Decision Under Uncertainty and Trust in Automation: Adaptive Systems Performance

Margery Doyle, Anil Raj, Joshua Cameron

2010

Overview

- Decision under uncertainty, time constrains, and increasing complexity
  - Implications
  - Uncertainty
    - Drivers, types, and some causes
    - Decision under uncertainty
- Trust in Automation
  - Reliance, reliability, and trust
  - Building trust in human-synthetic systems
    - Modeling the interaction between WL, SA-SU, DE
      - For the agents SA-SU of human performance
      - In-turn human trust of agent behavior
    - Architectures that include trust factors
    - Finite State Machine Induction
- Summary

Implications (Mabry, 2004).

- ***Today**, often the **human understanding*** of **the state of the system** must be achieved under greater time constraints.

- Combined with increased complexity in networks requires improved technologies for understanding and controlling complex system behaviors.

- Improvements are needed now at the interface between ***human understanding** of system state and **machine understanding** of system state*

- ***Inclusive of the human cognitive state*** *(Raj, Doyle, Cameron, 2010)*

Drivers of Uncertainty

- Lack of good probabilistic knowledge
- Lack of information
- Lack of situation awareness/understanding
- Differences in opinion
- Misdiagnosis (diagnostic uncertainty)
- Recollection of hypotheses (cause for propagation)
- Acceptance/rejection of hypothesis
- Selection of goal
- Selection of means to achieve a goal
- Execution of Means

Additional Causes of Uncertainty (Wang and Roush; 2000)

- **Stochastic (inherent)**
  - Due to variability in system design or the environment
    - Creates random outcomes
- **Statistical**
  - Incompleteness of data; small sample size
- **Modeling**
  - Resulting from the simplification of nature
    - Large number of assumptions are made during modeling
    - Fidelity
      - Often unequal between agent and human
      - Not matched to needed level of fidelity

Types of Uncertainty Cont (Marvis et al., 1998).

- **Design uncertainty**
  - **Two distinct classes of design parameters emerge:**
    - **Control parameters**
      - Items that the designer has direct control over
    - **Noise parameters**
      - Effect the design; yet beyond the control of the designer
- **Situations and systems are non-deterministic**
    - Leading to an inability to analytically predict and engineer the outcome of an event, or the exact value of a parameter
    - Inability to create deterministic system states
- **Operational uncertainty**
  - Arises as a result of what are often called noise parameters that affect the performance of a system.

"There are many distinctions between different types of uncertainty and ways of looking at uncertainty. The most important result of including uncertainties in a (risk) calculation, like the result of making the (risk) calculation itself, is not the number, but the insight that the inclusion gives to the assessor." (Marvis, 1998)

Decision Under Uncertainty

- A positive correlation exists between sensitivity to risk and sensitivity too uncertainty.
  - The higher the [perceived] risk, the more uncertain operators become (Fox and See, 2003).
    - Monitor perceived risk (Elise Payzan Le Nestour and Peter Bossaerts)
- Irreducible uncertainty (Risk)
- Estimation Uncertainty
- Unexpected Uncertainty

Reasoning Under Uncertainty

- Forces assumptions about the nature, intentions and methods of elements in the environment (Chickering & Heckerman, 1996).
- If all elements behaved rationally and predictably
  - One could regain SA
  - Infer likely system behavior Performance, and future state changes
- Maximizing one's gain and/or minimizing losses depending on the conditions.
- Experience and historical knowledge can substitute for the missing information
  - Recognition of patterns
  - Trends
  - Analogical reasoning
  - Case-based inference
  - Evidential deductions (Cooper, 1995; Dagum & Chavez, 1993).

Decision Under Uncertainty cont.

- Willingness to act under uncertainty
  - Governed by perceived likelihood of favorable outcome
  - Attractiveness of potential consequences
  - Dependent upon the degree of uncertainty concerning probabilistic information
- Attractiveness of a prospective outcome generally decreases as ambiguity or vagueness increase (Ellsberg, 1961;2001).

Why is Trust an Important Area of Inquiry?

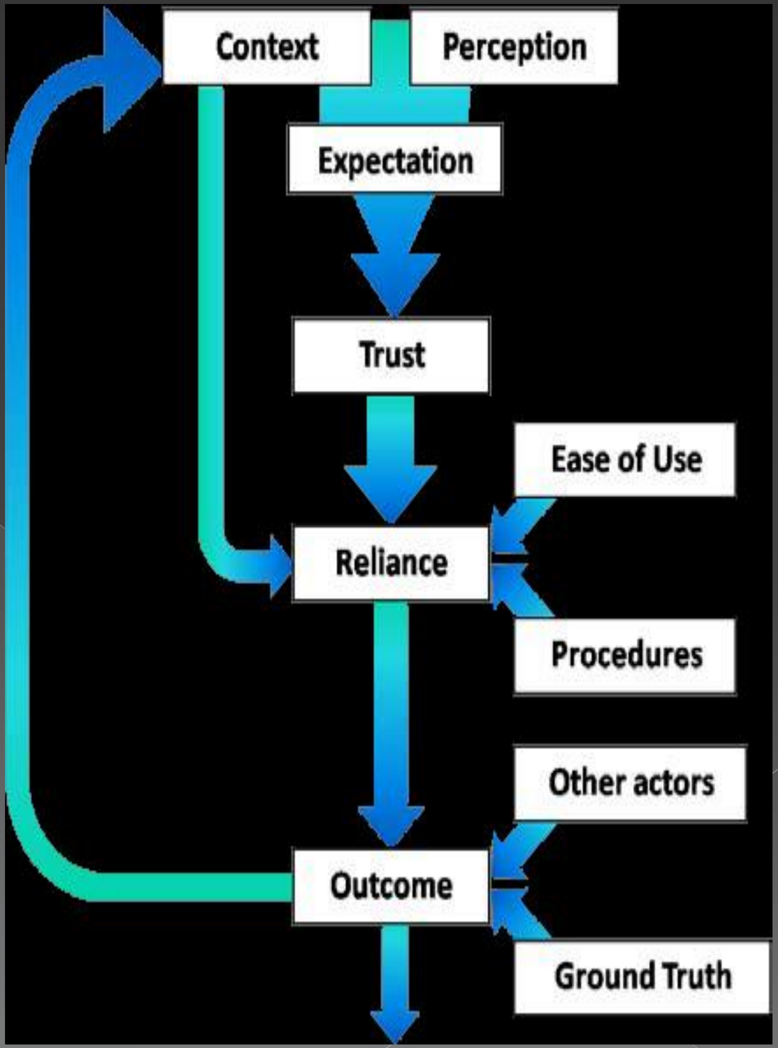
- Appropriate trust in and reliance upon automation is critical for safe and efficient operation (Gao and Lee, 2006).
  - A lack of trust in automation increases workload and decreases situation awareness because operators more closely supervise the system rather than the situation (Cummings & Mitchell, 2008).
  - Operators willingness so rely on automation impacts mission effectiveness.
- Obtaining optimal performance from an autonomous robot system requires good teamwork between the operator and the robot. Trust is an essential part of teamwork.

(Desai, Drury, and Yanco, 2008).

Automation and Trust

- Automation is "the technology that actively selects data, transforms information, makes decisions, or controls processes." (Lee and See, 2004)

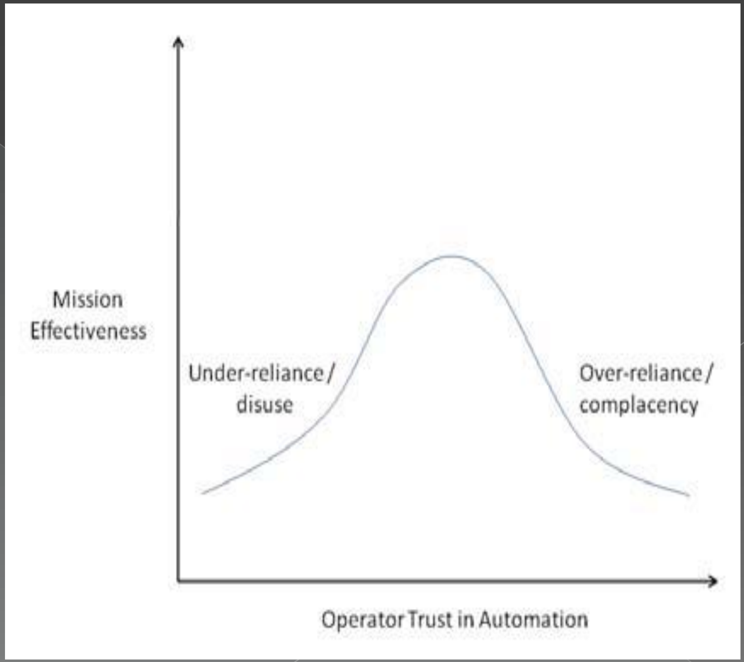
- Trust is "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability."



Trust Relationships (Cring and Lenfestey, 2009)

Reliance, Reliability, Understanding, and Trust

- Trust ≈ reliance when the operator has no choice
- Trust < capabilities an inappropriate reliance develops (Lee and See, 2004).
- Low levels of trust: Disuse
  - User distrust may lead to a 'fight' for control (Bruemmer, 2004)
    - Errors of commission.
- High levels of trust create over reliance
    - Errors of omission



Gaining and Maintaining Trust

- Trust
  - Can be difficult to achieve and maintain
  - Easy to lose
  - Difficult to recover
  - Trust is easier to build in human-human interactions than in human-automation/ synthetic agent interactions.
    - Why?
      - Humans when interacting with humans build trust through shared mental models of the world, a common knowledge base, similar goals, common motivating factors, time vested, and reliance.

Building Trust Cont.

- Create dependable and predictable systems.
  - Increase reliability and fault-tolerance (Few, 2004).
    - However, understanding robot actions and intentions may be more important than robot reliability or performance (Bruemmer, 2004).
- Synthetic systems need to provide status
  - Explain its own behavior
- Robot-Human communication etiquette improves trust
  - System should ask for help when needed (Yanco, 2004).
- Structured task environments increases trust, cooperation, and performance.

Agent Recognition of Patterns in Behavior

- "The ability to recognize patterns of operator behavior that could lead to poor outcomes is critical to monitoring the overall performance of the human-unmanned system team."
  - "Recognizing the onset of abnormal behaviors [non-optimal cognitive states], …allows for detection and prediction of the occurrence of potential critical events." (Boussemart and Cummings , 2009)

Building in Trust

- "By addressing operator trust explicitly during architecture development, system designers can incorporate more effective automation." (Cring and Lenfestey, 2009)

- To be effective, automation must be well-designed, reliable, and tailored to complement the capabilities of the human operator in varying supervisory roles (Cummings, Bruni, Mercier, & Mitchell, 2007; Cummings & Mitchell, 2008).

Model of Trust Relationships (Cring and Lenfestey, 2009)

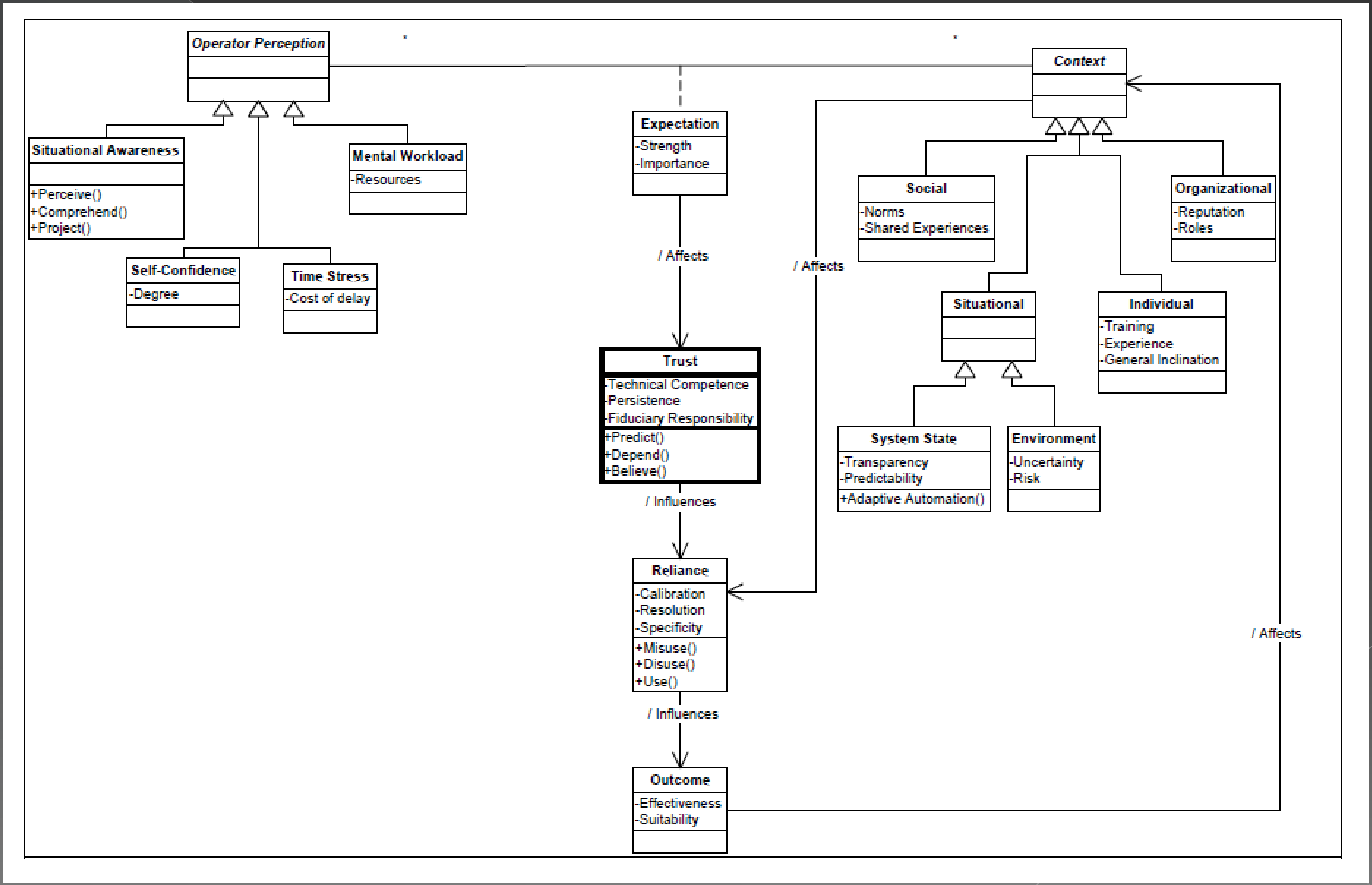
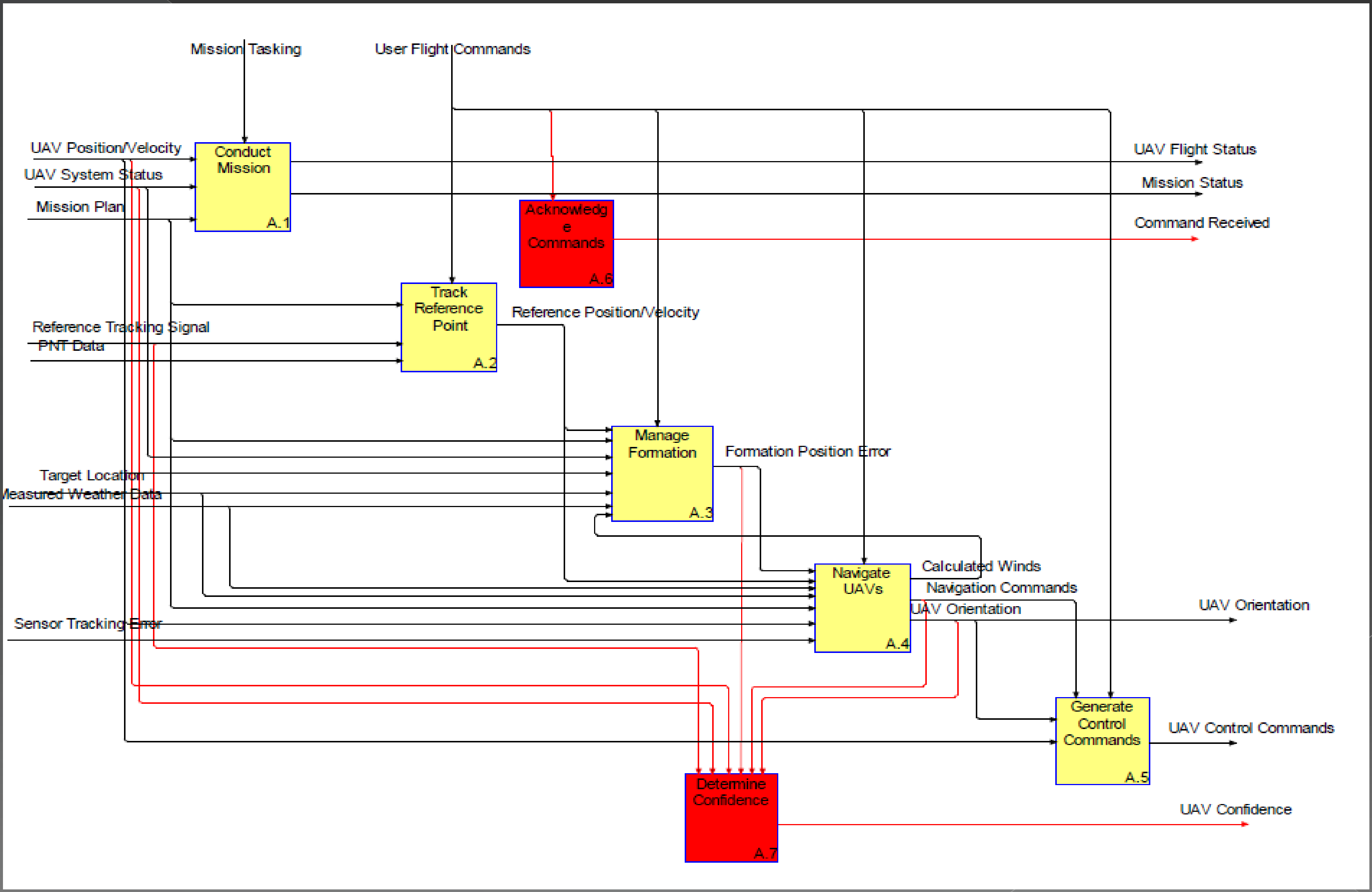
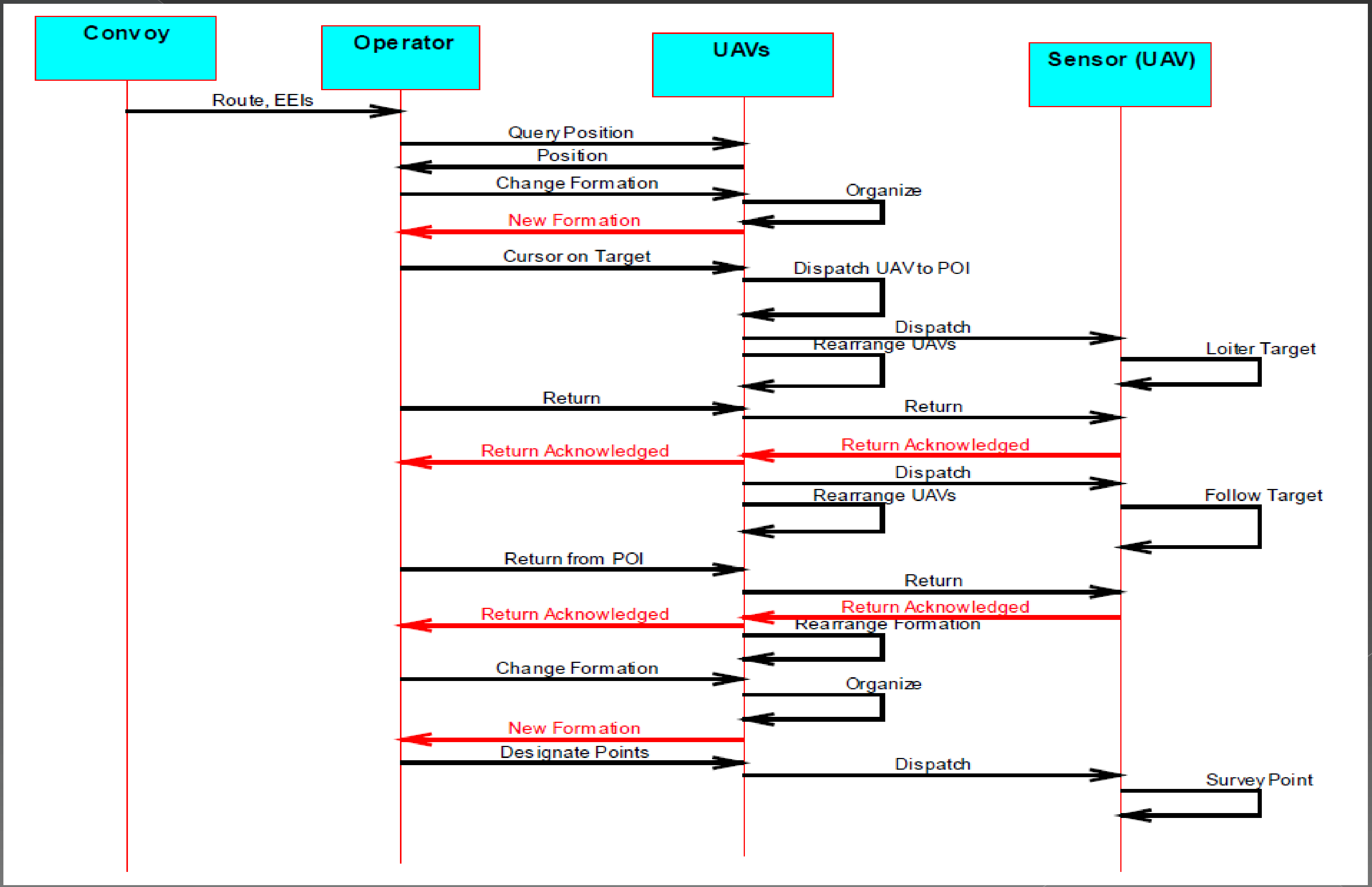


Diagram with Trust Factors (Cring and Lenfestey, 2009)



Operational Diagram: Trust Factors (Cring and Lenfestey, 2009)



Agent Recognition and Adaptive Responses to Human Performance

- Through the use of psychophysiological measures the agent is provided a window into the human operator's mental state
  - Level of cognitive workload
  - Level of attention
  - Level of frustration
  - Perceived Risk/Uncertainty/Trust*
- Agents can use this information to adapt itself to the needs of the user (i.e., adaptive automation/robotics).
- Robots can scale their own level of autonomy to support different levels of user trust (Swinson, 2004).

Modeling WL, SA-SU, DE interaction for use in Adaptive Automation

- Because situation awareness situation understanding and optimal levels of workload supports effective decision making and can inform synthetic agents about human performance
  - the upper and lower limits where the correlation between WL and SA-SU transitions from positive to negative and begins to negatively impact situation awareness, is sought.

The conceptual framework

(Doyle, 2008; Raj, Doyle, Cameron, 2010).

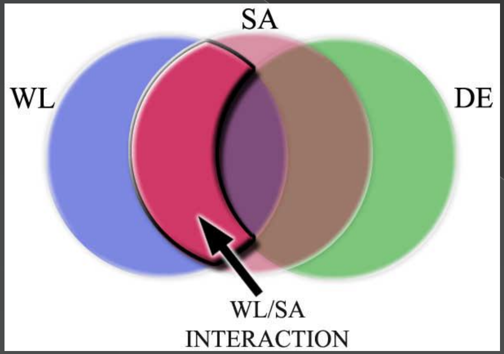


Figure 1

Figure 1: WL and WL associated with gaining SA-SU compete for limited mental capacity and interact.

Figure 2: illustrates the concept that DE is usually dependent on SA-SU and also competes for limited resources.

Figure 3: DE competes with SA-SU and task related WL for cognitive capacity.

Figure 4: SA-SU, and DE interact with WL and create an optimal cognitive state that utilizes workload appropriately, supports sufficient SA-SU, which in turn supports DE.

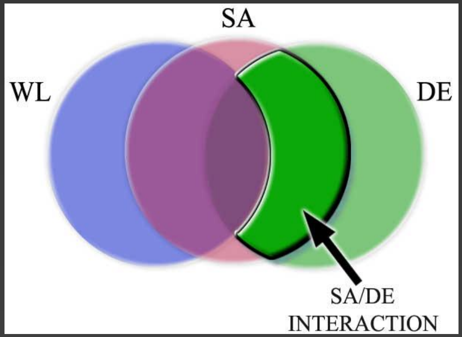


Figure 2

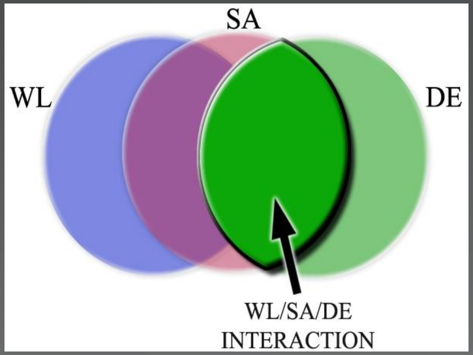


Figure 3

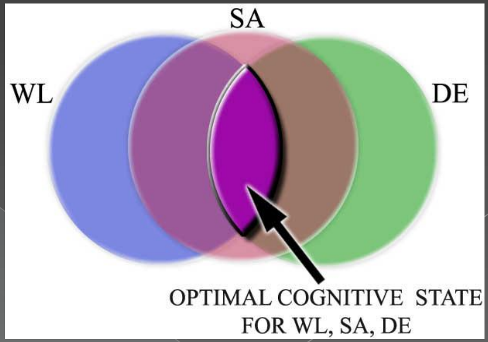
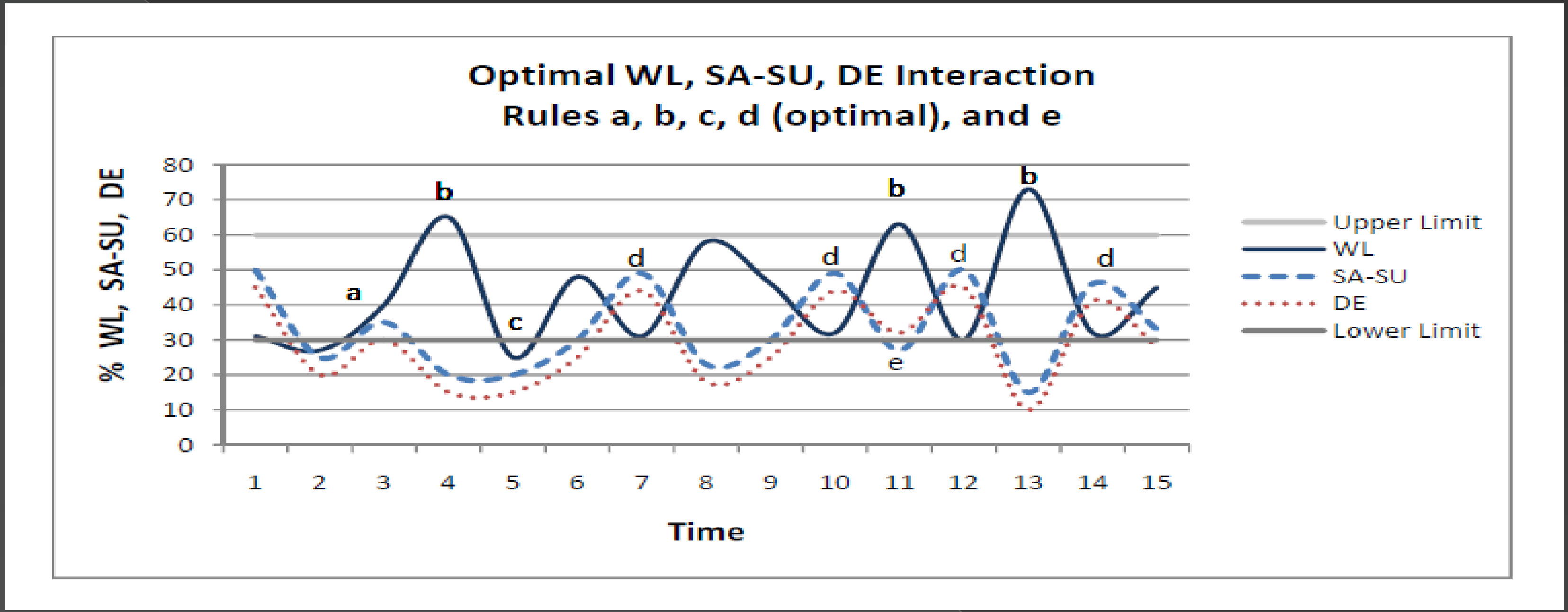


Figure 4

Hypothetical Interactions



Demonstrates that when WL, SA-SU, and DE are within the upper and lower limits and workload decreases, freeing up cognitive reserve, then both SA-SU and DE increase. This crossover within the upper and lower limits is considered the optimal state. The concept follows these rules: a) when WL starts within the upper and lower limits (UL/LL), then when WL increases, SA-SU and DE increase, b) when WL goes above the UL, SA-SU and DE decrease, c) when WL falls below the LL, SA-SU and DE are less than WL, d) when WL, SA-SU, and DE are within the UL/LL and WL decreases, SA-SU and DE will increase creating an optimal cognitive state supporting SA and DE, e) demonstrates that on occasion even though WL is above the UL and SA-SU are decreased, effective decisions can sometimes still be made (Doyle, 2008; Raj, Doyle, and Cameron, 2010).

Predict Behaviors That Lead to Human Error

- Predict the next type of operator behaviors
- Monitor operator behaviors and detect deviations from the expected norm.
  - Deviations (abnormal behaviors)which could eventually lead to human error.
- FSM used as a monitoring tool for human behavior but could also be used for human/ synthetic agent team behavior as well.

Modeling and Monitoring of Human Cognition: Game Theory and Bayesian Models

- Manage probabilities to identify stochastic connections between actions and consequences
- Overall, the data fusion approach for SA could benefit from both evolutionary game models estimating state determinations and solutions of the mappings between the state space and the representation hypotheses
- Classical neural net and Bayesian classifier “black box” methods, however, suffer from two problems
  - They often fail to accurately predict cognitive state when the context of the task changes significantly
  - They produce cryptic, difficult to interpret models.
  - Dimensionality reduction methods such as principal components analysis (PCA) or independent components analysis (ICA) express data more compactly, but they do not provide much insight into the data’s underlying structure, particularly its time-varying structure.

Modeling and Monitoring of Human Cognition: EEG

- Statistical techniques have shown the feasibility of classifying a person's overall cognitive state into a small number of categories from EEG data (Koska et al., 1997; Schmorrow & Stanney, 2009; Trejo et al., 2003)
  - However, classification performance suffers dramatically when the data comes from a future point in time where the operational context or environment has changed (Berka, et al., 2004; Lan, et al., 2005).
  - Understanding and modeling the EEG signal (or SA, workload, etc.) as a collection of independent components, where some components vary with external stimuli and others vary with internal cognitive states, would improve model performance.
    - The data signatures of these separate components vary with time, even with the time frame of a single task, which makes them difficult to identify.
    - Learning algorithms that assume a stationary distribution cannot handle constantly shifting sensory signals.

Modeling and Monitoring of Human Cognition: EEG Cont.

- Because brain activity associated with processing external stimuli does not remain stationary over time, traditional statistical methods assuming a stationary process can fail.
- While HMM do not assume a stationary process
  - Accuracy depends on the model designer correctly specifying the spatial and temporal structure of the underlying process generating the data
- FSM induction, however, models a multidimensional data stream as a collection of independent, time-varying components with computational efficiency while learning patterns in the data (Hong, Turk & Huang, 2000).

Modeling and Monitoring of Human Cognition: EEG-FSM Induction

- FSM induction can explicitly represent the modeling process and structure visually and in real-time to enhance understanding of the underlying process.
- The model produces an internal representation, or "memory", by
  - Segmenting the time-varying signal in terms of the signal's most frequently occurring parts
  - Detecting temporal relationships between those parts, which may or may not share the same set of spatial variables.

Modeling and Monitoring of Human Cognition: EEG-FSM Induction Cont.

- Similar to data compression, individual neuron-like processing elements come to represent the most frequent components within a signal.
- FSM induction can quickly partition a multidimensional signal with many variables into groups of correlated variables without any prior information about variable relationships.
  - If the signal represents a collection of independently evolving state trajectories, the algorithm learns to track each trajectory in the group.
  - FSM models can continue tracking changes in cognitive state despite changes in the sensory environment because it can decompose a signal into familiar and unfamiliar parts.

Modeling and Monitoring of Human Cognition: EEG-FSM Induction Cont.

- The FSM approach has the potential to model brain activity, (i.e., SA, workload etc.) as a collection of weakly dependent, stochastic processes
  - Where one or more processes correspond to the socio-technical team internal cognitive state, and other processes map to sensory processing or noise in the signal.
  - By basing the probability of future events on the occurrence of prior events without imposing a limit on how far in the past the prior events have occurred, it can model non-Markov processes such as operator and system cognitive state (Raj et al., 2009).
  - The FSM can remain robust to physiologic perturbations in the system and continue to improve its underlying model over time using new data, as it does not require a closed training set for operation.

Modeling and Monitoring of Human Cognition: EEG-FSM Induction Cont.

- Chaotic transitions likely emerge in a wide variety of cognitive phenomena (Bob et al, 2006).
  - Nonlinear observers can improve FSM models of dynamic socio-technical systems by identifying and tracking chaotic attractors corresponding to different mental states automatically.
    - Estimate the state of a nonlinear system
    - Identify mismatches (anomalies) between the model and the actual system.
    - Sliding mode variable structure observers (VSOs) can identify chaotic attractors in the data and augment the FSM of a given model by providing a longer-term memory of anomalous events and structural changes in the underlying system (Drakunov, 1984, 1992; Drakunov & Utkin, 1995).

Modeling and Monitoring of Human Cognition: EEG-FSM Induction Cont.

- VSOs can update the model when they detect an anomaly
  - Which then becomes a known pattern
  - Thus creating a change in the model structure.
  - This property allows reconstructing/estimating categories of cognitive state that can arise in different situations, such as stress, surprise, uncertainty, etc
- An adaptive automation system that combines FSM and VSO methods could identify natural variations in human-generated actions and data from background noise, as well as identify infrequent events and shifts in the structure of the system.

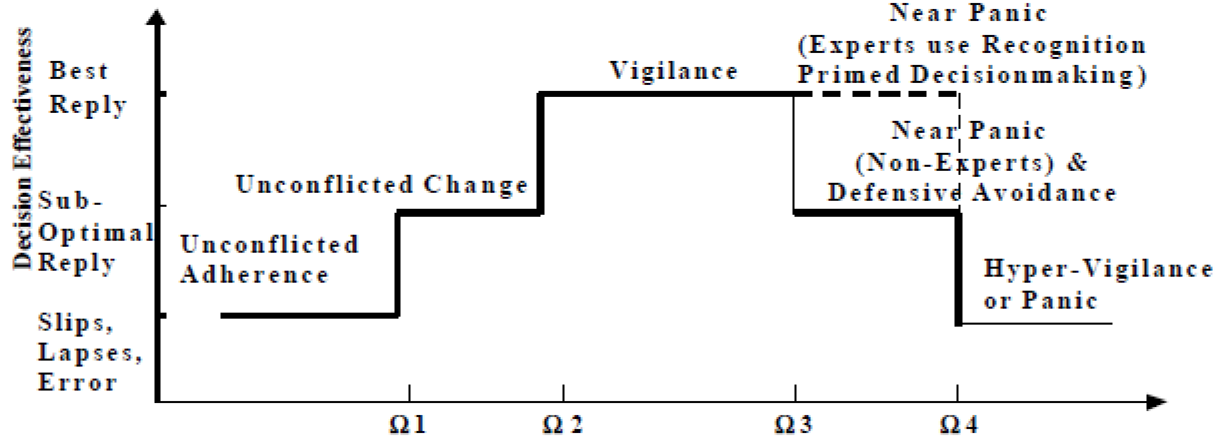
Summary

- Human Synthetic agent trust is critical to safe operation and mission effectiveness
- Increases in risk / uncertainty decreases trust
- Trust in human synthetic agent interaction is different than human-human interactions
- Synthetic agents must ask for help and acknowledge commands
- Adapt to human level of understanding and current willingness to trust in the agents capability
- Synthetic agents should be able to predict patterns in human behavior
- Perhaps, architectures that include trust factors, adaptive automation, and FSM techniques might bridge the symbiotic relationship gaps that exist in human agent/automated agent interactions today.

Backup Slides

Backup

Figure 3 - The Classic Performance Moderator Function is an Inverted-U



Use of all Avail.Info	Very Low	Medium	Complete	Low	Very Low
Time Avail. to Do Better	Very High	Adequate	Adequate	Insuf-ficient	None
Risks if don't change	None	Low or Medium	High	High	Very High

In particular, we use the algorithm to derive the values of what we call integrated stress, or the iSTRESS variable: