Cognitive Task Analysis:
Current Research

Hayley Davison

Background of CTA

- Task analysis: describing the physical tasks and cognitive plans required of a user to accomplish a particular work goal
- No one method to perform task analysis, but Hierarchical Task Analysis appears to be the preferred framework due to its flexibility
- Method chosen based on:
  - Purpose (task description, task simulation, etc.)
  - Design phase (CDIO)
  - Expertise of the researcher or domain experts

Discussion Papers

- Militello & Hutton (1998): Adapts CTA to be more usable to industry practitioners and tests reliability/validity of the new method
- Shrayne, et al. (1998): Uses the framework of HTA along with convergent methods to produce a task analysis of a safety-critical software design

Need for CTA

- Advances in technology have increased, not decreased mental demands of workers (Howell & Cooke, 1989)
- CTA describes & represents cognitive elements underlying:
  - Goal generation
  - Decision making
  - Judgments

Militello & Hutton

Problem

- Current CTA strategies require considerable time & resources:
  - Hall, Gott & Pokorny (1994) spent several years developing the Precursor, Action Result and Interpretation (PARI) method of CTA
  - Roth, Woods, & Pople (1992), Seamster, et al. (1993), Rasmussen (1986), and Rouse (1984)

Militello & Hutton

Potential Solution

- Transition the research efforts of CTA into an applied approach that can be used by system designers rather than human factors professionals
- ACTA
  - Streamlined CTA methods developed for system designers to elicit & represent cognitive task performance
  - Means to transform this data into design recommendations

Militello & Hutton

Applied Cognitive Task Analysis (ACTA)

- Task diagram interview
- Knowledge audit
- Simulation interview
- Cognitive demands table

Militello & Hutton

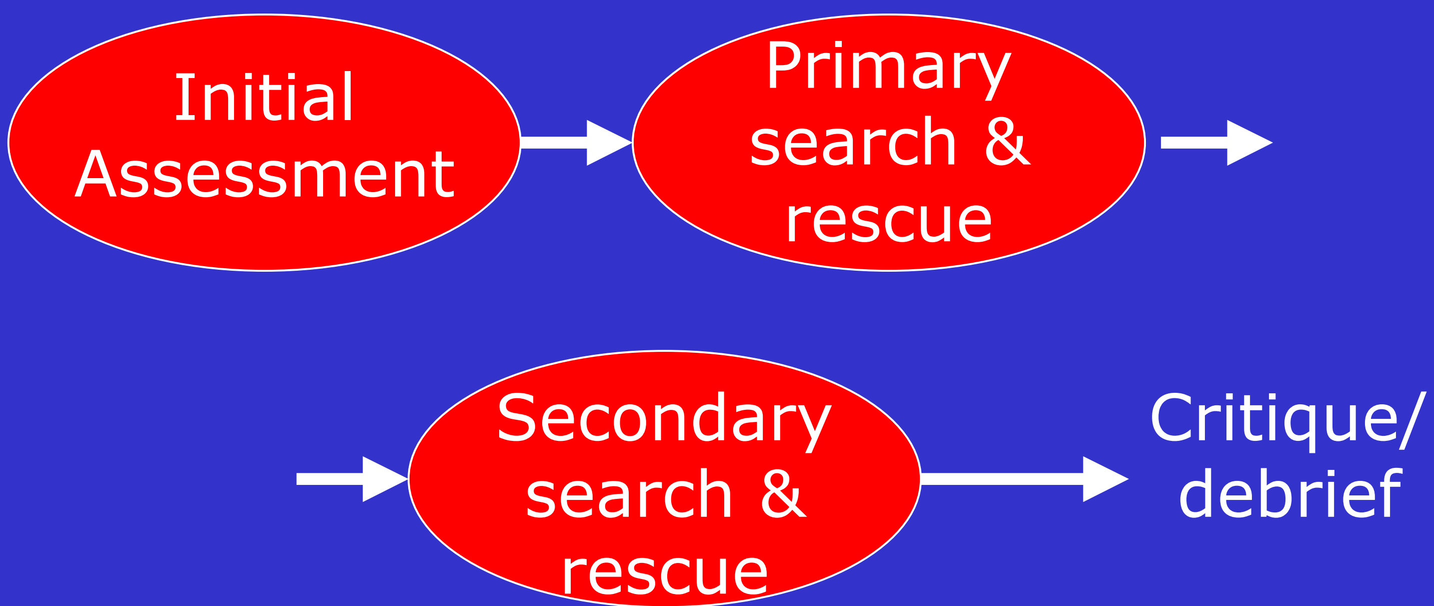
Part 1: Task diagram

- Purpose: provide broad overview of the task & highlights difficult cognitive portions of task to further investigate
- How:
  - Subject matter expert is asked to break the task down into steps or subtasks (3-6 subtasks, ideally)
  - Ask the expert which subtasks require "cognitive skill"– e.g., judgments, assessments, problem-solving
  - Diagram subtasks, indicating sequence and need for cognitive skill

Militello & Hutton

Task diagram example

Fireground Command



Militello & Hutton

Part 2: Knowledge Audit

- Purpose: surveys the expertise required for a task through probing concrete examples in the job context
- How:
  - Probes into knowledge categories surrounding expertise (diagnosing & predicting, situation awareness, perceptual skills, developing and knowing when to apply tricks of the trade, improvising, metacognition, recognizing anomalies, and compensating for equipment limitations
  - Probes result in an inventory of task-specific expertise

Militello & Hutton

Knowledge Audit example

- Past & Future
  - *Is there a time when you walked into the middle of a situation and knew exactly how things got there and where they were headed?*
- Big Picture
  - *Can you give me an example of what is important about the Big Picture for this task?*
- Noticing
  - *Have you had experiences where part of a situation just "popped" out at you; where you noticed things going on that others didn't catch? What is an example?*
- Job Smarts
  - *When you do this task, are there ways of working smart or accomplishing more with less– that you have found especially useful?*
- Self-Monitoring
  - *Can you think of a time when you realized that you would need to change the way you were performing in order to get the job done?*

Militello & Hutton

Knowledge Audit Table

Aspects of expertise	**Cues & strategies**	**Why difficult?**
Past & Future e.g., Explosion in office strip; search the office areas rather than source of explosion	Material safety data sheets (MSDS) tell you that explosion in area of dangerous chemicals and information about chemicals	Novice would be trained to start at source and work out May not look at MSDS to find source
Big Picture Big picture includes source of hazard, potential location of victims, ingress/egress routes, other hazards	Senses, communication with others, building owners, MSDS, building pre-plans	Novice gets tunnel vision, focuses on one thing, e.g., victims

Militello & Hutton

Part 3: Simulation interview

- Purpose: allows interviewer to probe the cognitive processes of the expert within the context of a specific scenario
- How:
  - Identify a "challenging scenario" either from a training scenario or a self-constructed scenario (based on a previous CTA)
  - Show the expert the scenario and ask them to identify major events and be prepared to answer questions about the scenario
  - Probe expert's assessment, actions, critical cues, and potential errors

Militello & Hutton

Example Situation Interview Probe & Table

Probe: *As the (job) in this scenario, what actions, if any, would you take at this point in time? What do you think is going on here? What is your assessment of the situation at this point in time? What pieces of information led you to this situation assessment and these actions? What errors would an inexperienced person be likely to make in this situation?*

Events	**Actions**	**Assessment**	**Critical Cues**	**Potential Errors**
On-scene arrival	Account for people (names) Ask neighbors Must knock on or knock down to make sure people aren't there	It's a cold night, need to find place for people who have been evacuated	Night time Cold: <15 degrees Dead space Add on floor Poor materials	Not keeping track of people (could be looking for people who are not there)

Militello & Hutton

Part 4: Cognitive Demands Table

- Purpose: means to consolidate & synthesize data
- How:
  - Gather data from previous 3 interviews
  - Arrange the data in terms of the type of information that the designers will need to develop a new course or design an new system

Militello & Hutton

Example Cognitive Demands Table

Difficult cognitive element	Why difficult?	Common errors?	Cues and strategies used
Knowing where to search after an explosion	Novices may not be trained in explosions. Other training suggests to start at the source. Not everyone knows about MSDS.	Novice would be likely to start at source of explosion. Starting at source is a rule of thumb.	Start where likely to find victims. Refer to MSDS. Consider type of structure.
Finding victims in a burning building.	There are lots of distracting noises. Breathing distracts.	Novices think that own breathing is a victim.	You & partner stop, hold breath, & listen. Crying, talking, bumping into things.

Militello & Hutton

Is ACTA valid and reliable?

- Little attention has been devoted to determine validity and reliability of methods of eliciting a knowledge base
- An evaluation study was conducted to address issues of validity & reliability for ACTA as well as more general assessment of reliability within the context of real-world tasks

Militello & Hutton

Evaluation Method

- Participants: graduate students with no previous CTA experience
  - 12 firefighting
  - 11 electronic warfare
- Groups: random assignment to either
  - Unstructured- 2 hr. workshop on CTA concepts and application of CTA to developing instructional materials
  - ACTA- 6 hr. workshop on ACTA techniques

Militello & Hutton

Evaluation Method (ct'd)

- Task:
  - Conduct 1 interview with an expert and observe 1 other student's interview focusing on domain-specific task (size-up task or signal threat task)
  - Attend 4 hr session on analyzing interviews & developing training materials
  - Consolidate data
  - Develop 10 learning objectives for a novice
  - Revise or add to training materials based on interview material
  - Complete a questionnaire about participation in study

Militello & Hutton

Data transformation

- Data were evaluated by experts and cognitive psychologists
- Validity: data were evaluated to determine
  - Whether ACTA tools produced information that was predominantly cognitive in nature
  - Whether information produced was domain-specific and relevant
- Reliability: multiple coders were used to ensure inter-rater reliability

Militello & Hutton

Validity indices

- Cognitive demands table
  - Items must address a cognitive skill or a cognitive challenge that a firefighter/EW operator encounters on the job, NOT declarative knowledge *single rater used*
  - Information elicited across relevant cognitive categories (Rasmussen's decision making model) *74% reliability*
  - Domain-specific content (firefighting only) into categories of task *81% reliability*
  - Expert evaluation: percentage of info. known only to experienced personnel & percentage of info. relevant to highly experienced personnel

Militello & Hutton

Validity indices

- Instructional materials
  - Expert evaluation on materials' importance (3-point scale) *87.8% reliability*
  - Expert evaluation on materials' accuracy (2-point scale) *71.4% reliability*

Militello & Hutton

Results

- Usability: participants rated usability of the ACTA techniques on average above a "3" on a 5-point scale
- Validity:
  - 93% of the ACTA group generated cognitive issues
  - 90% of the material was judged by experts to be material only highly-experienced personnel would know
  - 95% of the firefighting learning objectives were deemed important, 83% of the EW objectives were considered important
  - 92% of the firefighting learning objectives were found to be accurate, 54% of the EW objectives were found to be accurate

Militello & Hutton

Results

- Reliability: no well-established metric to identify reliability of CTA tools
  - 80% of the cognitive demands tables included information collection as a category of cognitive information
  - 62% of the information gathered from the firefighting group focused on the size-up task (the focus of the CTA)
  - No identical cognitive demands were generated by the participants, which is not surprising given the flexibility of the ACTA method and the small sample size
- Group differences: few differences were found between ACTA and Unstructured groups due to small sample sizes & large intra-group variability, indicating importance in CTA concepts

Militello & Hutton

Conclusions

- Although an attempt was made to evaluate validity and reliability, the authors concede that no proven techniques exist to do so
- Streamlined ACTA techniques appear to produce less comprehensive information than other more systematic techniques (e.g., Klein’s Critical decision method & Rasmussen’s Cognitive analysis)
- ACTA does provide method of eliciting critical information

Militello & Hutton

Human Error Identification

- In safety-critical systems, human error identification is particularly important
- Probabilistic Risk Assessment has been used to determine a set of undesirable outcomes based on the paths leading to the outcomes and the probability of their occurrence
- The authors claim that PRA is useful in performing HEI on many mechanical systems, but physical constraints do not exist in software systems

Shrayne, et al.

Problem

- The effectiveness of other HEI methods are dependent on the expertise of the analyst
- These techniques may not take into consideration a poor mental model on the part of the expert

Shrayne, et al.

Potential Solution

- A convergent methods approach to task analysis must be used to combat the issues with analyst bias and expertise
- Methods using task performance as well as expert opinion must also be used to address the problem of a faulty mental model

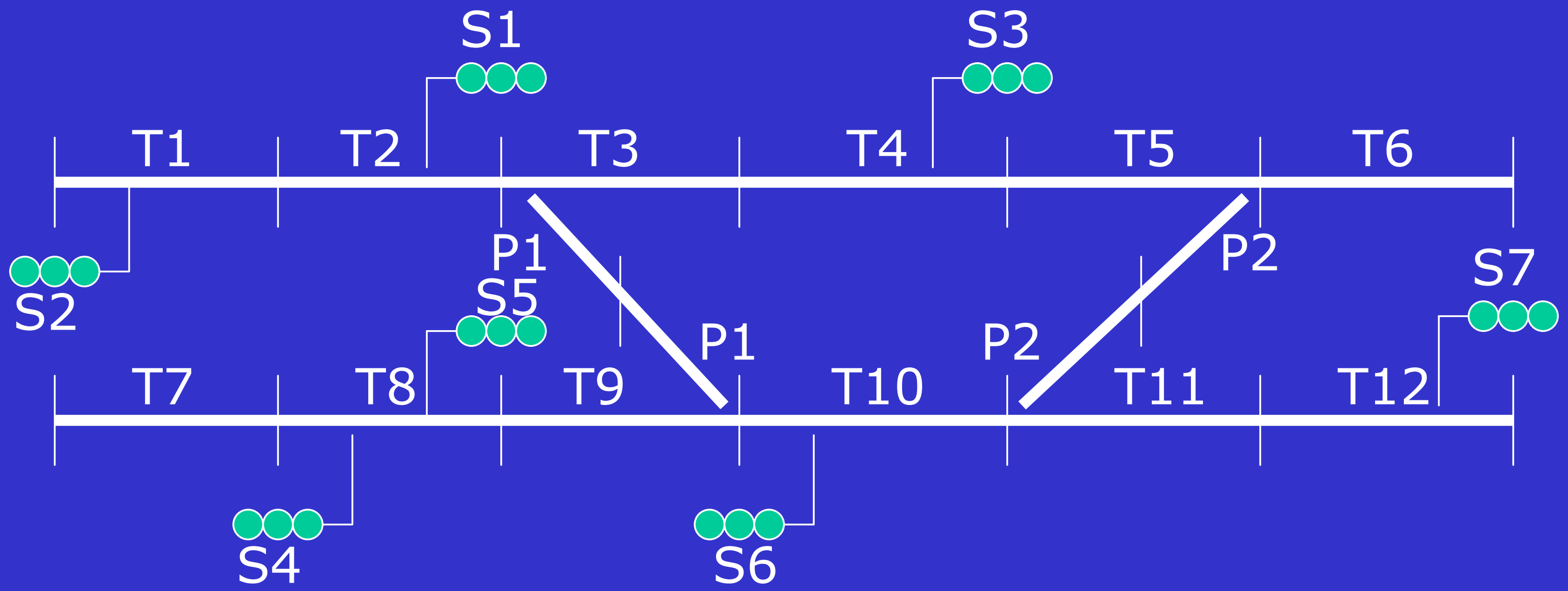
Shrayne, et al.

Case study: Railway signaling safety system

- Solid State Interlocking (SSI): automatically controls signals, points, etc. in a section of the railway allowing only safe movements of trains to occur
- Each SSI is triply hardware redundant, but has only one geographical database

Shrayne, et al.

SSI system



Shrayne, et al.

Analysis

- Hierarchical Task Analysis
  - Interviews
  - Documentation
  - Observation
- Error Analysis
  - Error log audit
  - Error observation
  - Work sample test
  - Laboratory experiment

Shrayne, et al.

Hierarchical Task Analysis

- Process analysis: described production process, functioning of equipment, jargon, & general task environment
- Data sources:
  - Interviews: accessible & flexible; purely subjective & liable to error & bias
  - Documentation: detailed info.; how task *should be* done, not how *is* done
  - Observation: allows features & problems of task performance to be discovered; "observer effect" affects expert negatively

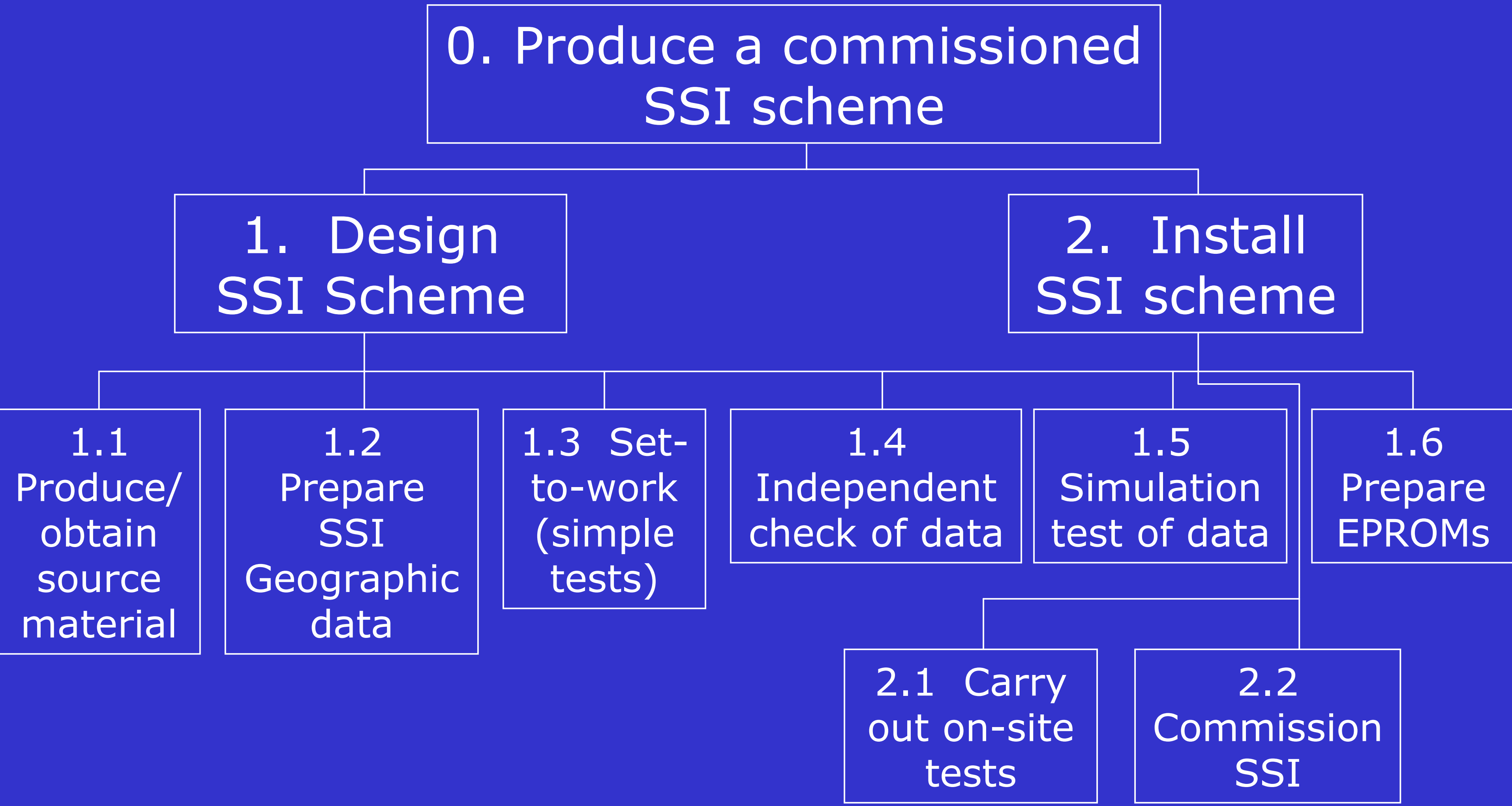
Shrayne, et al.

HTA Results

- 7 levels of hierarchy, 150 individual operations structured by 40 plans of various complexity
- Task division: office-based production & site-based production
- Programming process: Preparation, Set to work, Bench checking, Testing

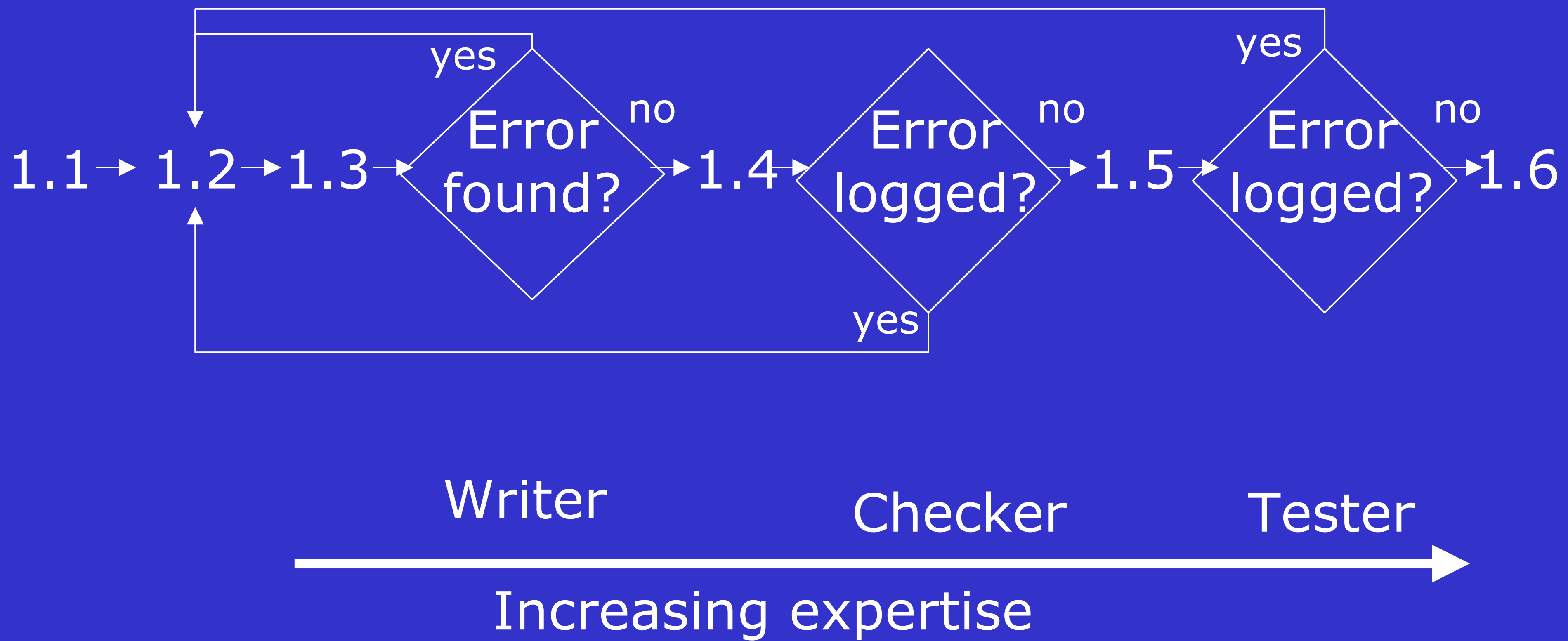
Shrayne, et al.

HTA for SSI system



Shrayne, et al.

Plan



Shrayne, et al.

Problems found by HTA

- Automation introduction
  - Automation introduced to write all simple rule-based code
  - Eliminates training that novices receive on simple tasks, leaving them less equipped to handle complex tasks
- Parallel checking and testing
  - Time pressure sometimes forces testing to occur in parallel rather than in serial fashion
  - With sloppy version control, unchecked code could be signed-off as safe

Shrayne, et al.

Need for Error Analysis

- HTA provided useful framework to identify potential problem areas of a task
- HTA does not reveal how all of the variables that may affect task performance will actually combine to produce task error

Shrayne, et al.

Error Analysis

- HTA used to identify key stages of SSI data production: writing, checking & testing
- Error analysis performed to generate empirical data
  - Error log audit
  - Error observation
  - Work sample test
  - Laboratory experiment

Shrayne, et al.

Error Log Audit

- Method
  - Logs from data checkers to the data writer were collected and errors were analyzed
  - Errors not in log
    - Caught by data writer
    - Not caught by either data writer or data checkers
- Results: 580 faults from checking & testing
  - 12.4% False alarms
  - 9.7% Identity and labeling errors
  - 33.4% Route setting errors
  - Problems: checking/testing not matched; faults caught at checking were not measurable at testing; no measure of faults escaping checking AND testing

Shrayne, et al.

Error observation

- Method
  - Different engineers were videotaped for several hours performing each of the three main task areas
  - Engineers were asked to verbalize the process of catching their own mistakes during code writing & checking & then categorize error in terms of skill, rule, or knowledge-based
- Results
  - 4.1 errors/hr in writing
  - 8.3 errors/hr in writing (editing only)
  - 0 errors in checking
  - .8 errors/hr in testing
  - No knowledge-based errors detected

Shrayne, et al.

Work Sample Test

- Method
  - Expert engineer and programmer chose a representative piece of code that included all main aspects of the writing task
  - Sections of the data were removed and partial code was given to 15 participants to complete
- Results
  - More rule-based errors were made than knowledge-based errors
  - Knowledge-based data were more time consuming to complete, with over 200s longer than rule-based data
  - Common-mode failure emerged4 times and once in 13 of 15 participants

Shrayne, et al.

Laboratory Experiment

- Method
  - Task simulations were developed for completion by novice participants to compare performance of checkers versus testers
  - 13 participants acted as checkers on the simulation, and 27 acted as testers
  - 16 faults were included per simulation that were chosen from actual logged faults
- Results
  - No main effect of task type (checking versus testing)
  - Significant interaction between task type and fault type, indicating poor performance of checkers in detecting “opposing locking I” faults and testers in detecting “signal aspect control” faults

Shrayne, et al.

Discussion

- Error type
  - HTA identified knowledge-based design tasks as the most problematic; work sample showed that this was the most difficult task, and time taken to complete this was greater than rule-based tasks
  - Error log audit discovered that the majority of faults found were rule-based
  - Introduction of automation may provide a solution to the rule-based fault
- Common mode error
  - HTA suggested that there was a similarity in the checking task to the writing task, which encouraged common mode error to occur during checking
  - Even highly trained engineers (87% in the work sample) make the same cognitive errors given the same task to complete

Shrayne, et al.

Discussion (ct’d)

- Task diversity
  - HTA suggested differences between task of checking and testing
  - Error log audit and Lab experiment confirmed that different errors were found by the 2 tasks

Shrayne, et al.

Conclusions

- Techniques varied in ability to capture event (and error) frequencies
  - HTA lacking
  - Error analysis helpful
- Empirical techniques differ in tradeoff between validity and control
  - Error logging & observation- highly externally valid
  - Lab experiment- internally valid & controlled
  - Work sample- somewhere in between
- Benefits found in some form from ALL of the techniques used

Shrayne, et al.

Discussion Questions

- What is the best method to get validity from our CTA?
- Is there a way of achieving reliability in our CTA techniques? Is reliability important?
- Can one person or team of persons compile a “good” CTA?
- When is a CTA “good enough”?