Cognitive architectures: What they are and future problems and perspectives [50+30]

Frank E. Ritter, College of IST, Penn State
frank.ritter@psu.edu
21apr16
papers on line
Why cognitive architectures?

* Make computers like people except they will be my friend, Rob Ward, ~1990

* To create model users (CM&N, 1983; Booher & Minniger, 2003)

* To create model opponents and colleagues (video games, simulations)

* To provide a path towards reuse and cumulation

* To create a unified theory in psychology (UTC) (Newell, 1990) (memory, Fitts law, Hick’s law, reading, mental models, .....)

* To understand the mind
Simple mechanism
More complex mechanism: Place the cameras
More complex mechanism

- Rapid behavior
- Need to record it, too fast to analyze
- Not commonsense
- Complex
Range of ways of simulating cognition

* Eliza (if-then rules, applyer)
* Sims (objects, tasks-object pairs, applyer)
* Lisp, Java (code, interpreter)
* Fortran, assembler (code, compiler, hardware)
* AI planning languages (plans, constraints, planner)
* Expert systems (rules, interpreter)

...and more....
Yet more complex mechanism

How to Build a BRAIN
A Neural Architecture for Biological Cognition

Chris Eliasmith

OXFORD
What is a Cognitive Architecture?

* A unified theory of cognition (UTC)
* A unified theory of behavior (UTC+PM)
* The mechanisms of cognition, the wheels and gears and levers, and buffers and storage....
* A computer program that is architecture + knowledge
* An AI agent architecture that "will run slower and make mistakes!"
Components

- Input(s)  (some CAs are like a brain in a vat)
- Some storage (memories)
- Some information processing
  - typically it means (production) rule-based behavior
  - sometimes case-based or incidence-based
  - neural-level mechanisms, or combos
- Outputs (keystrokes, symbols, motor output)
Simple architecture:
The MHP is useful for HCI and design

Model Human Processor

Card, Moran, & Newell, 1983
Keystroke-Level Model

- Assumes
  - Expert, & Error free, & Single task
- Task time =

\[ \Sigma T_{\text{Mental OPs}} \approx 1.35 \text{ s (fixed, but varies)} \]
\[ T_{\text{Keystroke}} \approx 0.3 \text{ s (also IDs)} \]
\[ T_{\text{Mouse}} \approx 1.1 \text{ s (also Fitts)} \]
\[ T_{\text{Home}} \approx 0.4 \text{ s} \]
\[ T_{\text{Sys}} \approx 0.0 \text{ s} \]

...a few more operators

(Card, Moran, & Newell, 1983)
ACT-R

STARTS TO SHOW HOW MIND WORKS, HOW PARTS COME TOGETHER, WITH: LEARNING, ERRORS, MULTI-TASKS
Work with St. Amant, can be used to test designs
ACT-R 6: Anderson, 2007

HELPS EXPLAIN HOW THE MIND CAN EXIST IN THE PHYSICAL UNIVERSE

- Rule-based
- sorted by PS
- Decision cycle
- Timing from decision cycle
- Subgoaling with
- Chunking
Symbolic Long-Term Memories

Procedural

Semantic

Episodic

Reinforcement Learning

Chunking

Semantic Learning

Episodic Learning

Symbolic Short-Term Memory

Clustering

Perception

LT Visual Memory

ST Visual Imagery

Action

Decision Procedure

(Laird, 2008)
BDI: CoJack (Ritter et al., 2012)
ACT-R/Φ

(Dancy, Ritter, Berry, Klein, 2015)

<table>
<thead>
<tr>
<th>HumMod System</th>
<th># Vars</th>
<th>Example Var</th>
</tr>
</thead>
<tbody>
<tr>
<td>Body Fluids</td>
<td>214</td>
<td>Blood Plasma Volume</td>
</tr>
<tr>
<td>Circulation</td>
<td>426</td>
<td>Sinoatrial (SA) Node Rate</td>
</tr>
<tr>
<td>Electrolytes</td>
<td>140</td>
<td>Sodium Ion (NA+) Pool Mass</td>
</tr>
<tr>
<td>Hormones</td>
<td>534</td>
<td>Adrenocorticotropic hormone</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Secretion</td>
</tr>
<tr>
<td>Metabolism</td>
<td>321</td>
<td>Energy Stored (Calories)</td>
</tr>
<tr>
<td>Nervous System</td>
<td>187</td>
<td>Norepinephrine (NE) Pool Mass</td>
</tr>
<tr>
<td>Organs</td>
<td>2,349</td>
<td>Bladder Volume</td>
</tr>
<tr>
<td>Respiration</td>
<td>326</td>
<td>Breathing Tidal Volume</td>
</tr>
<tr>
<td>Other Systems (Lifestyle, Heat, etc.)</td>
<td>2,026</td>
<td>Skin Temperature</td>
</tr>
</tbody>
</table>

![Graph showing data analysis](image-url)
Other Architectures

* There are at least four substantial reviews (Pew & Mavor, 1998; Ritter et al., 2003; Morrison, 2003), with perhaps >100 architectures

* Remaining major architectures include: Spawn (neural levels: Eliasmith), Epic (no cognitive constraint, just IO: Kieras & Meyer), Icarus (hierarchical knowledge representation: Langley), Clarion (hierarchical structures: Sun), APEX (engineering applications: Freed); Psi (Bach & Doerner)
Summary

* Common components:
  * Input and output
  * Memory(s)
  * Knowledge

* Based on different perspectives, memory, attention, learning, knowledge representation, hierarchy, hybrid

* None are complete, many are useful
A Couple of Insights

* Grant (1962) Testing extreme hypotheses

* Hard to test these things, treat them like parachutes, worth taking seriously? and how to improve them

* Treat these models as theories

* Don’t sample their behavior, find it (Ritter et al. 2011)
App: Human-computer interaction

* ABCS of HCI

* The foundations for designing user-centered systems: What system designers need to know about people

* 380 pages of data

* 20 pages of theory
App: Networks and cognition
Creating Networks with a Large Number of Intelligent Nodes

* Provides visualizations of networks and network evolution with more human-like agents

Interaction: Bitmap based interaction

* Provides models access to simulations and thus knowledge


pb270194.mov

App: User modeling of Interaction SegMan

* Lets models test interfaces and theories
* could save 30 years/day


pb270194.mov
reifers Demo804.mov
App: Moderators Challenge and threatened

* Models of challenged and threatened behavior


App: Exploring Moderators, Social impacts on cognition

* Start to represent the effects of social aspects on cognition and behavior
* How can you break the will of an agent with no will?
* But, how does this interact with cognition? Can we reuse this? [dunno and no and needs more work]


\[
p_a = \frac{1}{1 + \left(1 + e^{\frac{g_{\text{size}} g_{\text{observer}} (c + k)}{1 + k} \sqrt{\text{size}} + \text{leader}} \right)}
\]
Resources for Usability: High-level Languages like Herbal

- A review of high-level languages
- Modeling differences in expertise
- 20 min. non-repetitive task
- 9 rules+540 facts/543 rules per model, 10K total learned rules
- N=40 human subjects


Remaining Issues for UTCs

* Language
* How does physiology support cognition?
* Multiple-levels of representation AND learning
* Individual differences
* Large scale learning
* ....

* A point is insights
* Another point is cumulation and Reuse

* Single model of multiple but not all 2-person games
* $US 10 for 50

http://www.amazon.com/Sterile-Eye-Pads-Box-50/dp/B002U2I4JA/ref=sr_1_2?ie=UTF8&qid=1404291564&sr=8-2&keywords=eye+pads
Conclusions

Issues for your work and for CognitiveScience

- Usability of theories matters
- Adding more types of human behaviors matters
  - Learning, forgetting, and the rest of the hard, traditional cognitive aspects for UTCs/UTBs
  - Social aspects of cognition, including networks
  - Moderated aspects of behavior and of will
- Interaction with the world, matters
- Explanations of the results (Newell, 1990, p. 503)
  - We need some good jokes as a way to present our stories 😅
  - We need good diagrams, displays, and scenarios to explain our models
  - We need good movies to explain our models
- We need models easy to use and reuse (Newell, 1990, p. 503)
  - Reuse seems to be architecture extensions (not knowledge), which is surprising
  - These seem to be general software issue
- CogArchitecture is a path towards better psychology and general, human-level AI with lots of applications
Behavioral Modeling with the Herbal High-Level Language [30 min.]

Frank E. Ritter

College of Information Sciences and Technology
Penn State

acs.ist.psu.edu/papers
frank.ritter@psu.edu

Acknowledgements:
ONR, Kim, Paik, Yeh, Cohen, D2P team.

This project has been supported by several ONR grants.
19apr2016
Overview

- Bit about using to model novice to expert using a HLBR language
- Conclusions
Implementing Large User Models with Multiple Levels of Expertise

- ACT-R compiler
- Data
- comparison

If there is this in the perceptual buffer & that in DM
--> Put this in motor

If this is in DM & this2 in DM
--> put this this+that in goal

If this is in perceptual buffer
--> put move-eye(loc(this)) in motor

How the compiler works

Declarative Memory HTA

Novice trace

Middle expert

Expert
Dismal task

(1) Open a file, named normalization.dis under the experiment folder
(2) Save as the file with your initials
(3) Calculate and fill in the Frequency column (B6 to B10)
(4) Calculate the total frequency in B13
(5) Calculate and fill in the Normalization column (C1 to C5)
(6) Calculate the total normalization in C13
(7) Calculate the Length column (D1 to D10)
(8) Calculate the total of the Length column in D13
(9) Calculate the Typed Characters column (E1 to E10)
(10) Calculate the total of the Typed Characters column in E13
(11) Insert two rows at cell A0
(12) Type in your name in A0
(13) Fill in the current date in A1 using the command dis-insert-date
(14) Save your work as a printable format

(Dismal, a spreadsheet in Emacs)
(RUI, a keystroke logger)
(Kim & Ritter, 2015)
Data on the Dismal Task

- Time (s) vs. Day
- Mouse

Graph showing a decrease in time over days.
## Model Performance

<table>
<thead>
<tr>
<th></th>
<th>Initial rules</th>
<th>Learned rules</th>
<th>DMs used on Trial 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Novice</td>
<td>29</td>
<td>253</td>
<td>1,152</td>
</tr>
<tr>
<td>0% Expertise</td>
<td>617</td>
<td>197</td>
<td>1,152</td>
</tr>
<tr>
<td>10% Expertise</td>
<td>617</td>
<td>199</td>
<td>1,091</td>
</tr>
<tr>
<td>20% Expertise</td>
<td>617</td>
<td>197</td>
<td>1,030</td>
</tr>
<tr>
<td>30% Expertise</td>
<td>617</td>
<td>199</td>
<td>968</td>
</tr>
<tr>
<td>40% Expertise</td>
<td>617</td>
<td>199</td>
<td>908</td>
</tr>
<tr>
<td>50% Expertise</td>
<td>617</td>
<td>199</td>
<td>845</td>
</tr>
<tr>
<td>60% Expertise</td>
<td>617</td>
<td>196</td>
<td>784</td>
</tr>
<tr>
<td>70% Expertise</td>
<td>617</td>
<td>199</td>
<td>723</td>
</tr>
<tr>
<td>80% Expertise</td>
<td>617</td>
<td>199</td>
<td>661</td>
</tr>
<tr>
<td>90% Expertise</td>
<td>617</td>
<td>198</td>
<td>600</td>
</tr>
<tr>
<td>100% Expertise</td>
<td>617</td>
<td>197</td>
<td>538</td>
</tr>
</tbody>
</table>

![Graph showing task completion time over trials for different expertise levels](graph)
Model (predictions) to Aggregate Data

- Human Data
- KLM
- Novice
- 10% Expert
- 50% Expert
- 100% Expert

Task Completion Time (sec.)

Trials
Model to individual data fits

Best

Avg

Worst
<table>
<thead>
<tr>
<th></th>
<th>BestFit</th>
<th>Nov</th>
<th>0</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>s7</td>
<td>Nov</td>
<td>0.959</td>
<td>0.949</td>
<td>0.952</td>
<td>0.952</td>
<td>0.965</td>
<td>0.965</td>
<td>0.964</td>
<td>0.964</td>
<td>0.963</td>
<td>0.963</td>
<td>0.952</td>
<td>0.951</td>
</tr>
<tr>
<td>s8</td>
<td>Ex0</td>
<td>0.938</td>
<td>0.945</td>
<td>0.944</td>
<td>0.941</td>
<td>0.940</td>
<td>0.940</td>
<td>0.939</td>
<td>0.939</td>
<td>0.938</td>
<td>0.938</td>
<td>0.937</td>
<td>0.936</td>
</tr>
<tr>
<td>s13</td>
<td>Nov</td>
<td>0.994</td>
<td>0.989</td>
<td>0.990</td>
<td>0.991</td>
<td>0.992</td>
<td>0.992</td>
<td>0.992</td>
<td>0.991</td>
<td>0.991</td>
<td>0.991</td>
<td>0.991</td>
<td>0.990</td>
</tr>
<tr>
<td>s17</td>
<td>Ex100</td>
<td>0.972</td>
<td>0.973</td>
<td>0.973</td>
<td>0.974</td>
<td>0.974</td>
<td>0.974</td>
<td>0.975</td>
<td>0.975</td>
<td>0.976</td>
<td>0.976</td>
<td>0.977</td>
<td>0.978</td>
</tr>
<tr>
<td>s18</td>
<td>Ex100</td>
<td>0.989</td>
<td>0.988</td>
<td>0.988</td>
<td>0.989</td>
<td>0.989</td>
<td>0.989</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
<td>0.991</td>
</tr>
<tr>
<td>s21</td>
<td>Nov</td>
<td>0.980</td>
<td>0.972</td>
<td>0.974</td>
<td>0.974</td>
<td>0.976</td>
<td>0.976</td>
<td>0.977</td>
<td>0.977</td>
<td>0.977</td>
<td>0.977</td>
<td>0.977</td>
<td>0.977</td>
</tr>
<tr>
<td>s24</td>
<td>Nov</td>
<td>0.979</td>
<td>0.972</td>
<td>0.974</td>
<td>0.974</td>
<td>0.976</td>
<td>0.976</td>
<td>0.977</td>
<td>0.977</td>
<td>0.977</td>
<td>0.977</td>
<td>0.977</td>
<td>0.977</td>
</tr>
<tr>
<td>s29</td>
<td>Nov</td>
<td>0.990</td>
<td>0.985</td>
<td>0.987</td>
<td>0.987</td>
<td>0.988</td>
<td>0.988</td>
<td>0.988</td>
<td>0.988</td>
<td>0.987</td>
<td>0.987</td>
<td>0.987</td>
<td>0.987</td>
</tr>
<tr>
<td>s30</td>
<td>Nov</td>
<td>0.997</td>
<td>0.994</td>
<td>0.995</td>
<td>0.996</td>
<td>0.996</td>
<td>0.996</td>
<td>0.996</td>
<td>0.996</td>
<td>0.996</td>
<td>0.996</td>
<td>0.996</td>
<td>0.996</td>
</tr>
<tr>
<td>s33</td>
<td>Nov</td>
<td>0.998</td>
<td>0.996</td>
<td>0.997</td>
<td>0.997</td>
<td>0.997</td>
<td>0.997</td>
<td>0.997</td>
<td>0.997</td>
<td>0.997</td>
<td>0.997</td>
<td>0.997</td>
<td>0.997</td>
</tr>
<tr>
<td>s35</td>
<td>Nov</td>
<td>0.980</td>
<td>0.973</td>
<td>0.975</td>
<td>0.975</td>
<td>0.977</td>
<td>0.977</td>
<td>0.977</td>
<td>0.977</td>
<td>0.976</td>
<td>0.976</td>
<td>0.976</td>
<td>0.976</td>
</tr>
<tr>
<td>s37</td>
<td>Ex100</td>
<td>0.998</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
</tr>
<tr>
<td>s38</td>
<td>Nov</td>
<td>0.970</td>
<td>0.961</td>
<td>0.963</td>
<td>0.963</td>
<td>0.966</td>
<td>0.966</td>
<td>0.965</td>
<td>0.965</td>
<td>0.964</td>
<td>0.964</td>
<td>0.964</td>
<td>0.963</td>
</tr>
<tr>
<td>s39</td>
<td>Ex100</td>
<td>0.993</td>
<td>0.995</td>
<td>0.995</td>
<td>0.995</td>
<td>0.995</td>
<td>0.995</td>
<td>0.995</td>
<td>0.995</td>
<td>0.996</td>
<td>0.996</td>
<td>0.996</td>
<td>0.996</td>
</tr>
<tr>
<td>s43</td>
<td>Ex100</td>
<td>0.976</td>
<td>0.981</td>
<td>0.980</td>
<td>0.980</td>
<td>0.979</td>
<td>0.979</td>
<td>0.979</td>
<td>0.979</td>
<td>0.980</td>
<td>0.981</td>
<td>0.981</td>
<td>0.982</td>
</tr>
<tr>
<td>s47</td>
<td>Nov</td>
<td>0.985</td>
<td>0.979</td>
<td>0.981</td>
<td>0.982</td>
<td>0.983</td>
<td>0.983</td>
<td>0.983</td>
<td>0.983</td>
<td>0.982</td>
<td>0.983</td>
<td>0.982</td>
<td>0.982</td>
</tr>
<tr>
<td>s52</td>
<td>Ex100</td>
<td>0.999</td>
<td>0.998</td>
<td>0.998</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
</tr>
<tr>
<td>s53</td>
<td>Ex0</td>
<td>0.994</td>
<td>0.997</td>
<td>0.996</td>
<td>0.996</td>
<td>0.996</td>
<td>0.995</td>
<td>0.995</td>
<td>0.996</td>
<td>0.996</td>
<td>0.996</td>
<td>0.996</td>
<td>0.996</td>
</tr>
<tr>
<td>s55</td>
<td>Ex100</td>
<td>0.992</td>
<td>0.993</td>
<td>0.993</td>
<td>0.994</td>
<td>0.994</td>
<td>0.994</td>
<td>0.994</td>
<td>0.994</td>
<td>0.995</td>
<td>0.995</td>
<td>0.995</td>
<td>0.995</td>
</tr>
<tr>
<td>s58</td>
<td>Ex10</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
</tr>
<tr>
<td>s60</td>
<td>Nov</td>
<td>0.950</td>
<td>0.939</td>
<td>0.942</td>
<td>0.943</td>
<td>0.946</td>
<td>0.946</td>
<td>0.945</td>
<td>0.945</td>
<td>0.944</td>
<td>0.944</td>
<td>0.944</td>
<td>0.943</td>
</tr>
<tr>
<td>s62</td>
<td>Nov</td>
<td>0.986</td>
<td>0.980</td>
<td>0.982</td>
<td>0.982</td>
<td>0.983</td>
<td>0.983</td>
<td>0.983</td>
<td>0.982</td>
<td>0.982</td>
<td>0.982</td>
<td>0.982</td>
<td>0.981</td>
</tr>
<tr>
<td>s64</td>
<td>Ex0</td>
<td>0.959</td>
<td>0.968</td>
<td>0.965</td>
<td>0.965</td>
<td>0.963</td>
<td>0.964</td>
<td>0.964</td>
<td>0.964</td>
<td>0.965</td>
<td>0.965</td>
<td>0.965</td>
<td>0.966</td>
</tr>
<tr>
<td>s65</td>
<td>Ex0</td>
<td>0.998</td>
<td>1.00</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
</tr>
<tr>
<td>s69</td>
<td>Nov</td>
<td>0.985</td>
<td>0.981</td>
<td>0.983</td>
<td>0.982</td>
<td>0.983</td>
<td>0.983</td>
<td>0.982</td>
<td>0.983</td>
<td>0.982</td>
<td>0.981</td>
<td>0.981</td>
<td>0.980</td>
</tr>
<tr>
<td>s70</td>
<td>Nov</td>
<td>0.999</td>
<td>0.997</td>
<td>0.998</td>
<td>0.998</td>
<td>0.998</td>
<td>0.998</td>
<td>0.998</td>
<td>0.998</td>
<td>0.997</td>
<td>0.997</td>
<td>0.997</td>
<td>0.997</td>
</tr>
<tr>
<td>s71</td>
<td>Nov</td>
<td>0.569</td>
<td>0.562</td>
<td>0.565</td>
<td>0.559</td>
<td>0.561</td>
<td>0.561</td>
<td>0.559</td>
<td>0.558</td>
<td>0.554</td>
<td>0.553</td>
<td>0.551</td>
<td>0.547</td>
</tr>
<tr>
<td>s74</td>
<td>Nov</td>
<td>0.987</td>
<td>0.982</td>
<td>0.983</td>
<td>0.984</td>
<td>0.985</td>
<td>0.985</td>
<td>0.985</td>
<td>0.985</td>
<td>0.984</td>
<td>0.984</td>
<td>0.984</td>
<td>0.984</td>
</tr>
<tr>
<td>s78</td>
<td>Ex100</td>
<td>0.961</td>
<td>0.968</td>
<td>0.966</td>
<td>0.967</td>
<td>0.965</td>
<td>0.965</td>
<td>0.965</td>
<td>0.966</td>
<td>0.967</td>
<td>0.967</td>
<td>0.968</td>
<td>0.969</td>
</tr>
<tr>
<td>s77</td>
<td>Ex100</td>
<td>0.985</td>
<td>0.983</td>
<td>0.984</td>
<td>0.985</td>
<td>0.986</td>
<td>0.986</td>
<td>0.986</td>
<td>0.986</td>
<td>0.987</td>
<td>0.987</td>
<td>0.987</td>
<td>0.988</td>
</tr>
</tbody>
</table>
Summary

- Learning, a large (30-10 min.), long term (4 trials), non-iterated task (N=40)
- Models novice to expert transition, validated
- Exploration of model knowledge levels
- Model Rapidly built (ie, 2 people x 4 hr.)
- Examines how to explain/explore such models
  - 6,198 initial rules and 10,239 learned rules
- Can use the learning theory to build better interfaces and tools for HCI/UCD

Future work

- Does not (yet) use Perceptual-Motor
  - Use PM to slow learning
- Examine more data aspects
  - Retention
  - Subtask learning, interact directly with task
  - Individual differences, errors
  - Duplicate tasks
Some References


Introduction to Cognitive Modeling:
How to run studies, which support testing models [20]

- Ideally, you use another study's data
- Not ideal, run study when
  - No other study
  - It is flawed
  - Your advisor or reviewer requires it
- As a modeler, you need to learn how to run studies so you can read and use them
- So, learn:
A Risk Driven Approach to Experimental Design and Practice [30]
Frank E. Ritter and Jonathan H. Morgan (slides)
The College of IST
Penn State
&
Jong W. Kim and Richard Carlson (book)
Psychology, U. of Central Florida, and Psychology, Penn State

Ritter, Kim, Morgan, & Carlson, 2013
acs.ist.psu.edu/reports/ritterKM09.pdf ++

Supported by ONR, Contracts N00014-11-1-0275, N00014-10-1-0401, & W911QY-07-01-0004
Overview

acs.ist.psu.edu/papers
acs.ist.psu.edu/reports/ritterKM09.pdf

www.frankritter.com/rbs/ rbs-handout-cogsci.pdf (TB, p. 3)

0900-0915 (0) Orientation
0915-0945 (1) An overview of risk-driven experimental design
0945-1015 (2) Preparation for running an experiment
1015-1040 break
1040-1115 (3) Ethical challenges in the experimental process
1115-1145 (4) Risks to validity, with class participation
1145-1200 Slack
1200-1215 (5) Conducting an experiment
1215-1230 (6) Concluding a study and reporting results,
Summary
Summary 1 of tutorial:
(Re)Looking at failure: What constitutes a failure/risk?

- Someone got hurt

- After committing significant resources, the study was never completed

- We have learned nothing new because our data is not repeatable or generalizable

- We have failed to communicate our results or their significance to anyone else
Sources of Failure?

Why did someone get hurt?

- We failed to do a risk assessment
- Being prepared for unanticipated problems
- We failed to screen participants properly
- We failed to either develop or follow procedures, either experimental procedures or data management procedures
- We did not anticipate or mitigate situational risks either in our experimental setting or outside of it that hurt our participants
- We ignored additional insights we could have learned from the participants through observation or debriefing
- Others?
Sources of Failure?

- Why we were unable to complete the study?
  - We were overly ambitious, perhaps because we failed to fit the research question or methods to the problem at hand
  - We ran out of time
  - We ran out of resources or lacked them in the first place
  - We lacked the people, either participants or staff, or trained staff
    (experiments appear to have less risk than modeling)
Sources of Failure?

- Why we were unable to reproduce our results or generalize them?
  - We failed to use the same experimental procedures or test under the same conditions for each S
  - We failed to achieve an adequate sample size or sufficient degree of representativeness in our sample
  - Our task fidelity was poor. We failed to construct an experimental task that was analogous with respect to its key points.
Sources of Failure?

- Why have we been unable to report our results or communicate their significance?
  - We failed to properly catalog or backup our data
  - We failed to write as we went. We no longer remember some of the critical, early details.
  - We made poor data analysis or display choices
  - We failed to identify a venue early, or understand who we should consider our audience
How do we avoid failure?

- We recognize that running a study is an incremental risk-driven process, similar in some respects to spiral development of systems (Boehm & Hansen, 2001; Pew & Mavor, 2007)

- To be successful, we need to:
  - Formulate a research question that meets our research goals
  - Have a theory of transfer effects that minimizes risks associated with confounding variables, and enables us to conserve time and resources
  - Pilot studies and study components
  - Be candid in our risk assessments and be willing to adapt and refine
What to get out of this Tutorial

1) Some feeling for how to run a study
   ➤ Cognitive science may be modeling + data
      So, to use data you have to know how it was gathered
   ➤ Modeling is slow, so data publication helps modelers
   ➤ If you are a computer scientist, you won’t have taste in this area
      => Help you develop a green thumb
   ➤ Not how to design a study, but related

2) Some tools to help you set up a study

3) Materials
   Book and report on this topic (please let me know if you use it for a class)
   Handout (available online)
   Example problems

5) A greater appreciation for mistakes to avoid and a theory of how to avoid them
Experimental Process Overview, linear

(TB, p. 11)

An iterative, and often overlapping process

- Identify research problem and priorities, design experiment
  
- Prepare IRB forms
  - Fill out IRB forms (local to institution): Note risks and how to address harms
  - Consent form
  - Debriefing form

- Develop the experiment environment
  
- Run pilot study

- Analyze pilot study data

- Prepare experimental script

- Advertise the experiment, recruit subjects

- Explain the experiment to participants (e.g., purpose, risk, benefits)

- Run the experiment

- Debrief subject and wrap up session

- Gather, analyze, and store data

- Report results
Summary: Lessons so Far

- More steps than I thought
- Iterative and risk-driven (if you pay attention)
- A process but not a set process
- Studies will overlap each other and inspire each other
- It is useful to have the RAs/Es pay attention
  - Ss suddenly ‘get it’
  - Ss don’t get some aspect
  - Ss comments
  - Ss ‘cheat’ somehow
Preparation for an experiment

(TB, p. 14)

Experiments are driven by their questions and shaped by the methods available to explore those questions and existing results/lessons in that area. This contributes to doing multidisciplinary work.
Summary: Piloting

- Write out method
- Used to check your work
- Use a script,
  Step 1, start program, Step 2 “Welcome to…”
- Start local, e.g., YOU, and then officemate, and then move further and further away
- Mount a scratch monkey
- Check your apparatus and data gathering and use of data
- Consider/reconsider, number of Ss to run
  - Previous studies
  - Power analyses (\text{\textcopyright}Cohen for Ss; \text{\textcopyright}Ritter et al. for models)
  - Why not prefer large effects?
Ethical Challenges Associated with the Experimental Process

Ethical problems can be decreased by deliberate proactive action.

A couple of bad examples and then a general view
A HCI Study Gone Wrong (circa 2008)

- No informed consent
- No privacy grantees or data management plan
- "You have no friends."
  Yes, a student researcher felt compelled to inform a participant and the S's teachers and Dean of this fact.
- Even "HCI" studies can hurt people
- Know your methods, protect Ss
Ethical Challenges Associated with the Experimental Process

Ethical problems can be decreased by deliberate proactive action.
Summary:
How to avoid ethical problems

- Recruit fairly
- Look out for your Ss
- Anonymise data at the beginning of each session by using subject IDs, not names
- Have a plan for surprising data (e.g., high BP)
- Communicate early and relatively often about publication plans and data ownership (Diguisto, 1994)
- Some argue that you have an obligation to use the data you gather
Challenges to Validity: Constraints on your study

Or: alternative hypothesis for results (TB, p. 21)

Challenges to validity can be anticipated and mitigated.
Conducting an Experiment (TB, p. 24-25)

Success in execution is directly correlated to careful preparation.
Summary: Running a session

- Use of piloting means no surprises (except for the data!)
- Script keeps treatment the same, it includes session set up
- Keep eyes open while running for further insights
- Anonymise data as soon as possible
Concluding an Experiment and Reporting Your Results (TB, p.27)

Concluding a Study and Relaying Results

Data Care & Backup

§§ 6.1 & 3.6

Analyzing Data & Reporting Results

Documenting Data Analyses

§§ 6.2.1 & 2.2.4

Using Descriptive & Inferential Statistics

§ 6.2.2

Planned vs. Exploratory Data Analysis

§ 6.2.3

Displaying Data

§ 6.2.4

Communicating Your Results

§§ 6.3

Kept raw data as a backup

Record all data transformations

Try numerous measures

Think about what you are aggregating

Don't be afraid to do additional analyses

Explore graphing your data

Consider your writing outlet

Debrief, debrief, debrief!
Summary: Concluding an Experiment and Reporting Your Results

- Concluding a session
  - Finish with the subject (thank, debrief, check paperwork)
  - Check the data was collected and saved
  - Comment on the data if anomalies
- Data care, security and privacy
  - Anonymizing removes nearly all ills
- Back up data (daily, weekly)
- Data analysis
  - Not how, but note how (document and keep track of)
  - Know your data if you are the RA that analyses
  - Save the analyses, time is not important, space is not important, the insights and results are important
  - Aside: we prefer regression
  - Aside: we prefer individual analyse
Ch 6.5 Communicating your results

- Start with a target in mind (if you can)
- Work to larger publications (workshop, conf, journal, book)
- Rewrite, rewrite, rewrite (the book was draft #53 turned in, revised twice in pageproofs)
Ch. 7 Afterward

- Appropriate behavior with subjects
- Insights
- Repeatability
- Reportability
Summary 2 of Tutorial

- There are steps to running a study separate from design and analysis
- These are practical, hands-on, implicit knowledge
- They are informed by previous studies
- To be successful, we need to:
  - Formulate a research question that meets our research goals
  - Pilot studies and study components
  - Be candid in our risk assessments and be willing to adapt and refine
  - Be aware of alternative hypotheses, and avoid what we can and control what we cannot avoid
  - Plan for reporting results early
If you will teach this....

- Full book available from Sage & Sage online
- Slides available as ppt or pdf
- Workbook available as pdf
References


