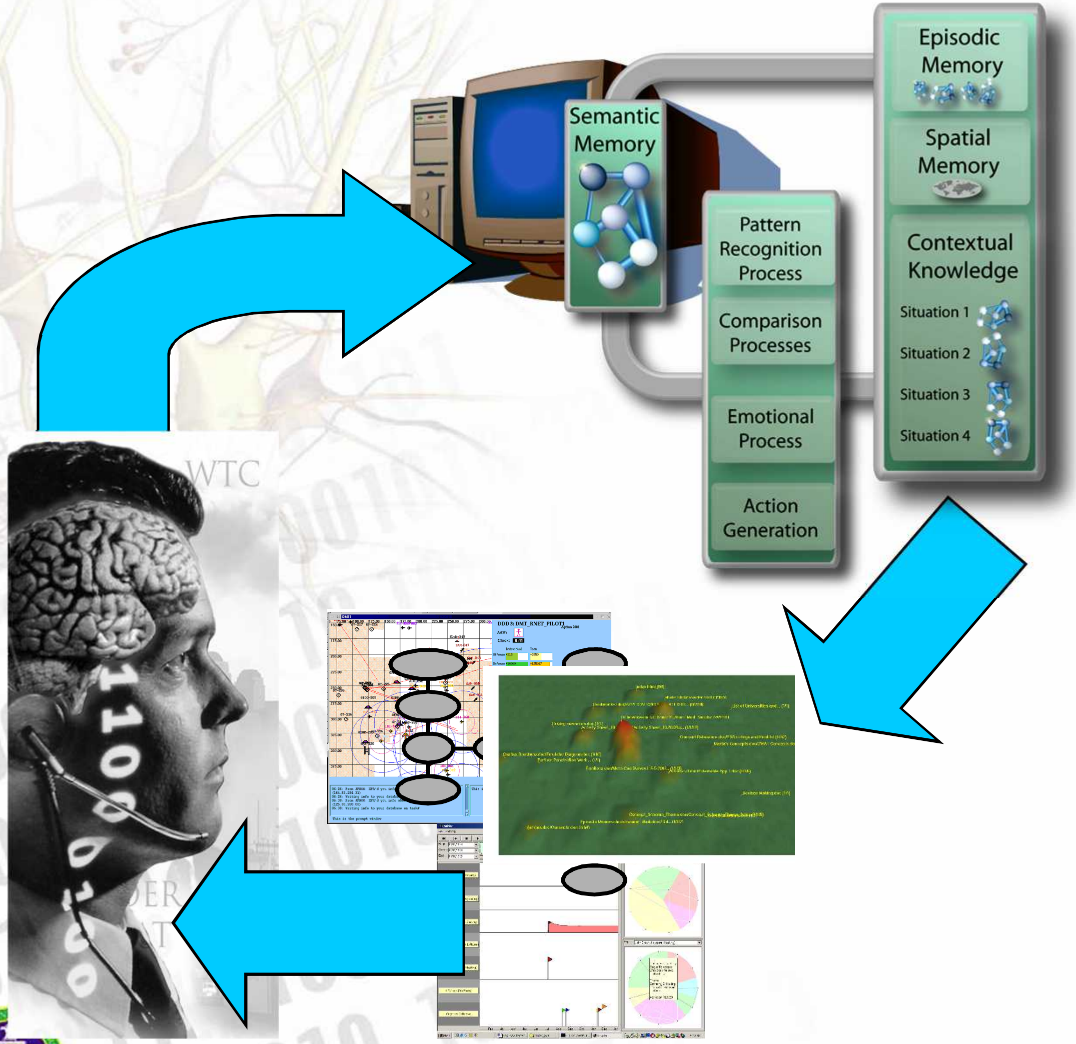
  (3) make readily accessible the knowledge and experience of diverse experts.

Transform the human-machine interaction to become more like an interaction between two cognitive entities

Cognitive systems are different from most artificial intelligence

- Intent is to model specific individual, instead of a generalized expert model, or absolute truth.
- While many AI systems have been inspired by human cognition, our systems' "human-likeness" is attained through a rigorous engineering process with no known counterpart in AI.
- Dynamic complex system that responds gracefully to anomalous events and may easily adapt to changing circumstances, as opposed to brittleness of rule-based expert systems.
- Knowledge is associative with emphasis on pattern recognition, as opposed to rule-based representations of knowledge and emphasis on logical operations.
- Machine interaction with human user based on adaptation to the unique knowledge and experience of individual with emphasis on systems that conform to the user, as opposed to one-size fits all approaches or customization based on statistical profiles.
- Ultimate goal is to augment the human, not replace the human.

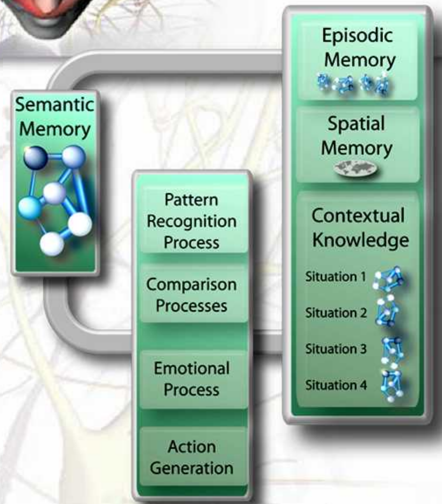
Our program has benefited from a Sandia National Laboratories Grand Challenge



The objective has been to capture the context understanding of individual humans in technologies that augment human detection and interpretation of meaningful patterns.

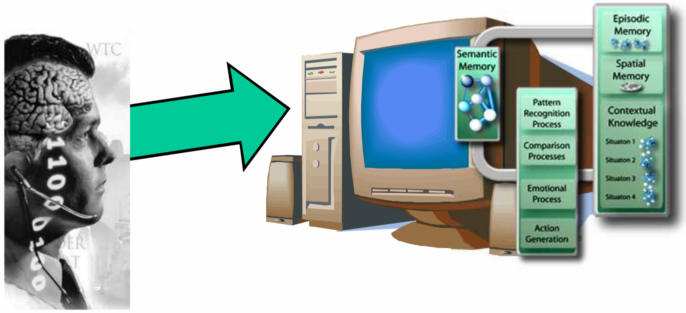
Cognitive Science & Technology

Our cognitive systems research and development has had three general technical objectives



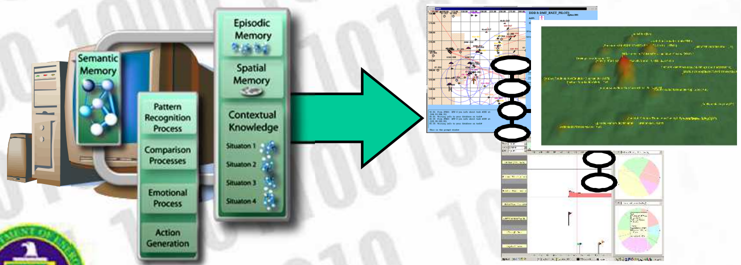
Computational model of human context understanding

(Target = 90% correspondence between model and individual).



Automated knowledge capture.

(Target = 90% correspondence between automated knowledge capture and task-based measures of knowledge)



Augmented human performance for context understanding.

(Target = 50-100% improvement relative to no augmentation)

Sandia National Laboratories

A conceptual depiction of a cognitive system to infer and adapt to a user's ongoing cognitive processes

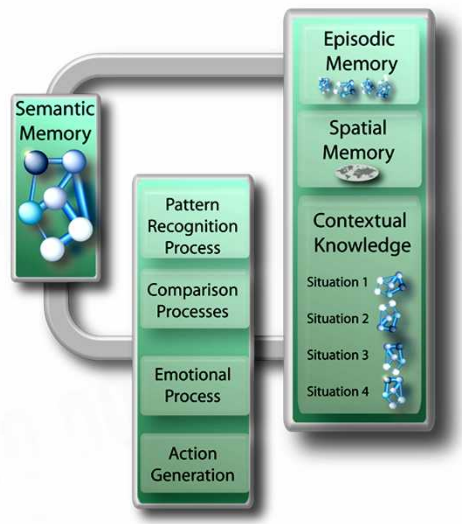
Context Manager utilizes real-time user cognitive model(s) to infer context and provide basis for system adaptation.

Data Sources
- BioSensors
- Cognitive State Assessor
- Man-Machine Interface
- Physical Platform
- Mission
- Knowledge of User

↔

Context Manager

User(s) Cognitive Model



↔

System Interface Director

Adaptive Automation Director

Cognitive Science & Technology

There are many variations on the concept for adaptive cognitive systems

Data sources
- user actuation of controls
- eye-tracking and pupillary response
- user head position
- user postural adjustments
- physiological recordings
- physical location (e.g. GPS) and associated intelligence
- system state variables
- external sensors
- communications
- user response to structured queries

Types of context
- user tasks and goals, including relative urgency and progress toward goals
- user awareness of stimuli or events
- user situation interpretation, or misinterpretation
- user emotional state
- impediments to normal cognitive functioning
- user knowledge of domain or task(s)
- user skill levels or capabilities
- actual and/or perceived physiological state
- perceived roles and responsibilities
- awareness of others

System adaptations
- adjust rate of information flow
- adjust time profiles of automated systems
- adaptive automation and/or allocation of tasks
- adjust saliency of information display elements
- system alerts
- augmented context recognition and interpretation
- user-tailored system support
- adaptive training

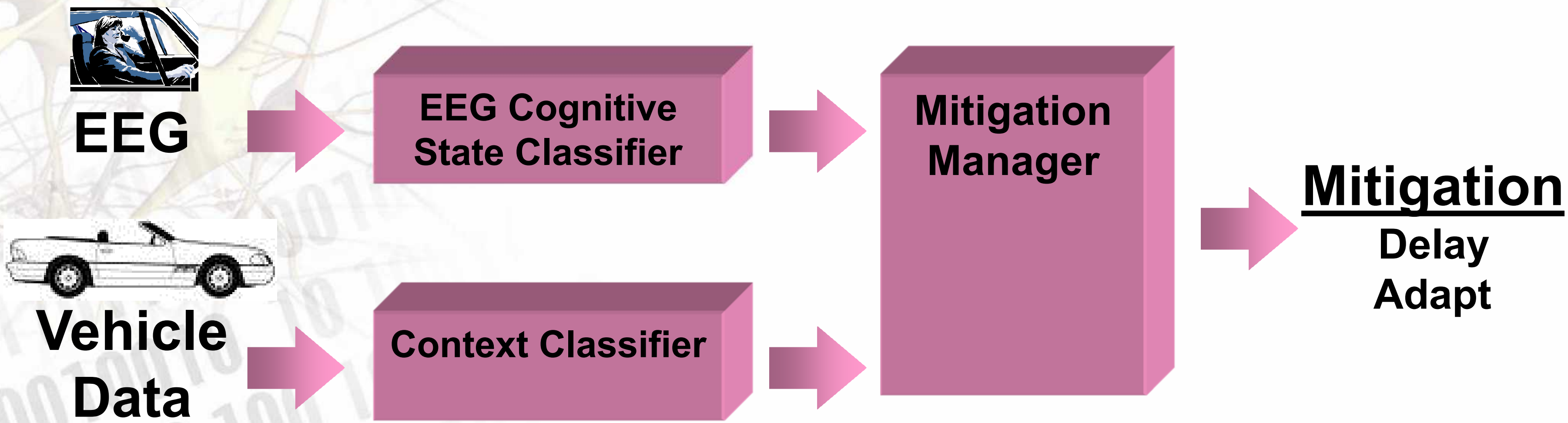
Sandia National Laboratories

It helps to think of human analogs to cognitive systems that infer and adapt to user cognitive processes

- Recognize a person is encumbered > hold door, offer to carry some of the load.
- Recognize a person is confused > alter or expand explanation, draw picture.
- Recognize a person does not recognize importance of what they are being told > shift emphasis in speech, state importance.
- Recognize a person does not see something > alert them to where to look, point or turn the person.
- Recognize a person does not remember > offer cues to prompt their recall.

Our development of an automotive cognitive systems offers a relevant example

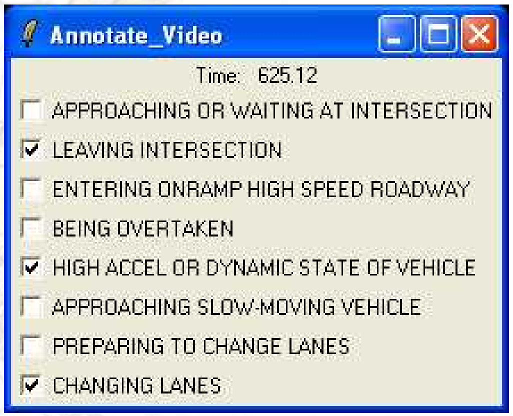
The objective is to provide real-time recognition of high-load driving contexts as a cue for the automobile to initiate measures to mitigate the high-load



Cognitive classifier operates as vigilant observer that knows when person is busy so mitigation manager can block distracting information

Our first step was to prove that real-time context recognition was feasible

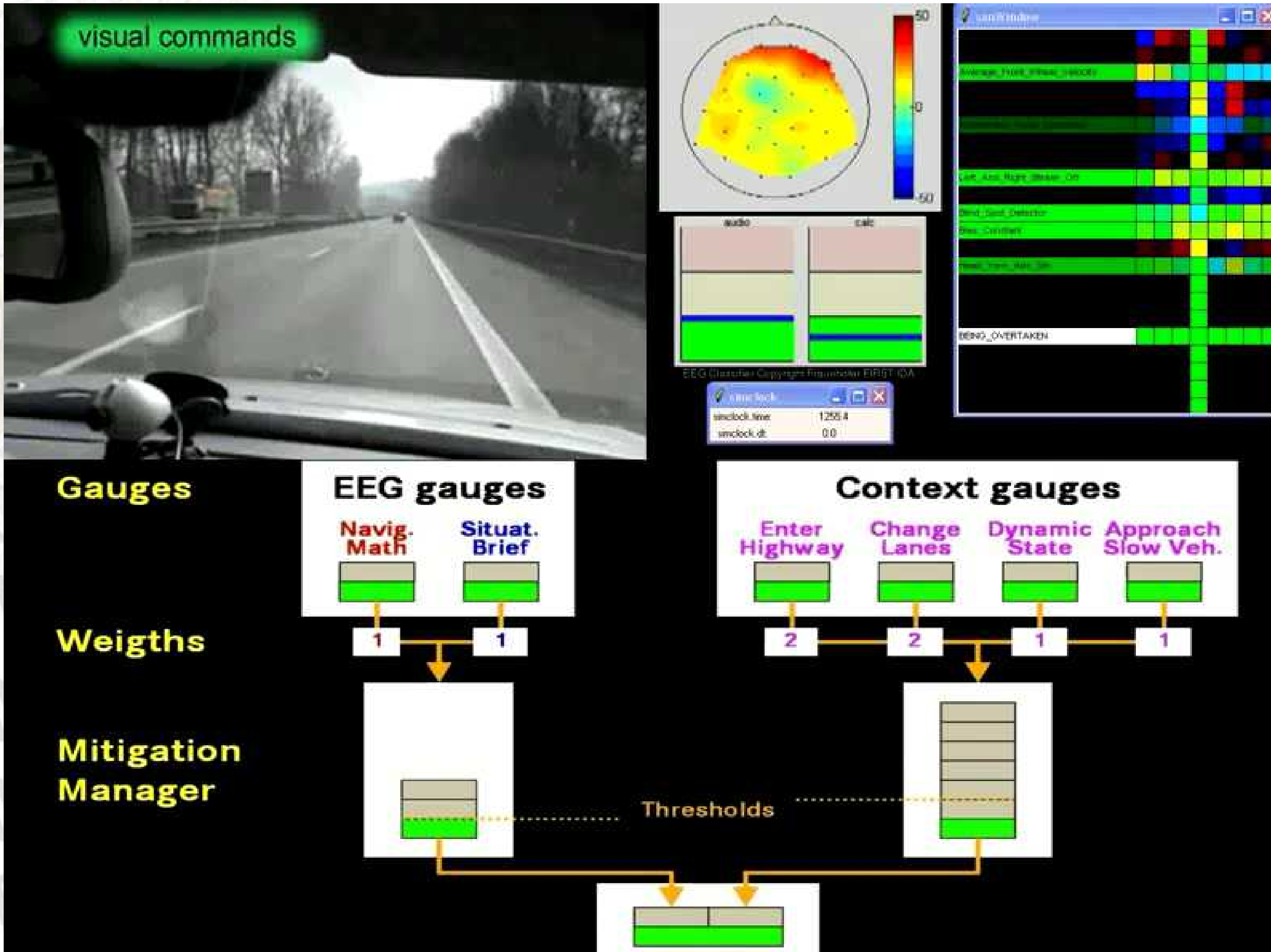
- Collected ~24 hours of real-world driving data
- Humans labeled with respect to predetermined driving contexts



- Trained context classifier and assessed relative to benchmarks – SNL cognitive model was 94% accurate vs. 84% for genetic algorithm 54% for Bayesian approach and 35% for neural net

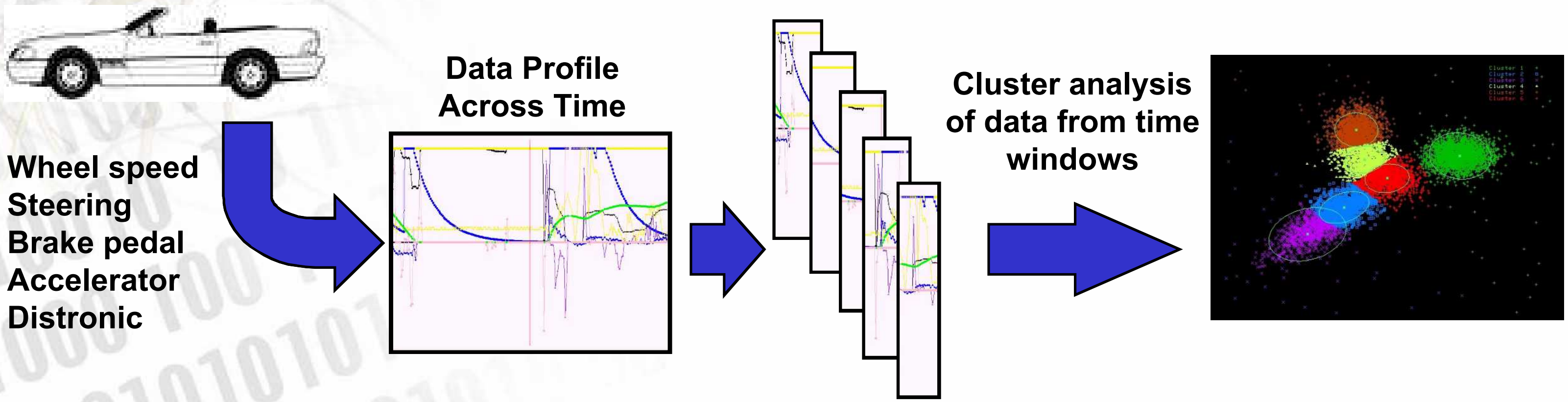
Proved real-time context classification was feasible, yet concluded that supervised learning was limited as a long-term solution

Real-time context recognition has been integrated with EEG measures of cognitive load to demonstrate improved performance



We next developed an approach to train context classifier using unsupervised learning

- Conclusions concerning supervised learning
  - Effectiveness of developers in selecting contexts inherently limited overall success
  - Impractical to adapt to individual
- Adapted unsupervised learning approaches for automated context extraction - better to let data tell *us* what contexts are important



Proved that we could identify regularities in data that correspond to meaningful driving contexts

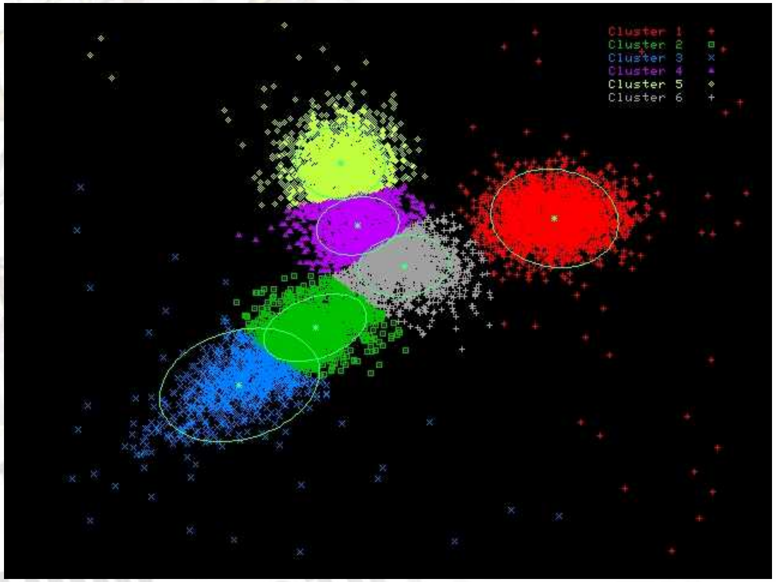
Cognitive Science
& Technology

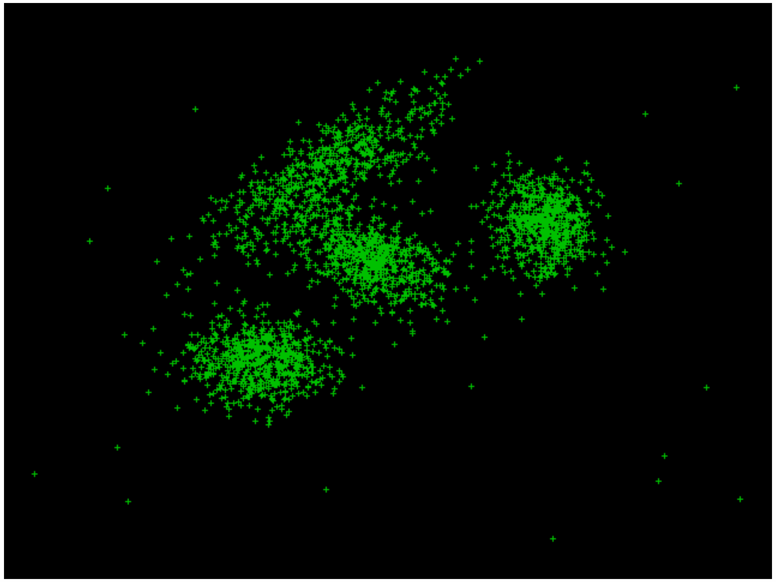
The model quickly hones in on a stable set of clusters that provide contexts for context classifier

National
Laboratories

We next extended automated context learning for adaptation of classifier to individual operator

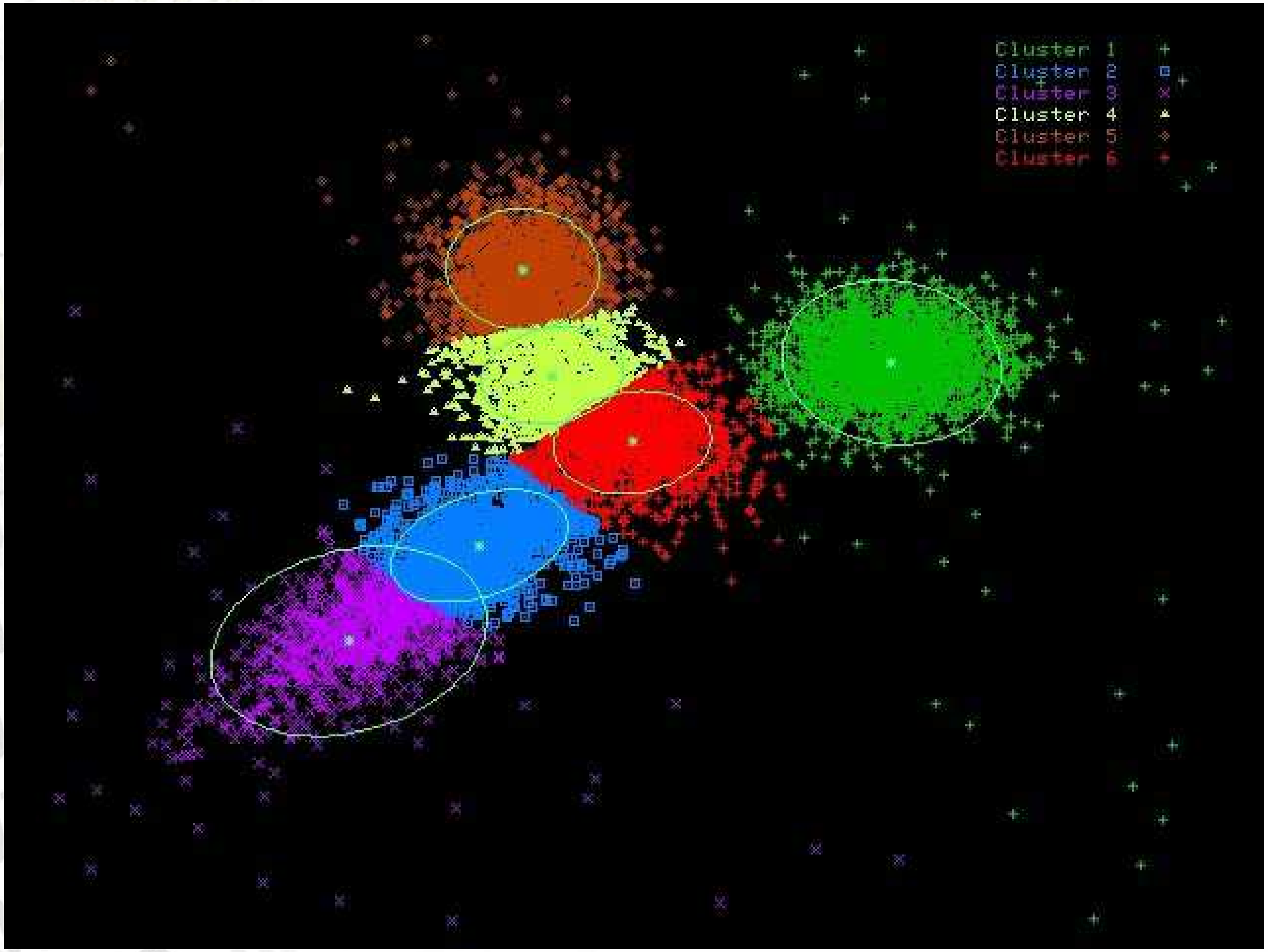
- We develop a generalized model based on data obtained from a group





- But now want to *adapt* the model to a new user

The model quickly adapts to idiosyncrasies of individual providing an individualized context classifier



Cognitive Science & Technology

We next developed model to predict cognitive load on basis of ongoing driving contexts

- **Obtained human ratings concerning the level of cognitive load associated with a broad range of driving situations**
- **Used supervised learning to train model to predict cognitive load based on corresponding driving contexts**
- **Assessed performance by comparing model predictions to human estimates**
  - Mean Difficulty Error (L1): 4.99, on 100 point scale
  - Mitigation Agreement: 93.17%

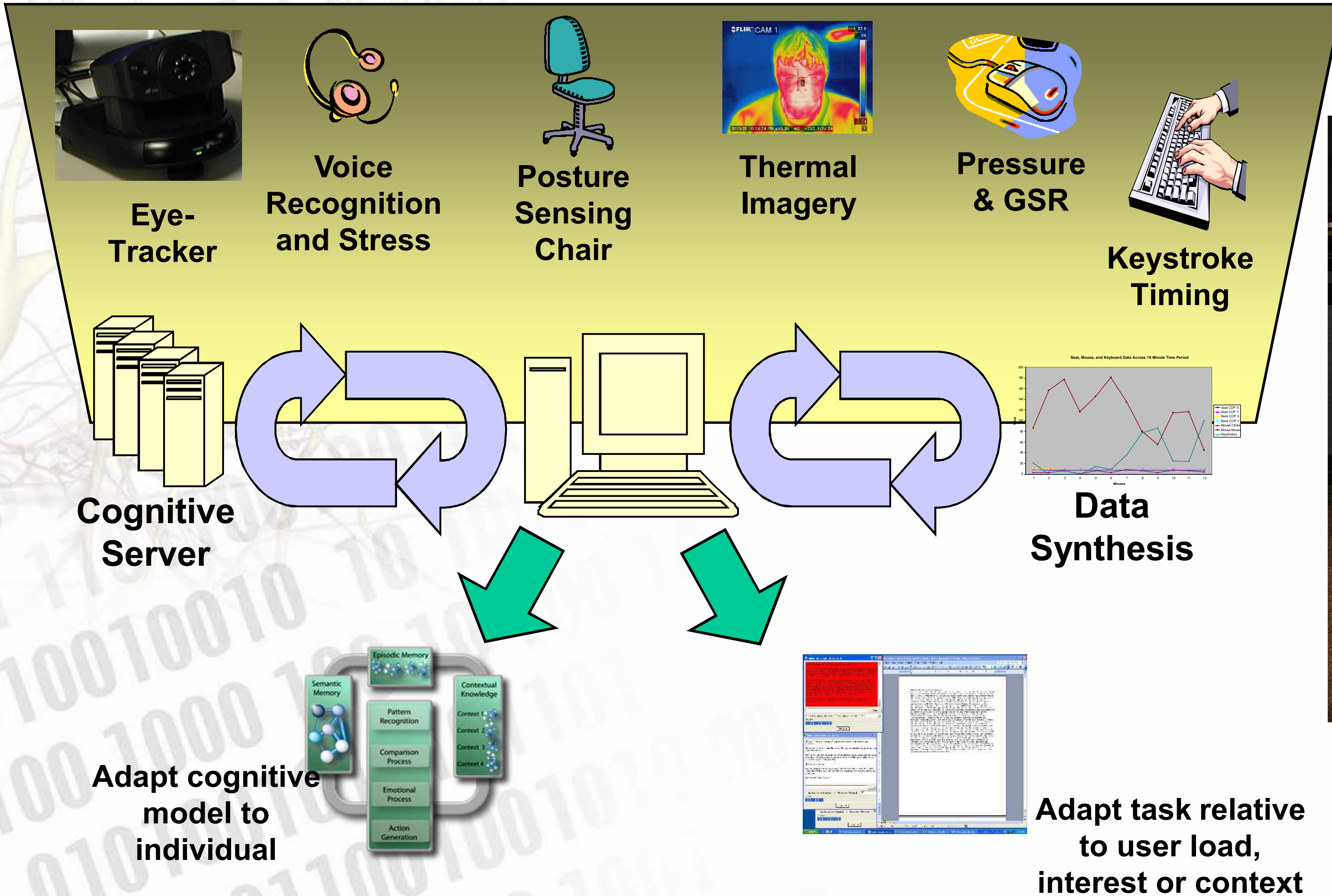
Proved that real-time context classification was viable basis for cueing system adaptations to operator cognitive load

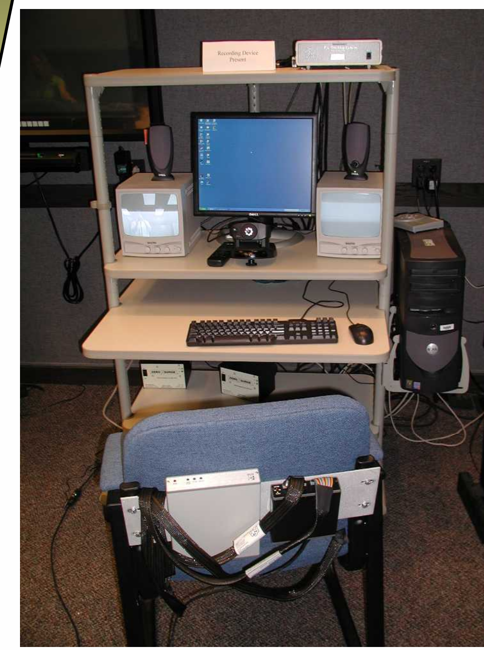
Sandia National Laboratories

Context classifier provides real-time estimate of the cognitive load associated with driving conditions

Cognitive Science & Technology

Non-intrusive sensors provide basis for real-time detection of cognitive state and system adaptation



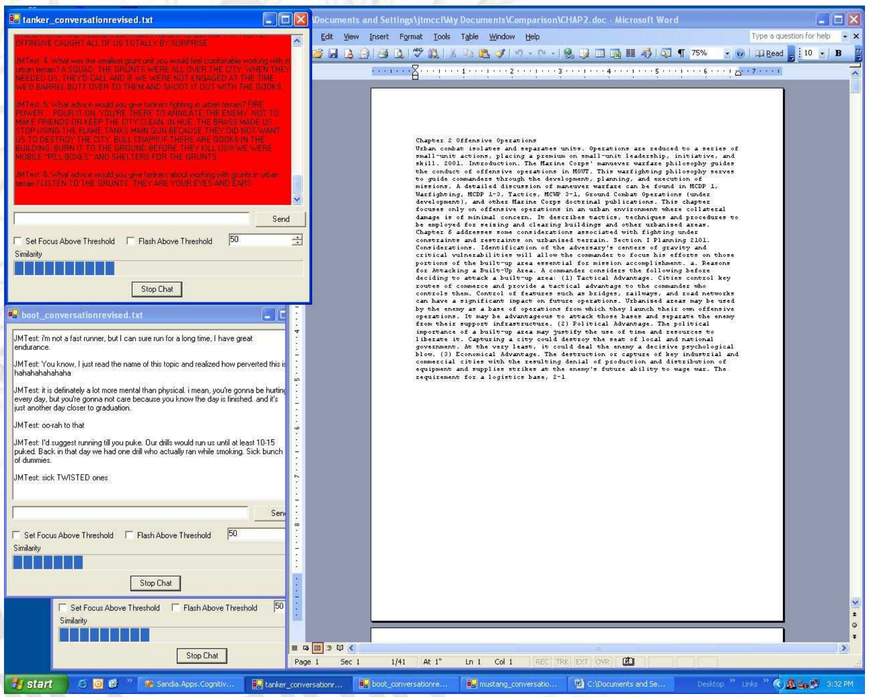


Sandia National Laboratories

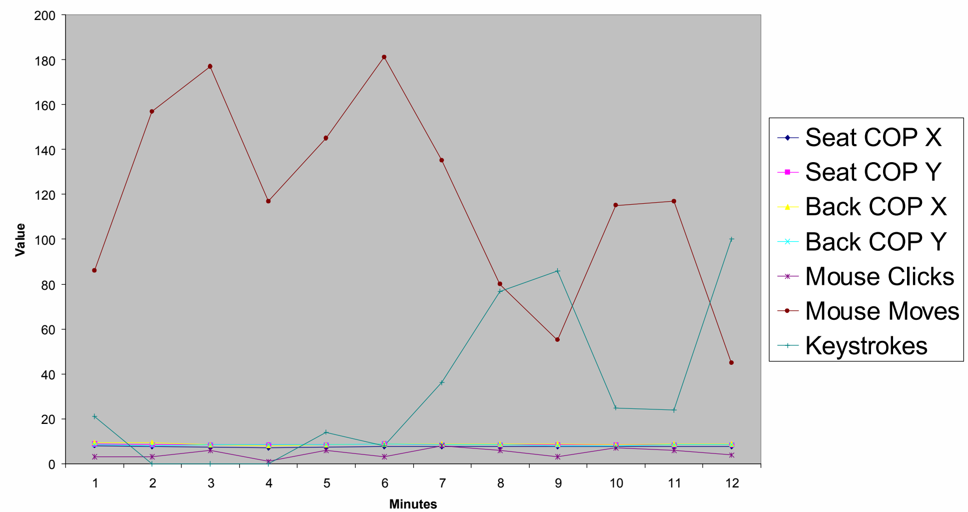
Cognitive Science & Technology

Individual information bandwidth is expanded through attentive systems utilizing cognitive models

Individual's cognitive model monitors multiple data sources to highlight transactions of interest or relevance to the current task.



Operator cognitive model monitors multiple Instant Messaging windows and alerts operator when it detects an item of interest or relevance to the primary task.



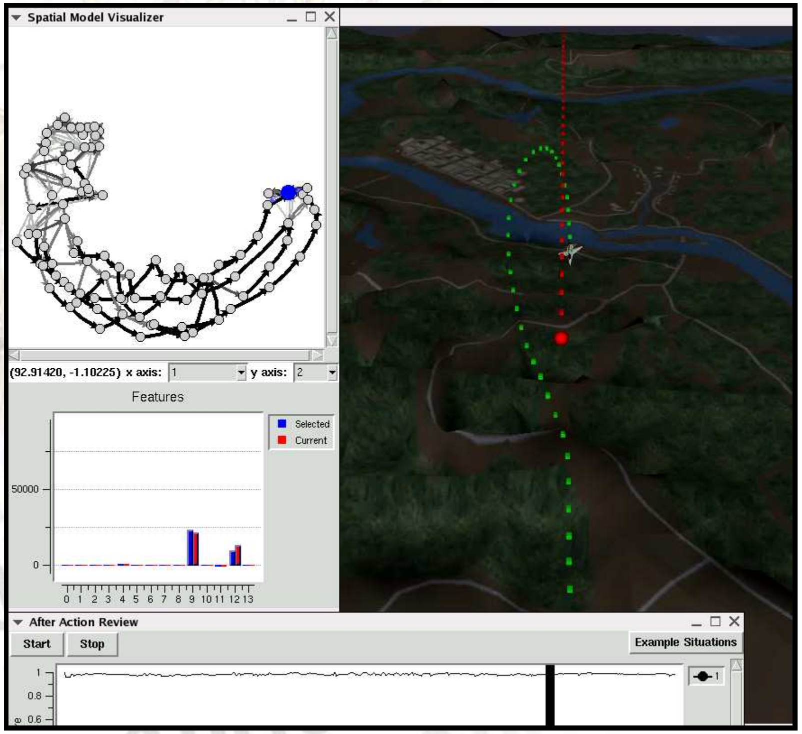
Considering data sources in combination reveals patterns that occur across time

Sandia National Laboratories

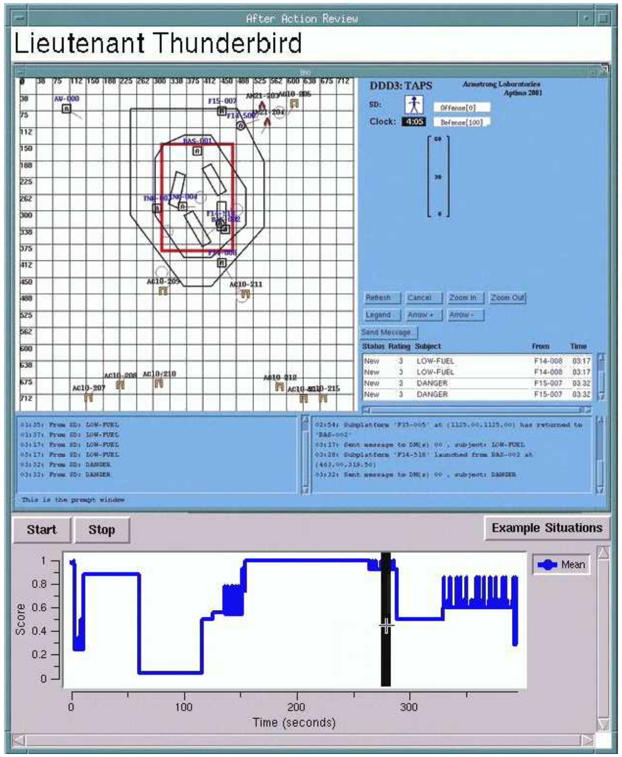
Cognitive Science & Technology

Discrepancies in task performance may be detected in real-time by comparing model to observations

As operator performs tasks, their performance is compared to cognitive models of experts to identify discrepancies



Upper left figure illustrates performance of expert in aircraft stern convergence task and lower graph compares expert to trainee performance.



Lower graph illustrates performance of operator relative to an air traffic control expert over the course of training session.

Sandia National Laboratories

Our philosophy for operations system safety presents opportunities to benefit from cognitive systems technologies

Within an engineered system

(1) At some level, there is *no escaping human involvement*

(2) *Most complex* system component

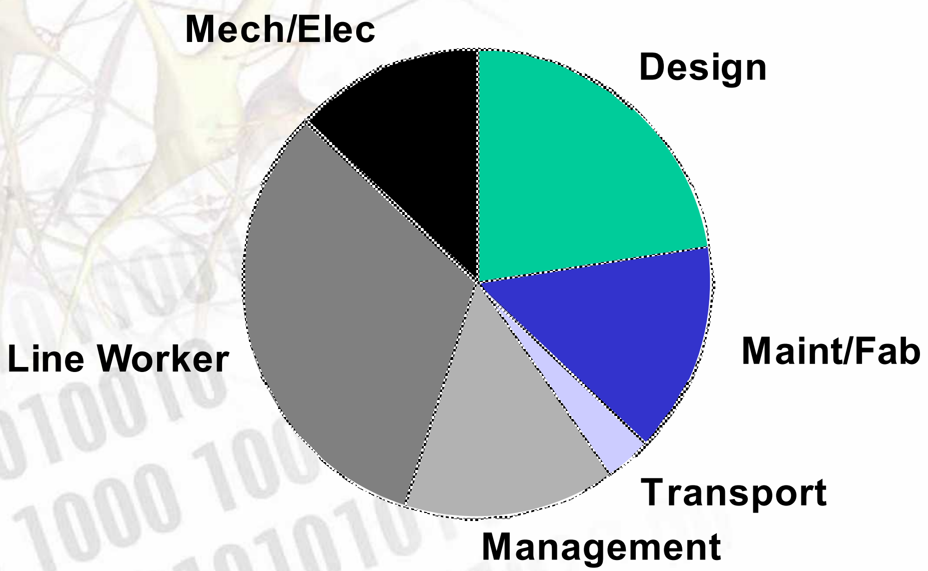
(3) *Least understood* system component

(4) System component *most vulnerable* to failure

(5) Humans present a *remarkably diverse* set of failure modes

Where do our experiences tell us human errors occur when we consider an operation from design to operation?

Disproportionate Attention is Focused on Line Workers Neglecting Engineering, Maintenance and Management



Of those accidents involving Line Worker error, over half could be traced to an inadequate user interface design

Cognitive Science & Technology

Errors were rarely simplistic omissions and commissions, but involved complex cognitive processes

Failed To Anticipate Hazard

Events

Adequacy of Controls

System Interaction

Significance

Failed To Anticipate Human Error

Failure Discriminate

Fail Perceive Risk

Incomplete Mental Model

Misinterpret Intent

Mistaken Mental Model6

Physical Constraints

Placekeeping

Variance in Performance

General Cognitive Errors

Authorization

Dynamic Properties

Functional Requirements

Material Properties

Response Latency

Failure Discrimination

Incomplete Mental Model

Mistaken Mental Model

Placekeeping

Failed To Anticipate Requirements

Communicate Hazard

Environment

Functions

Maintenance

Sandia National Laboratories

Cognitive Science & Technology

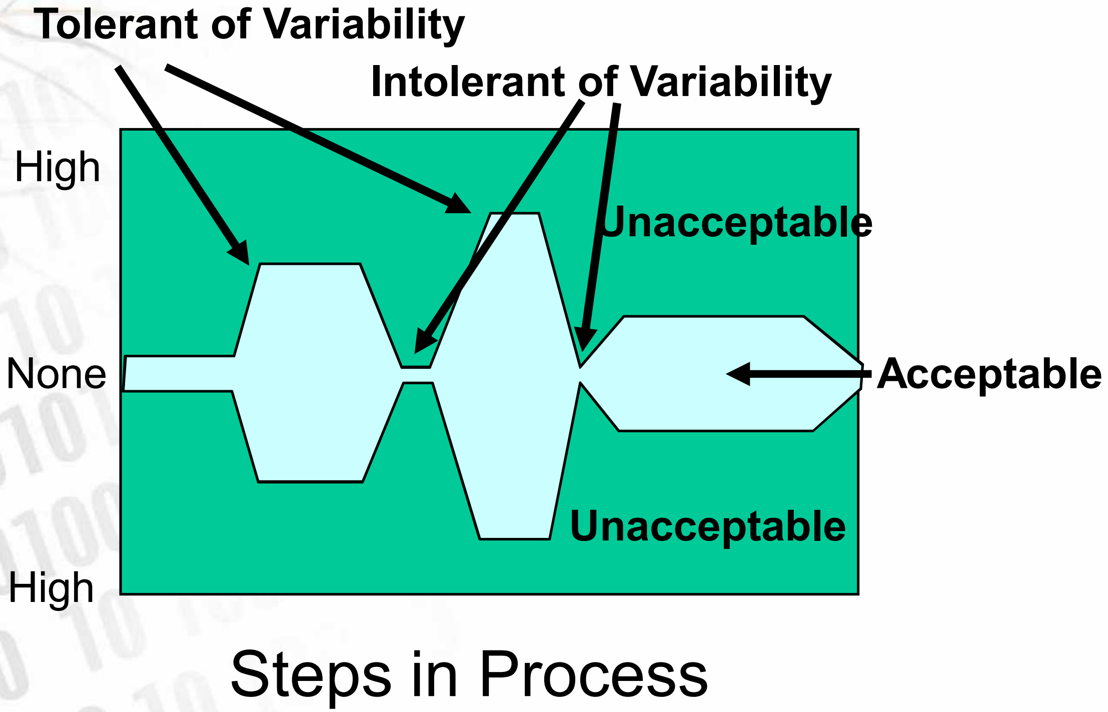
There are several common misconceptions about human performance and human error

1. **Human as Machine** - mechanistic perspectives allow many failure modes to be easily deduced, but lead to erroneous simplification of human elements
2. **Random Behavior** - a willingness to attribute unusual, unexpected behavior to random unexplainable processes leads to conclusions that human behavior cannot be predicted, nor controlled
3. **Isolation of Individual Human Agents** - consideration of only formalized human interactions leads to critical informal interactions being overlooked
4. **Predictability of Human Agents** - variability in human behavior is assumed to be uncommon, attributable to random events, leading to a sense that people are all/always the same
5. **Just So World** - tendency to assume the system will operate as designed (i.e., people will read and follow instructions, training will accomplish the desired outcome, personnel selection will weed out the "bad apples")
6. **Erroneous Attribution of Mental Models** - designers and managers often overestimate the degree to which others share their mental models leading to expectations that people will behave rationally per the assumed mental model

Sandia National Laboratories

Humans introduce variability to engineered systems designed and intended to operate in a specified manner

Humans are organic systems (i.e., living organisms) and once they are inserted into a system that otherwise, could be understood using properties from the physical sciences, that system begins to exhibit properties of an organic (i.e., biological) system.



Cognitive systems technologies represent a controls that may be incorporated into a system

Ways in which human variability may be controlled

1. **Increased Tolerance for Variability – derivations may occur, but deviations are either detected and corrected, and/or due to the design of the system, deviations have no consequence**

2. **Lessen Variability - mechanisms are put in place to minimize the potential for deviations**

3. **Isolate Variability – mechanisms prevent deviations in one area of the system from impacting other areas**

Discussion