Presented at the FIERCES School, April 2016, MePHI, Moscow.

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

Frank E. Ritter, College of IST, Penn State
frank.ritter @ psu.edu

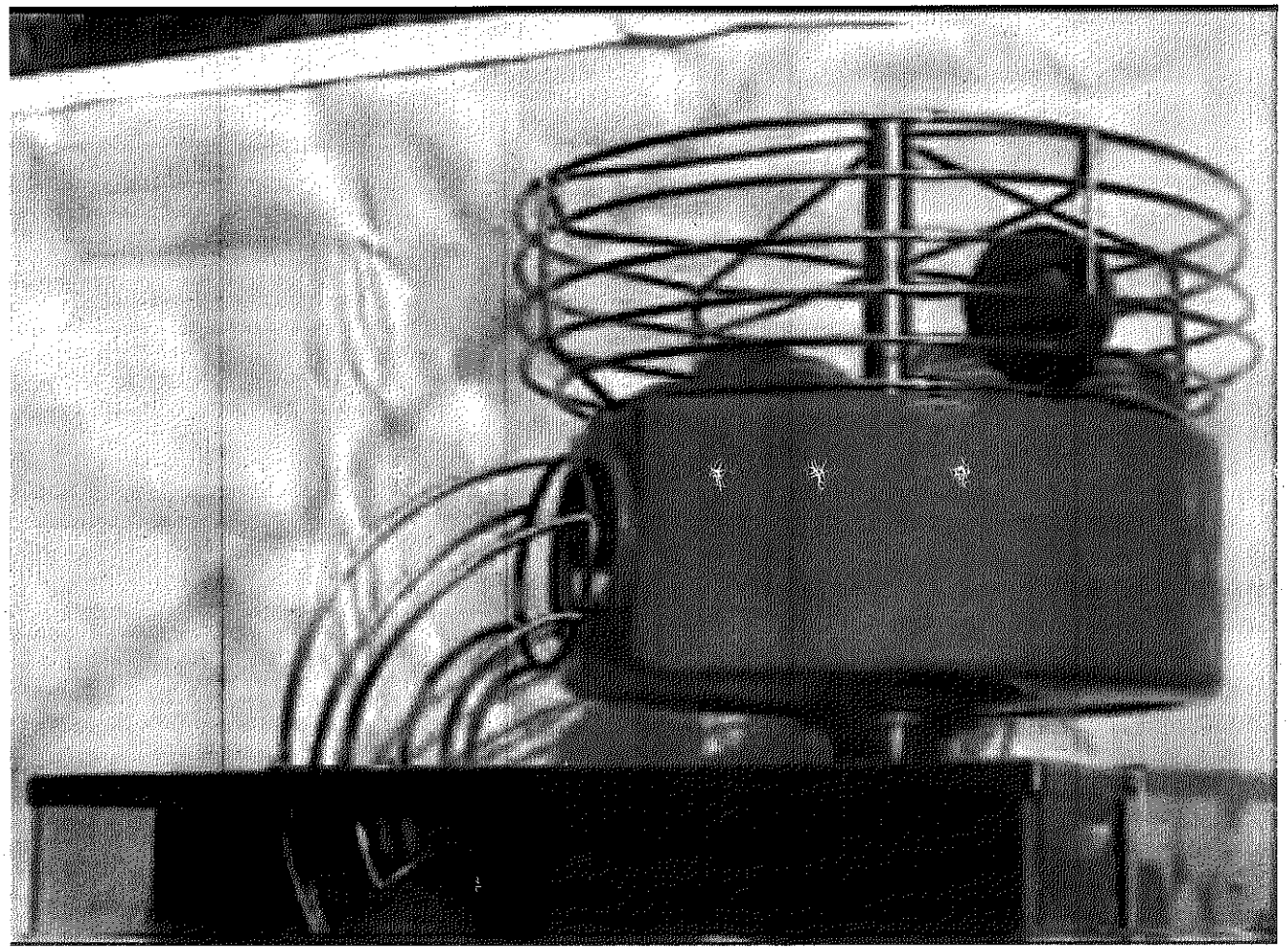
21apr16

📖 papers on line

Why cognitive architectures?

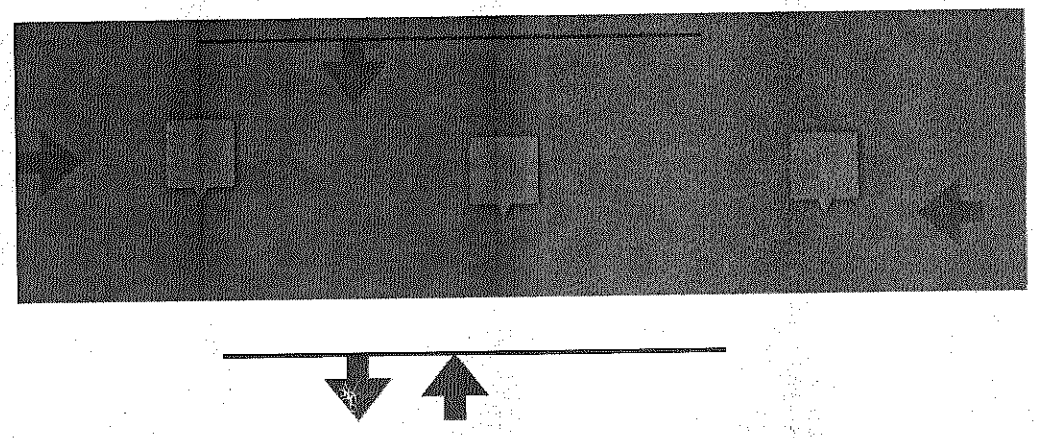
- Make computers like people except they will be my friend, Rob Ward, ~1990
- To create model users (CM&N, 1983; Booher & Minniger, 2003)
- To create model opponents and colleagues (video games, simulations)
- To provide a path towards reuse and cumulation
- To create a unified theory in psychology (UTC) (Newell, 1990) (memory, Fitts law, Hick's law, reading, mental models, ...)
- To understand the mind

Simple mechanism



More complex mechanism: Place the cameras

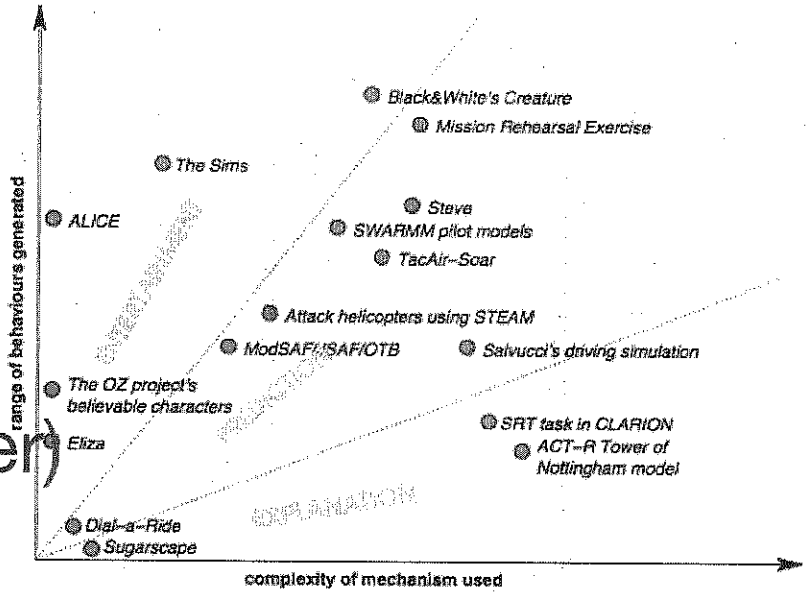
More complex mechanism



- Rapid behavior
- Need to record it, too fast to analyze
- Not commonsense
- Complex

Range of ways of simulating cognition

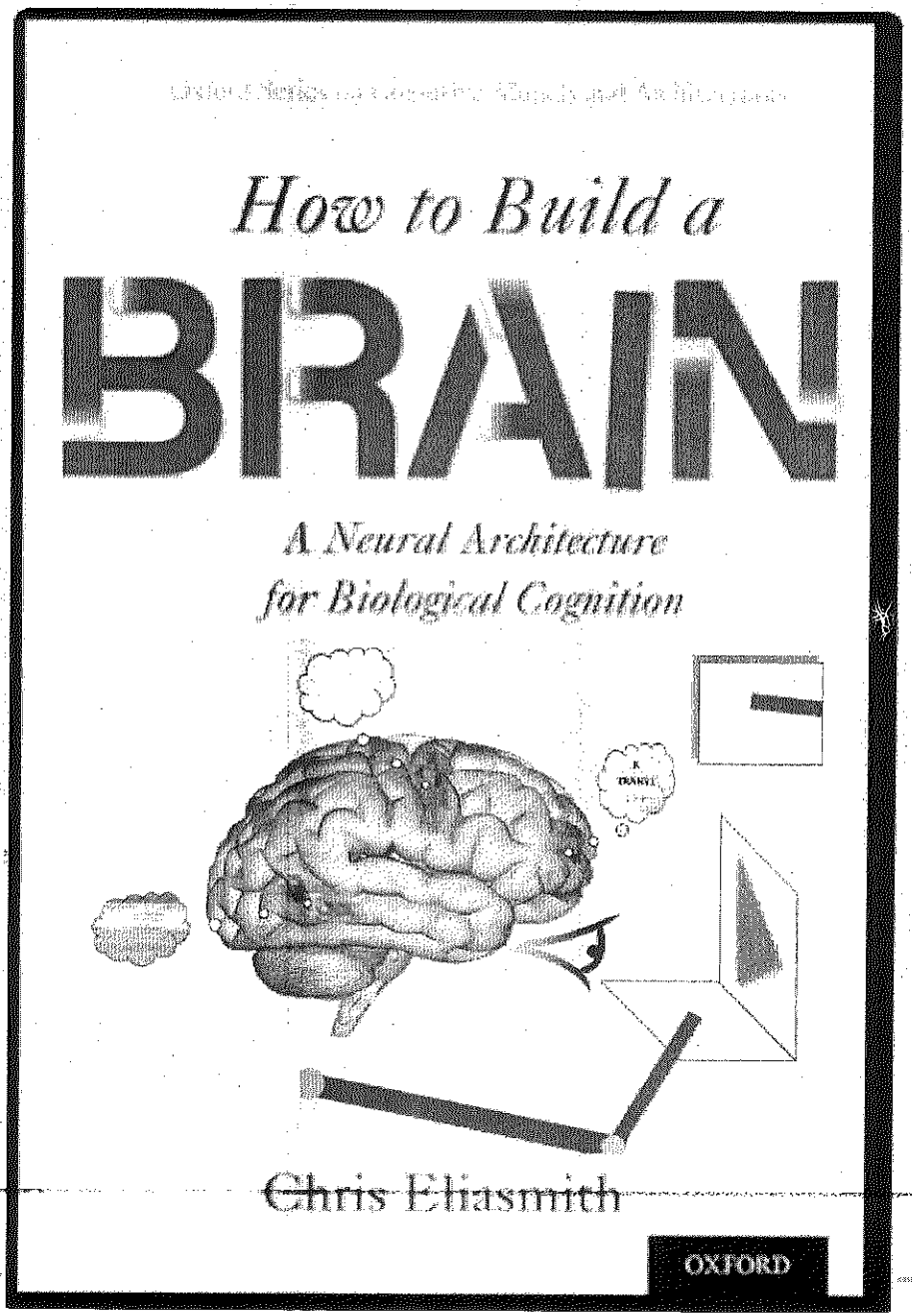
- Eliza (if-then rules, applier)
- Sims (objects, tasks-object pairs, applier)
- Lisp, Java (code, interpreter)
  - Fortran, assembler (code, compiler, hardware)
- AI planning languages (plans, constraints, planner)
- Expert systems (rules, interpreter)



Norling, 2009, thesis

.....and more....

Yet more complex mechanism

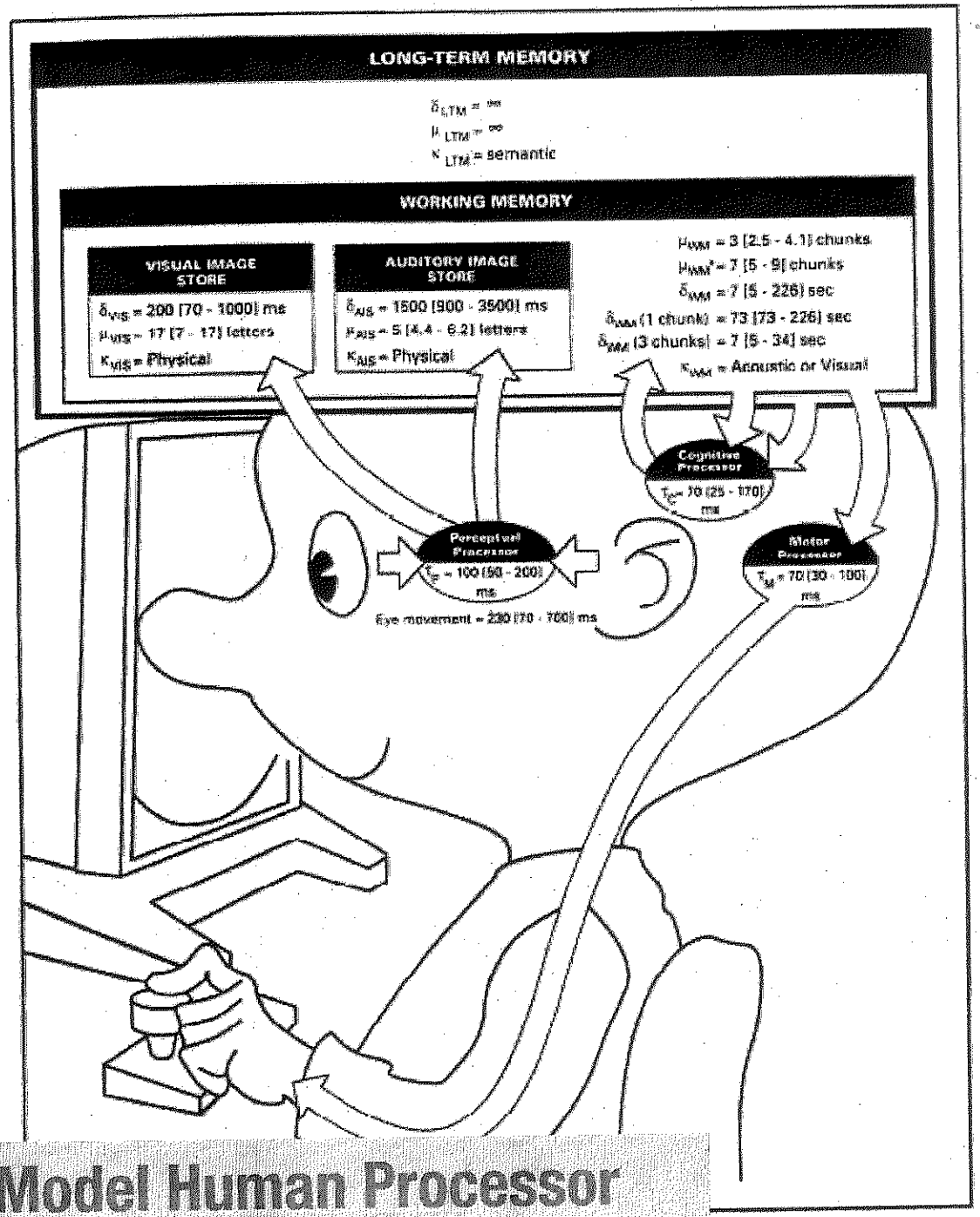


What is a Cognitive Architecture?

- 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)

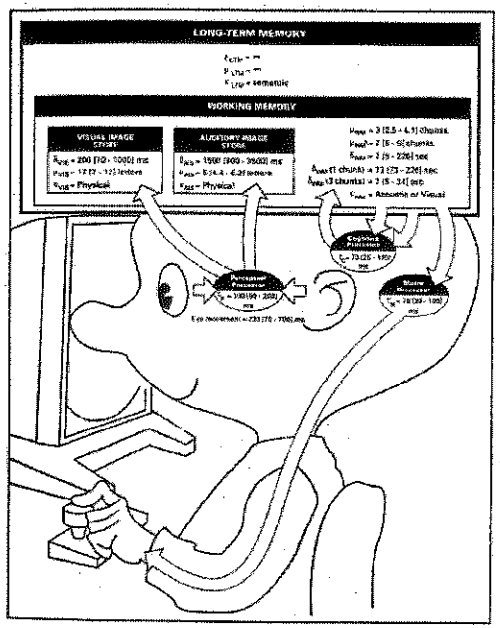


Model Human Processor

Card, Moran, & Newell, 1983

Simple architecture: The MHP is useful for HCI and design

Keystroke-Level Model



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

$$\Sigma T_{\text{Mental OPs}} = \sim 1.35 \text{ s (fixed, but varies)}$$

$T_{\text{Keystroke}} = \sim 0.3$ s (also IDs)

$T_{\text{Mouse}} = \sim 1.1$ s (also Fitts)

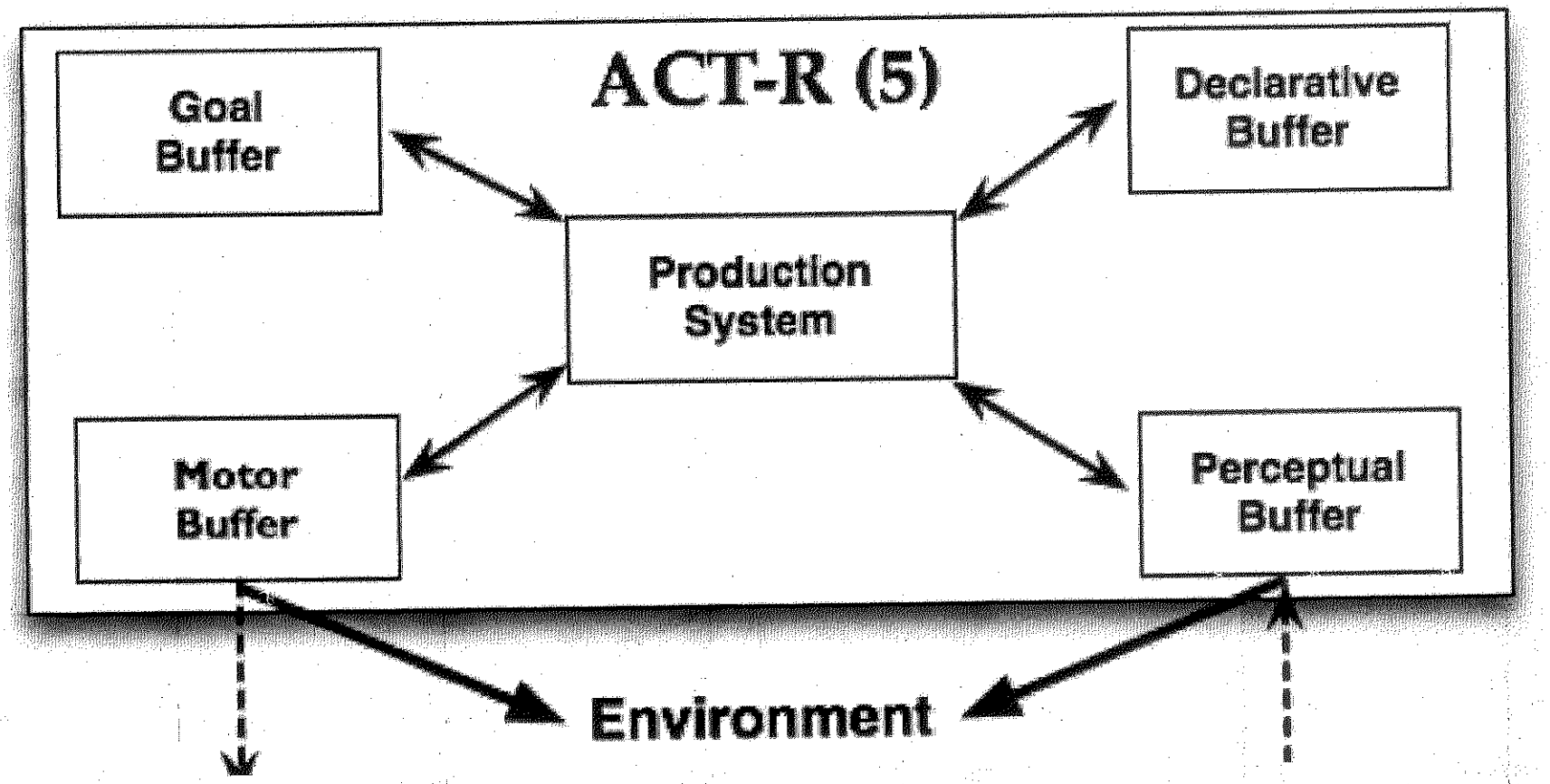
$T_{\text{Home}} = \sim 0.4$ s

$T_{\text{Sys}} = \sim 0.0$ s

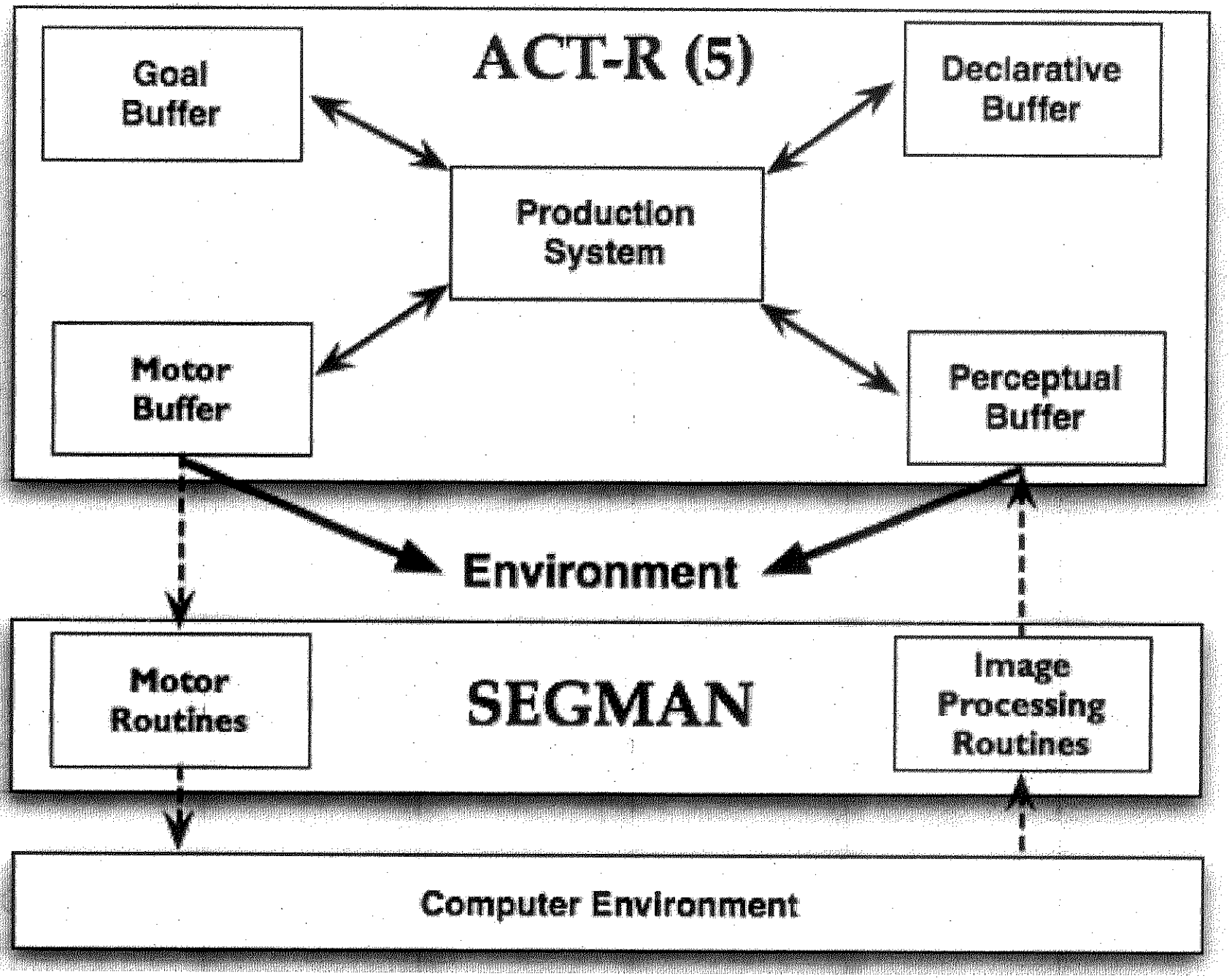
....a few more operators

(Card, Moran, & Newell, 1983)

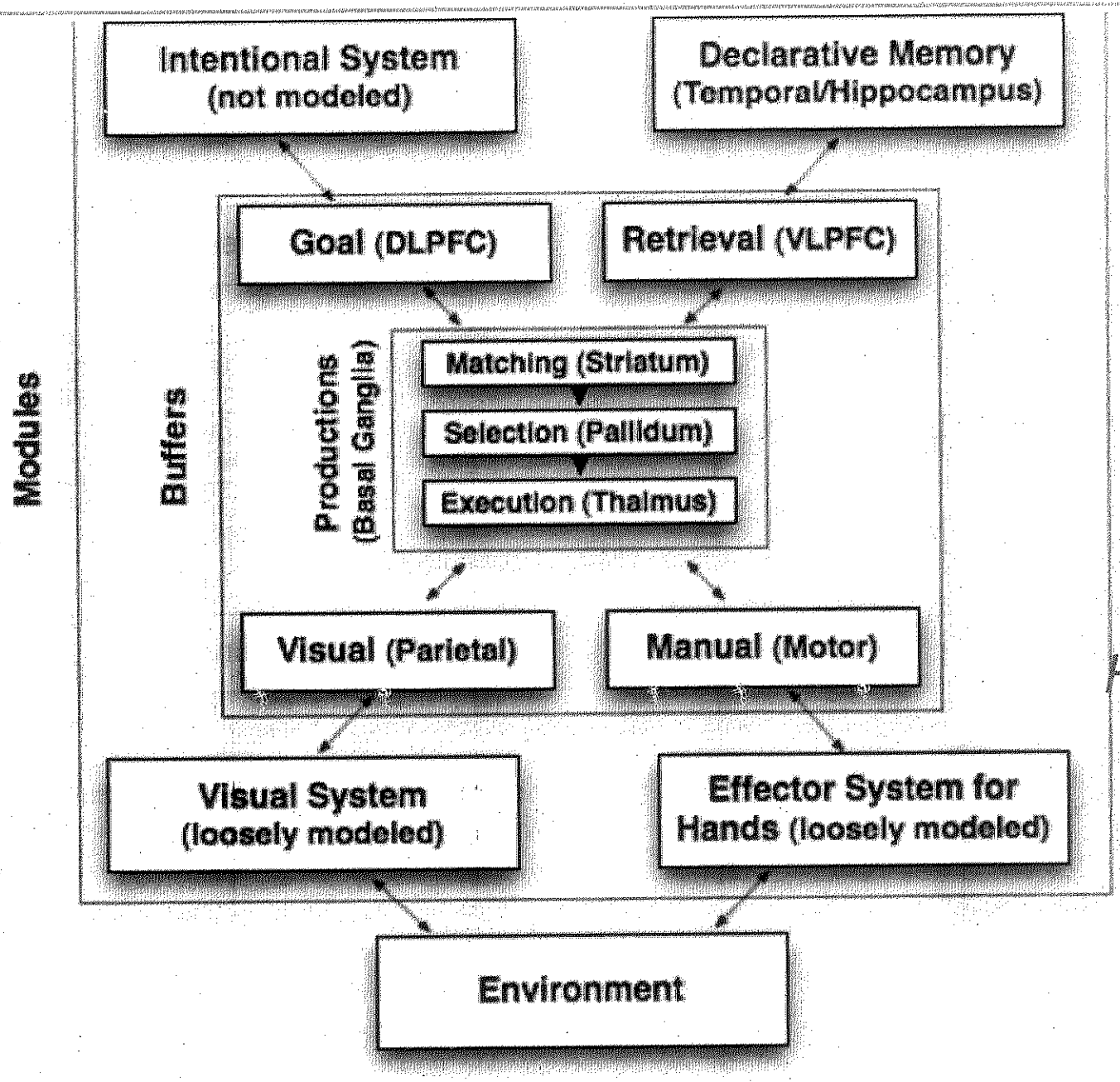
ACT-R



STARTS TO SHOW HOW MIND WORKS, HOW PARTS COME TOGETHER, WITH: LEARNING, ERRORS, MULTI-TASKS

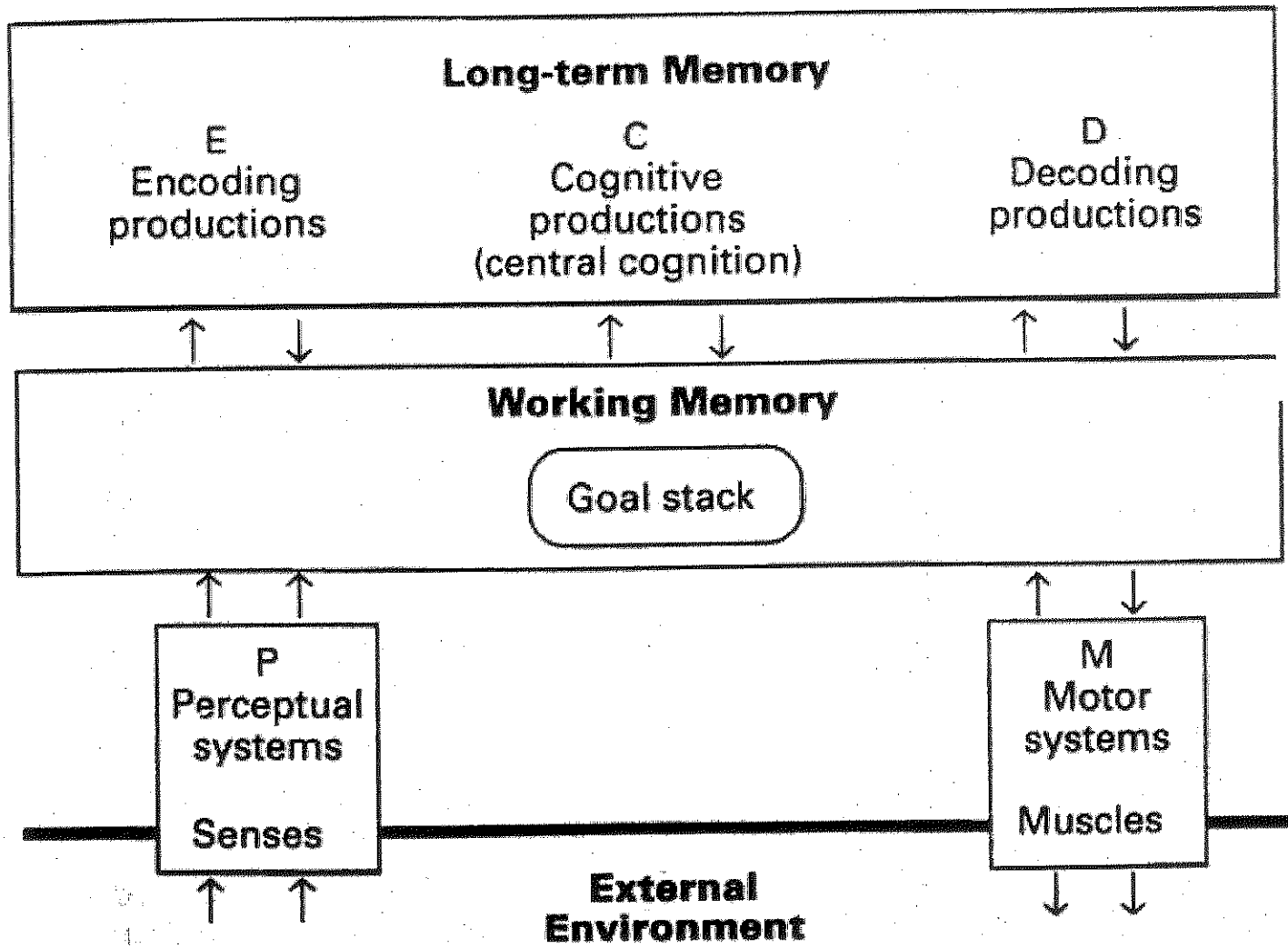


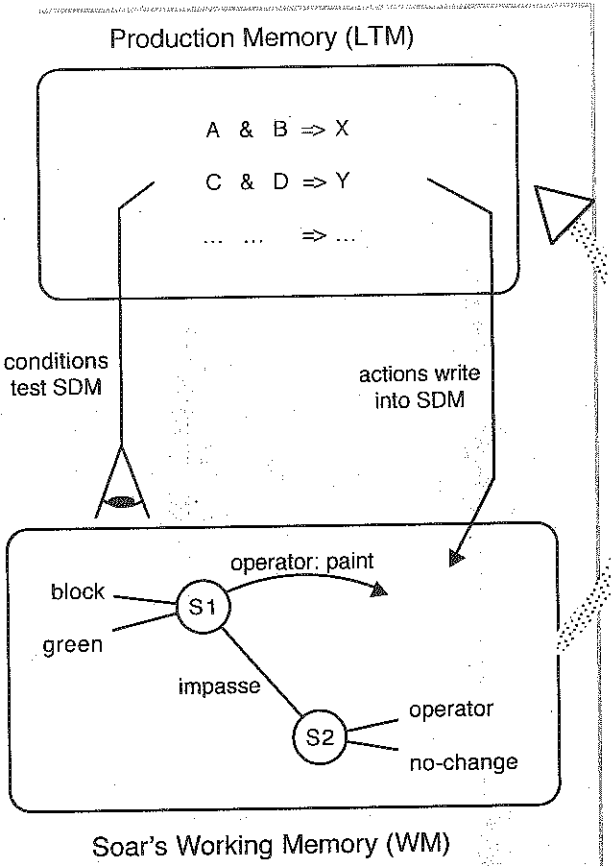
- Work with St. Amant, can be used to test designs



HELPS EXPLAIN
HOW THE MIND CAN EXIST IN THE PHYSICAL UNIVERSE

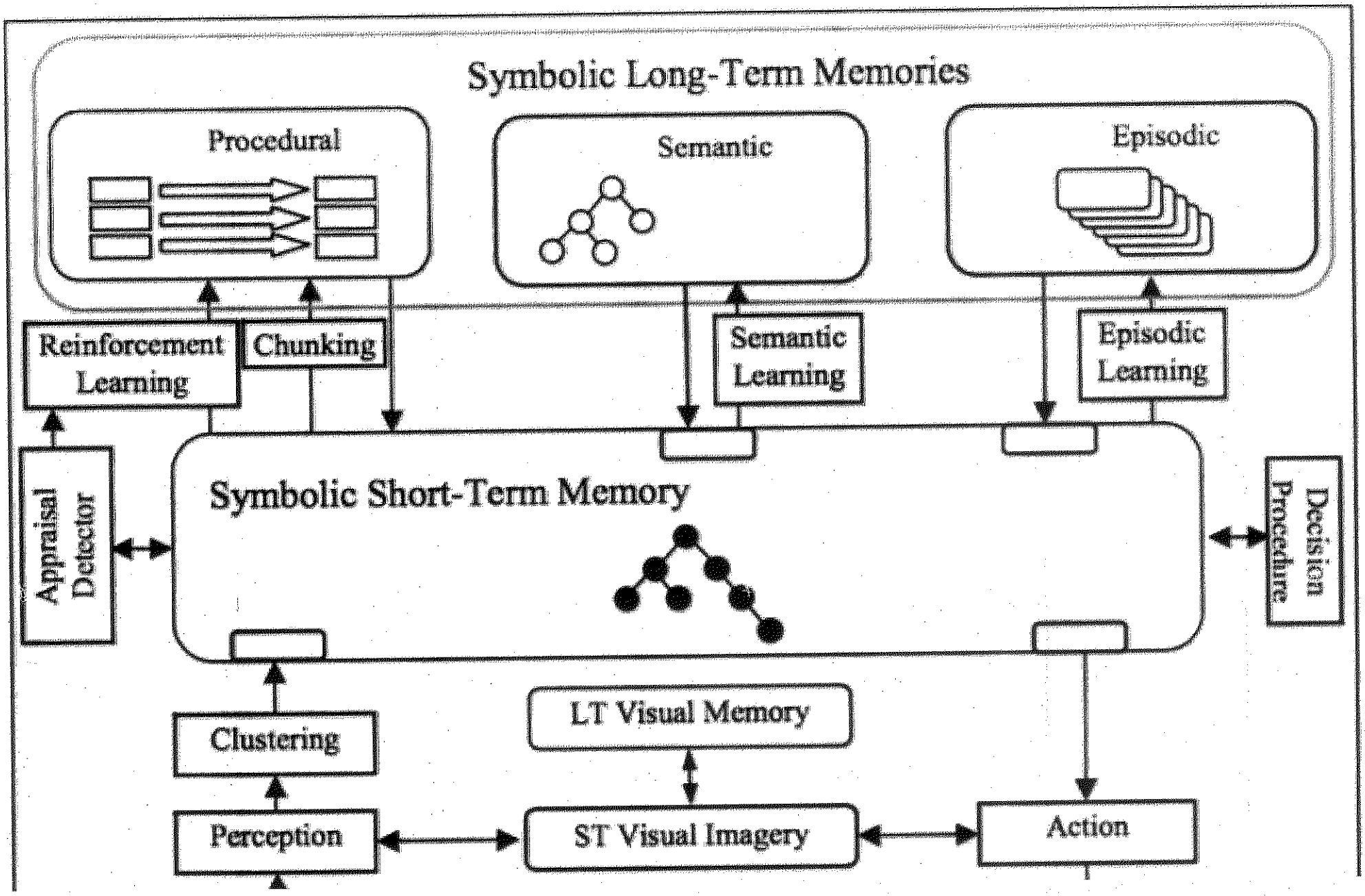
ACT-R 6: Anderson, 2007



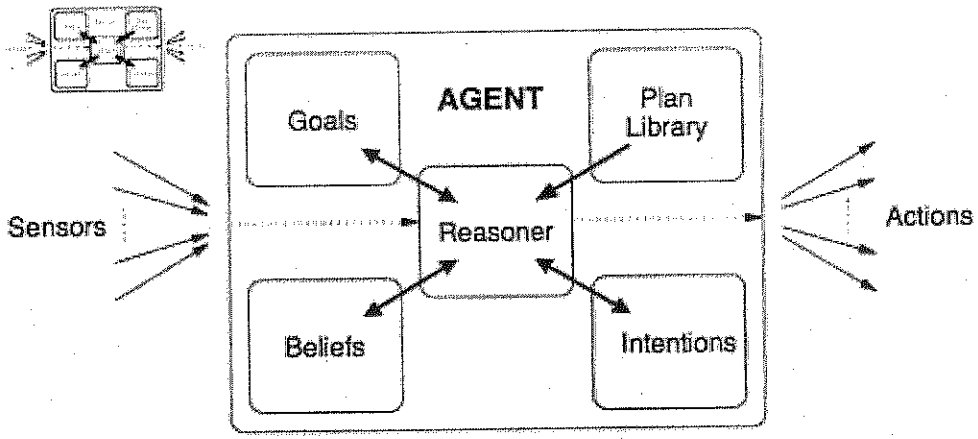


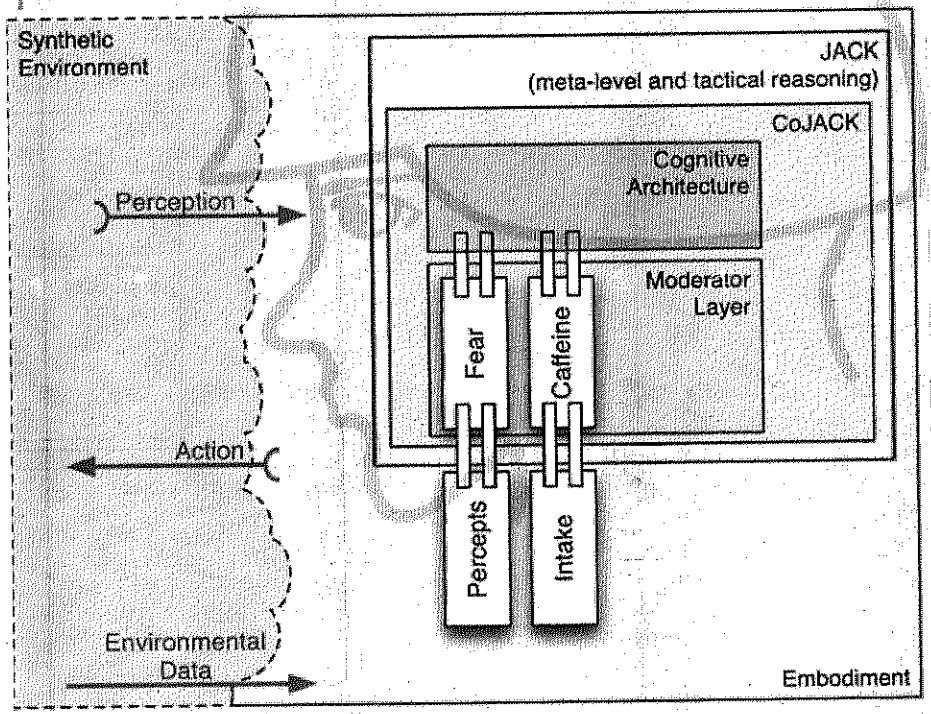
Soar: (Newell, 1990, Laird, 2012)

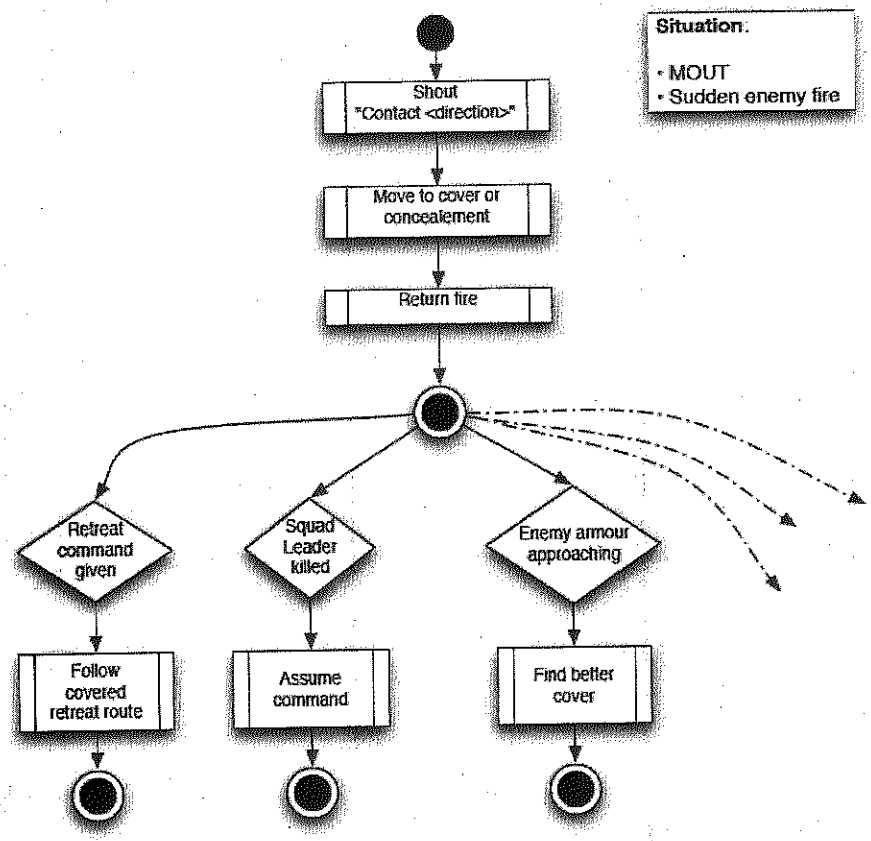
- Rule-based
- sorted by PS
- Decision cycle
- Timing from decision cycle
- Subgoaling with
- Chunking


Symbolic Long-Term Memories
Procedural
Semantic
Episodic
Reinforcement Learning
Chunking
Semantic Learning
Episodic Learning
Appraisal Detector
Symbolic Short-Term Memory
Decision Procedure
Clustering
LT Visual Memory
Perception
ST Visual Imagery
Action

(Laird, 2008)



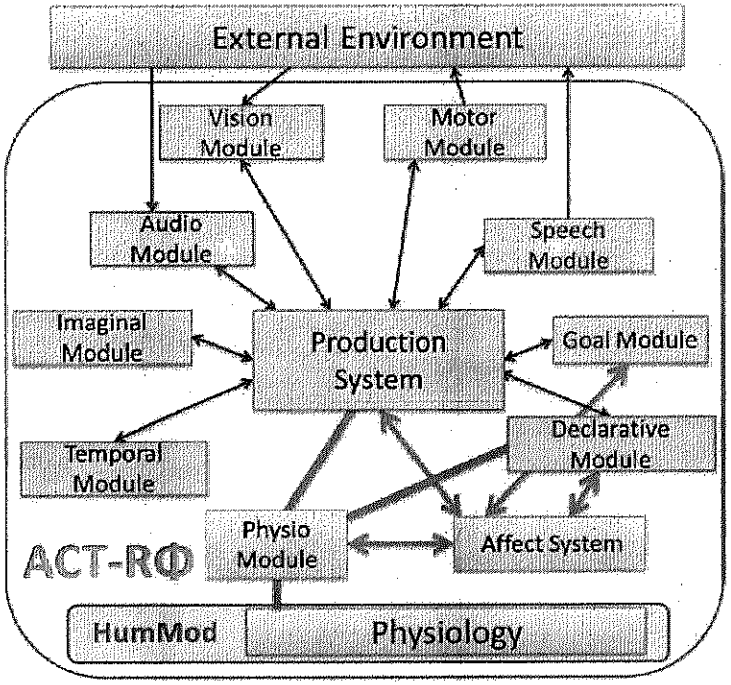




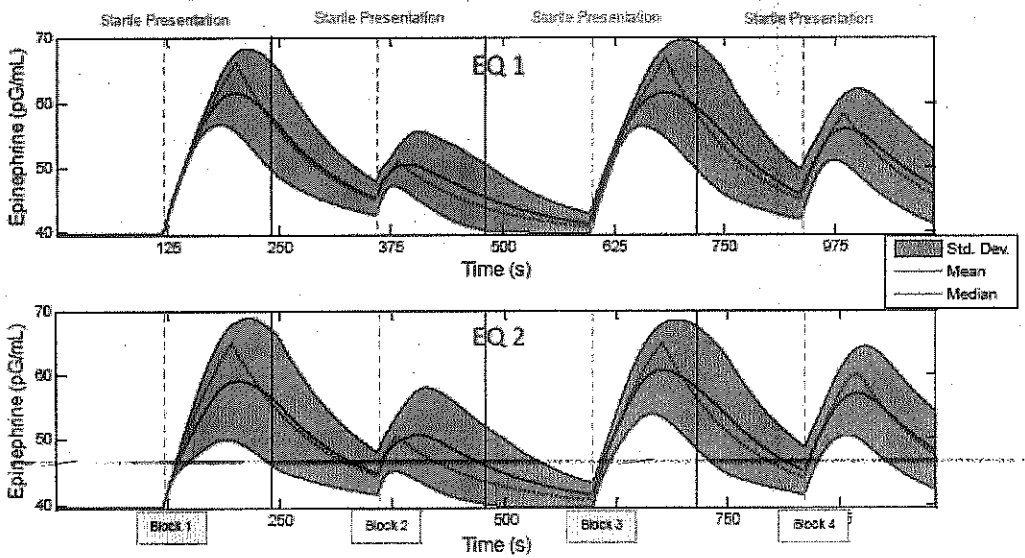
BDI: CoJack (Ritter et al., 2012)

ACT-R/Φ

📖(Dancy, Ritter, Berry, Klein, 2015)



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



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

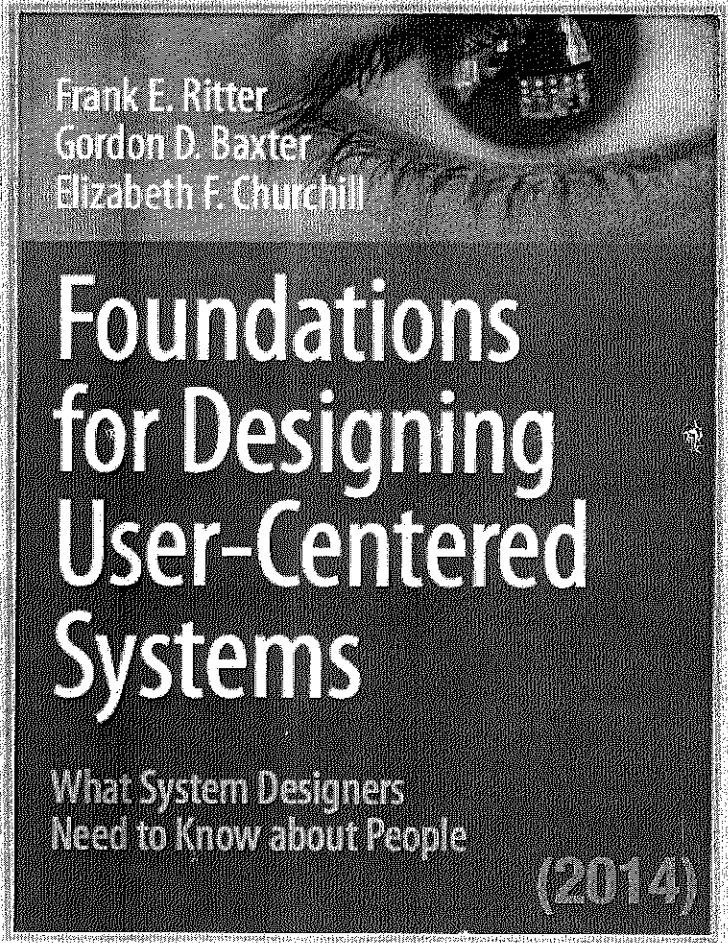
- Common components:
  - Input and output
  - Memory(s)
  - Knowledge
- Based on different perspectives, memory, attention, learning, knowledge representation, hierarchy, hybrid
- None are complete, many are useful

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

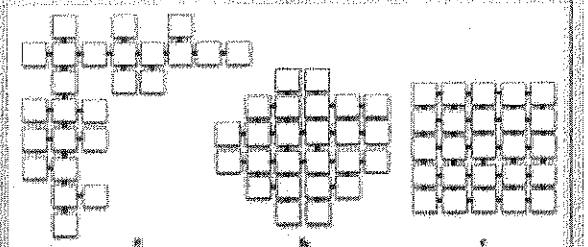
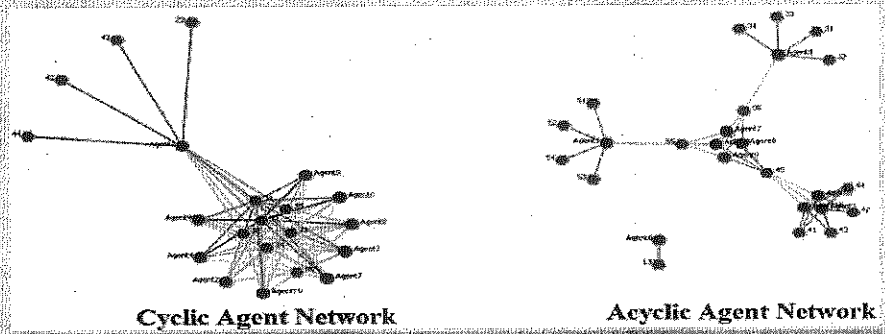
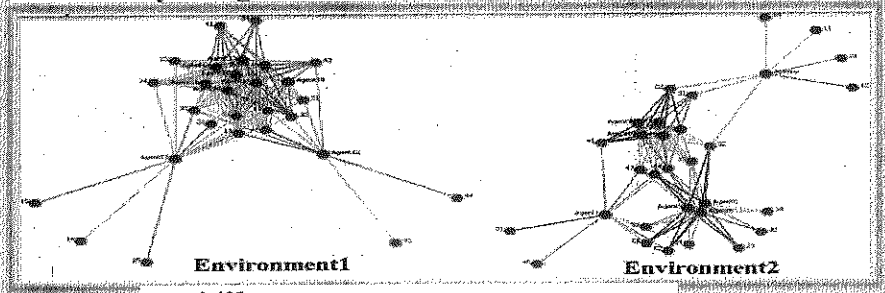


Figure 1. Maps (hallway, central, full grid)

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





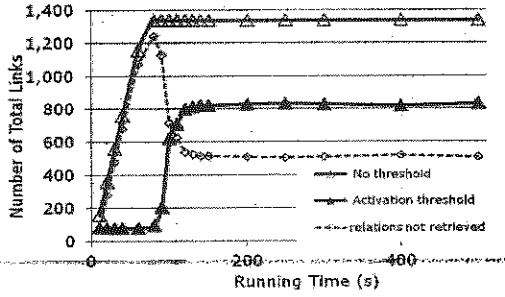
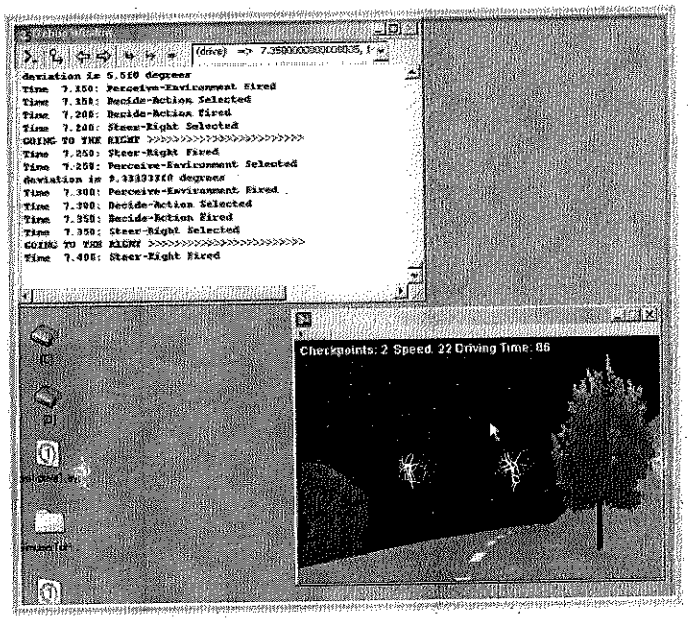


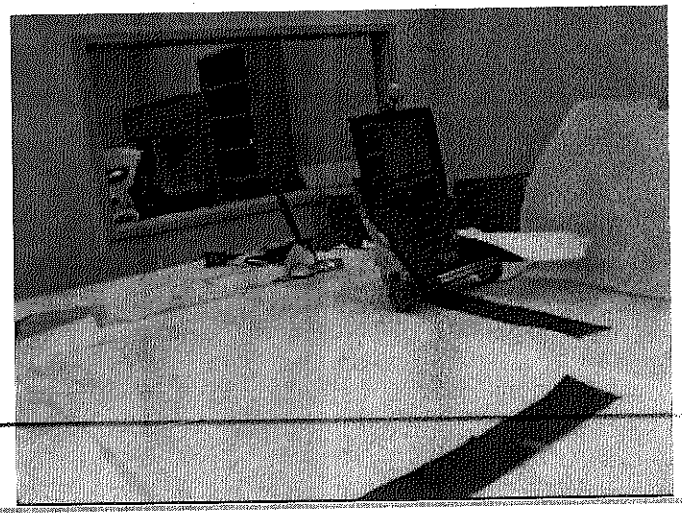
Figure 2. The effect of memory threshold on network formation over time for the fixed path

[📖 Zhao, Cw., Hiam, J. W., Morgan, J. H., Ritter, F. E. (2011). A multi-strategy spatial navigation model in a text-based environment. In *Proceedings of the 20th Conference on Behavior Representation in Modeling and Simulation*. 251-258. 11-BRIMS-036.] (submitted)

Interaction: Bitmap based interaction

- Provides models access to simulations and thus knowledge



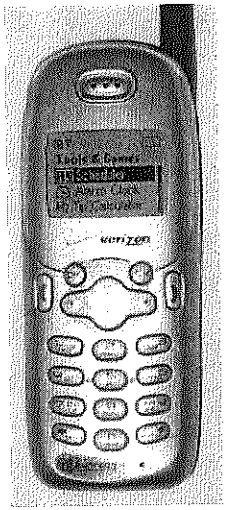


Ritter, F. E., Baxter, G. D., Jones, G., & Young, R. M. (2000). Supporting cognitive models as users. ACM Transactions on Computer-Human Interaction, 7(2), 141-173

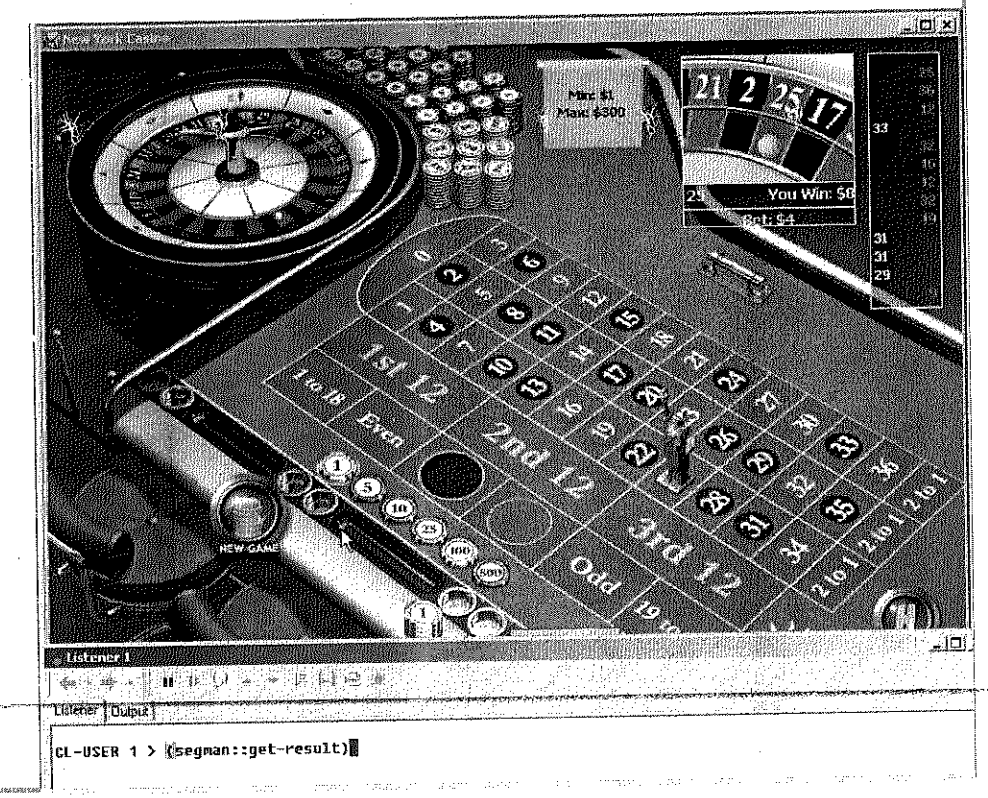
pb270194.mov

Ritter, F. E., Kukreja, U., & St. Amant, R. (2007). Including a model of visual processing with a cognitive architecture to model a simple teleoperation task. Journal of Cognitive Engineering and Decision Making, 1(2), 121-147.

App: User modeling of Interaction
SegMan



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



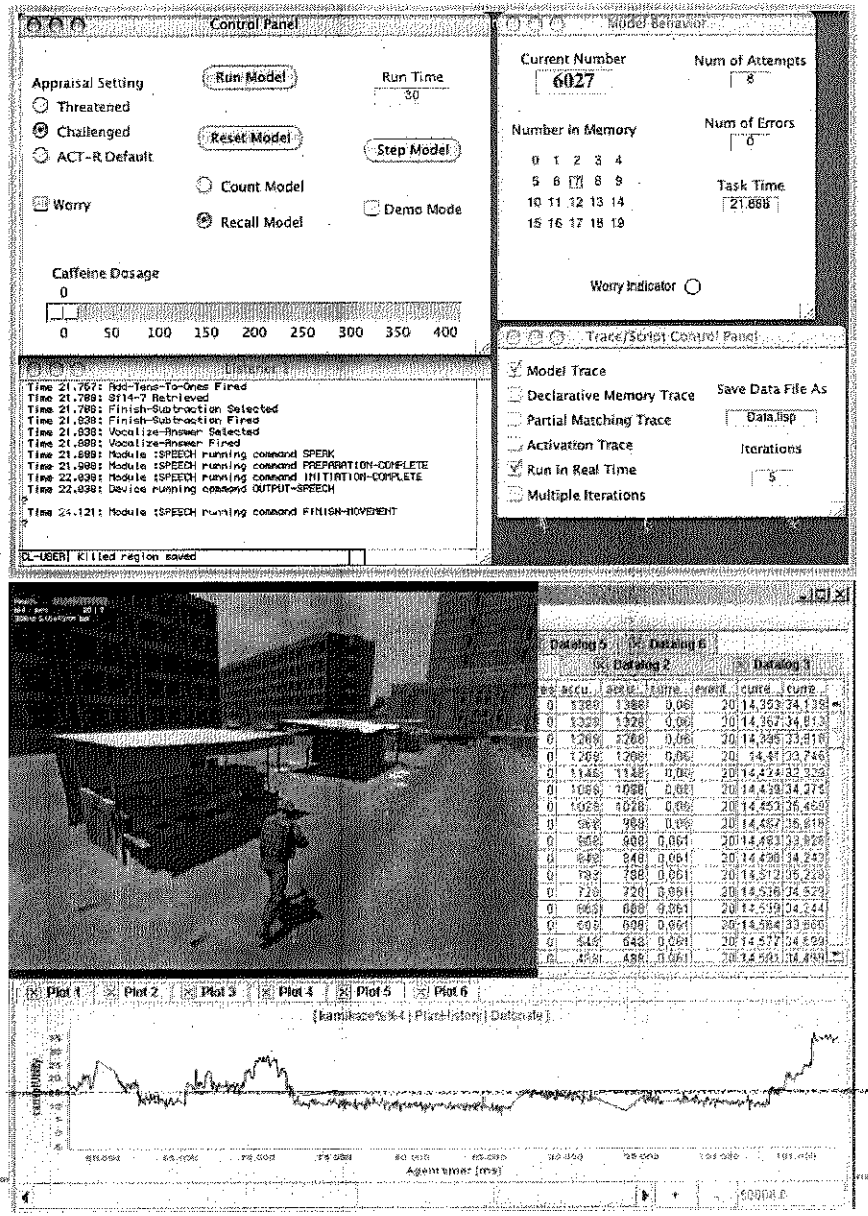
St. Amant, R., Horton, T. E., & Ritter, F. E. (2007). Model-based evaluation of expert cell phone menu interaction. ACM Transactions on Computer-Human Interaction, 14(1), 24 pages.

pb270194.mov

reifers-demo804.mov

App: Moderators Challenge and threatened

- Models of challenged and threatened behavior



Ritter, F. E., Reifers, A. L., Klein, L. C., & Schoelles, M. J. (2007). Lessons from defining theories of stress for architectures. In W. Gray (Ed.), Integrated models of cognitive systems (pp. 254-262). New York, NY: Oxford University Press. challenge-no-worry5nov04.mov

Evertsz, R., Pedrotti, M., Busetta, P., Acar, H., & Ritter, F. E. (2009). Populating VBS2 with realistic virtual actors. In Proceedings of the 18th Conference on Behavior Representation in Modeling and Simulation, 09-BRIMS-04. civilian_fear.mov

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 resuse this? [dunno and no and needs more work]

$$p_a = \frac{1}{1 + \left(\frac{1}{t_d} * e^{\frac{1*\left((d_{friend} * \sqrt{g_{size}}) + \frac{d_{leader}}{1+k} \right)}{g_{size} * d_{observer} * (c+k)}} \right)}$$

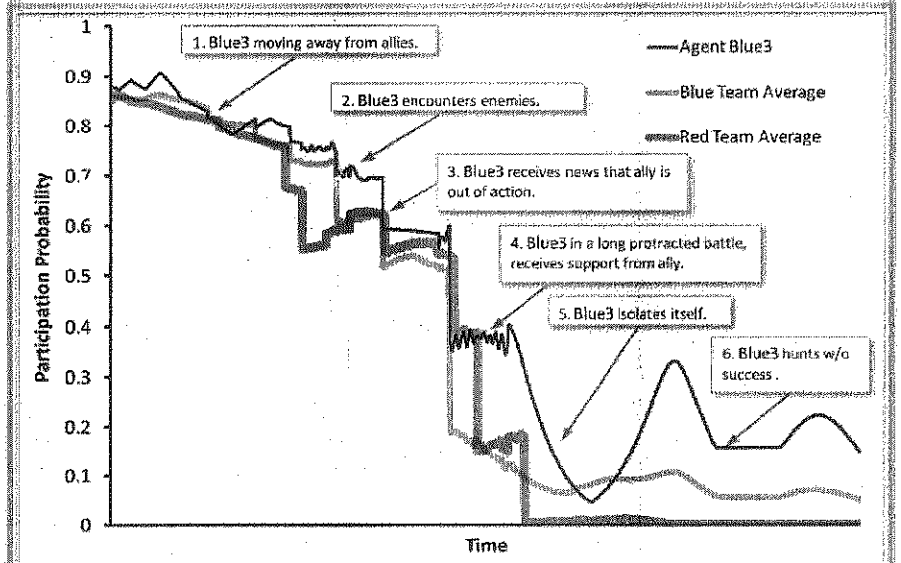


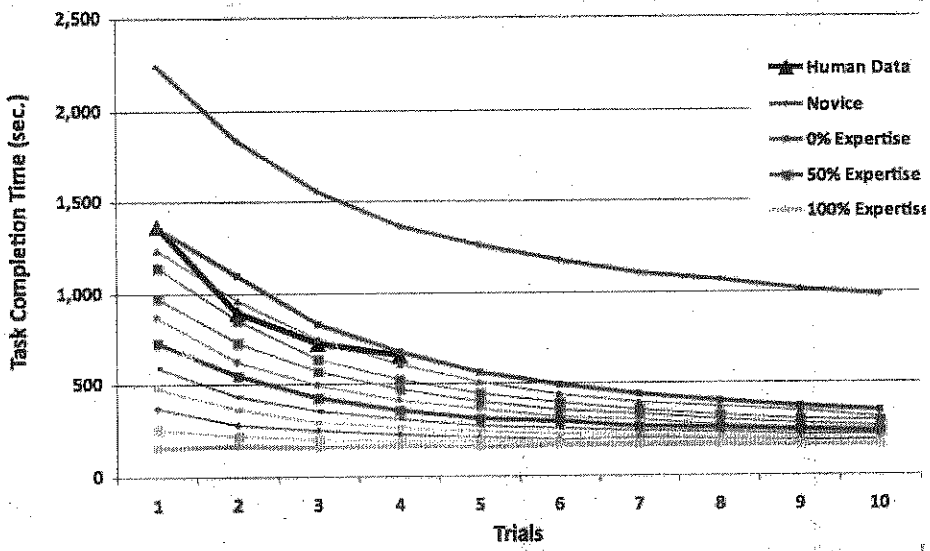
Figure 4: Trace of agent's participation values in an 8x8 battle.

[📖 Morgan, J. H., Morgan, G., & Ritter, F. E. (2010). A preliminary model of participation for small groups. *Computational and Mathematical Organization Science, 16,* 246-270.]

Grossman, D. (1996). *on killing: The psychological cost of learning to kill in war and society.* New York, NY: Back Bay Books, Little Brown.

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



[📖 Ritter, F. E., Haynes, S. R., Cohen, M. A., Howes, A. John, B. Best, B., Lebiere, C., Jones, R. M., Crossman, J. Lewis, R. L., St. Amant, R., McBride, S. P., Urbas, L. Leuchter, S., Vera, A. (2006). High-level behavior representation languages revisited. In *Proceedings of ICCM - 2006- Seventh International Conference on Cognitive Modeling*, 404-407. Edizioni Goliardiche: Trieste, Italy.

[📖 Paik, J., Kim, J. W., Ritter, F. E., Morgan, J. H., Haynes, S. R., & Cohen, M. A. (2010). Building large learning models with Herbal. In D. D. Salvucci & G. Gunzelmann (Eds.), Proceedings of ICCM - 2010- Tenth International Conference on Cognitive Modeling (pp. 187-191).

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
  West, R. L., & Lebiere, C. (2001). Simple games as dynamic, coupled systems: Randomness and other emergent properties. Cognitive Systems Research, 1(4), 221-239.

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

* $US 10 for 50

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

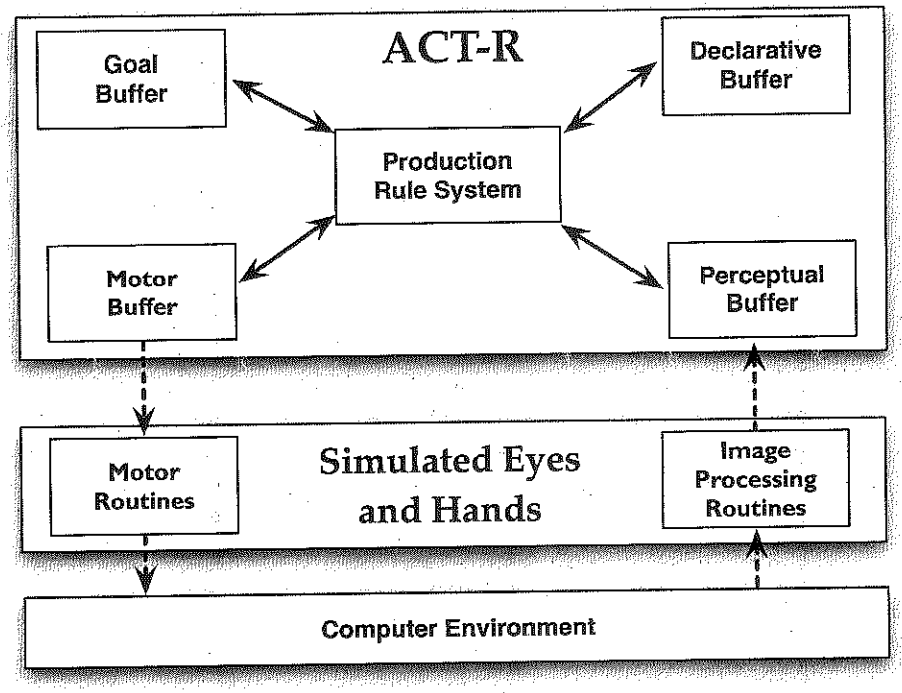
Implementing Large User Models with Multiple Levels of Expertise

- ACT-R compiler
- Data
- comparison

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

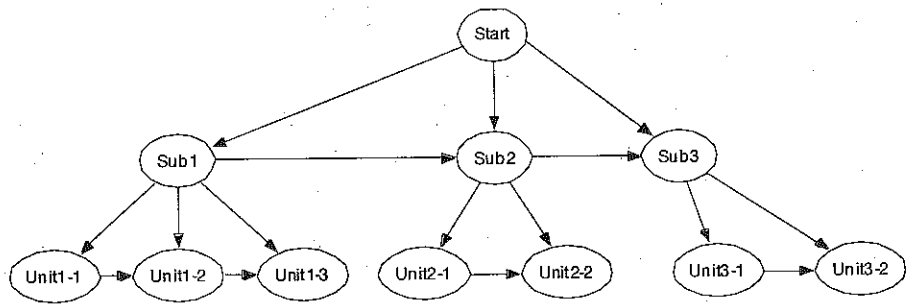
If this is in DM & this2 in DM
--> put this this+that in goal

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

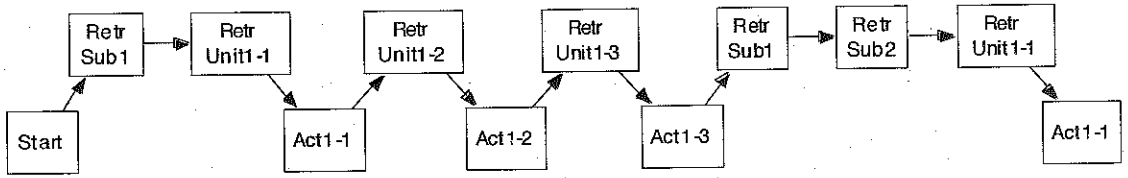


Paik, Kim, Ritter, & Reitter (2015)

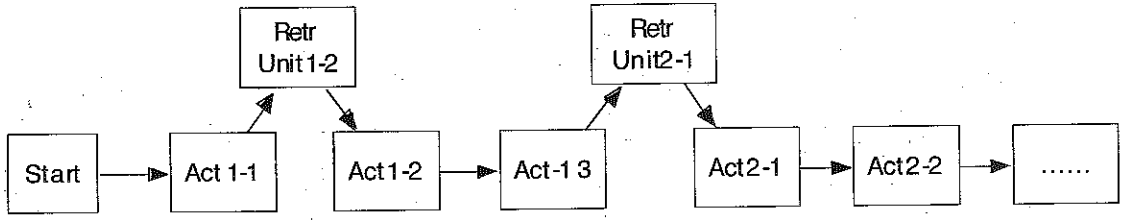
How the compiler works



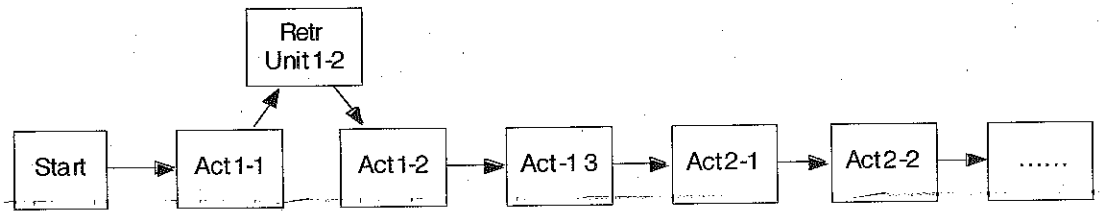
Declarative Memory HTA



Novice trace



Middle expert

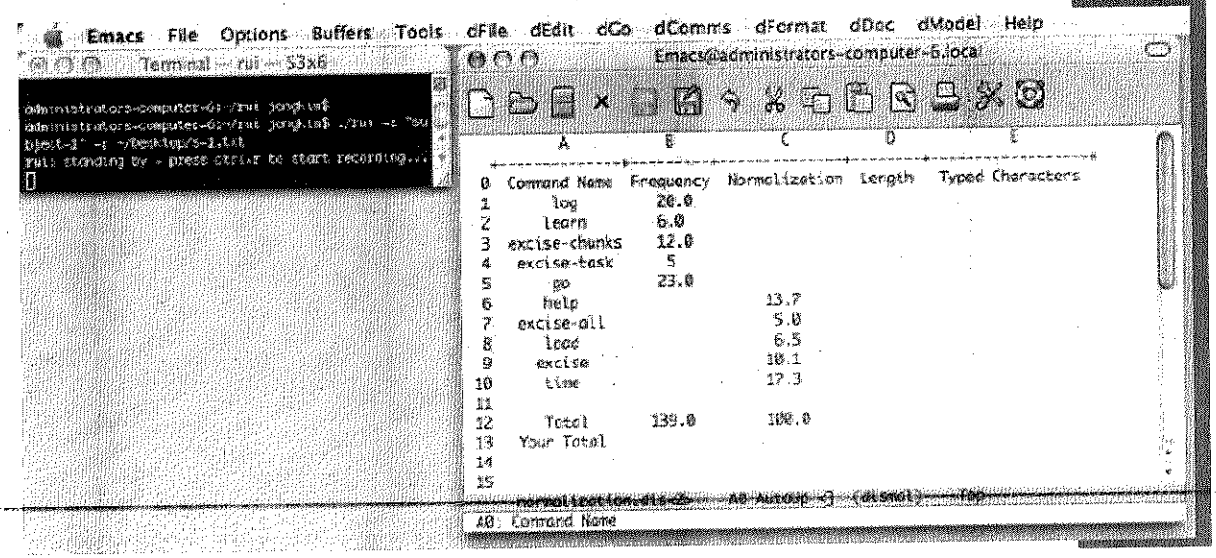


Expert

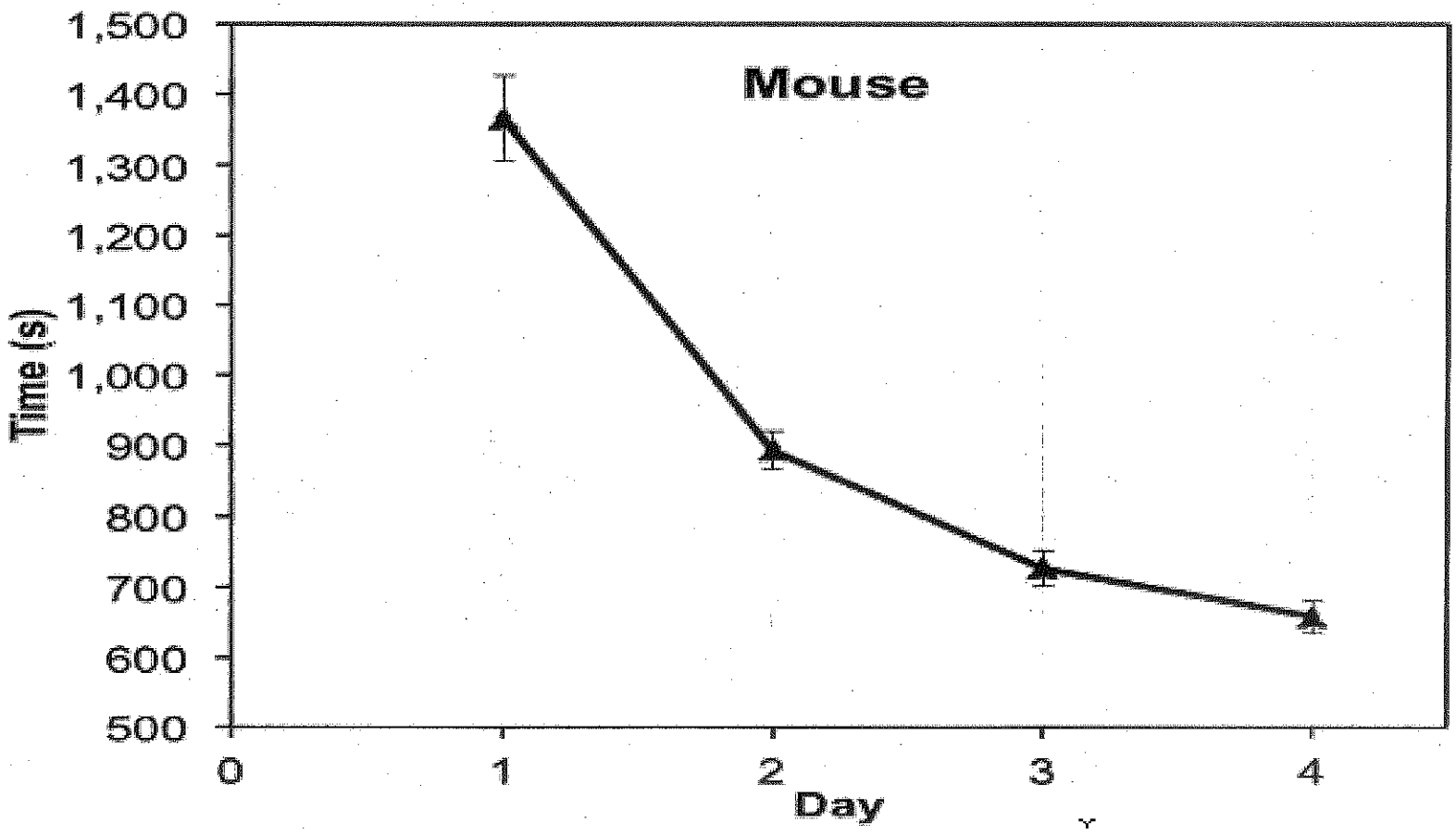
Dismal task

(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)

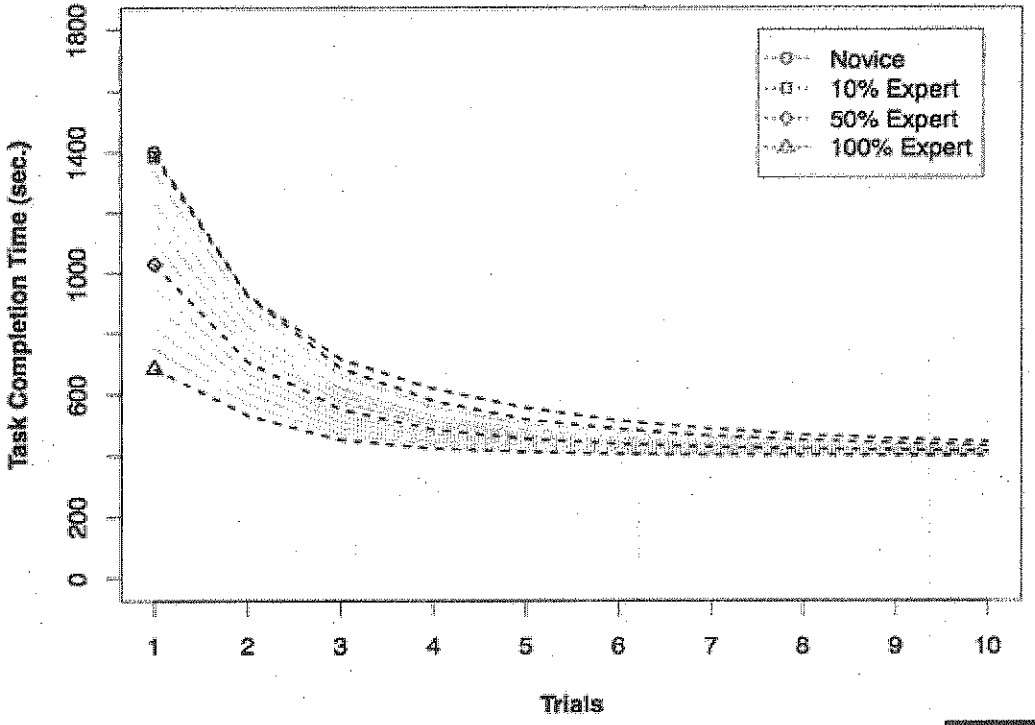


Data on the Dismal Task

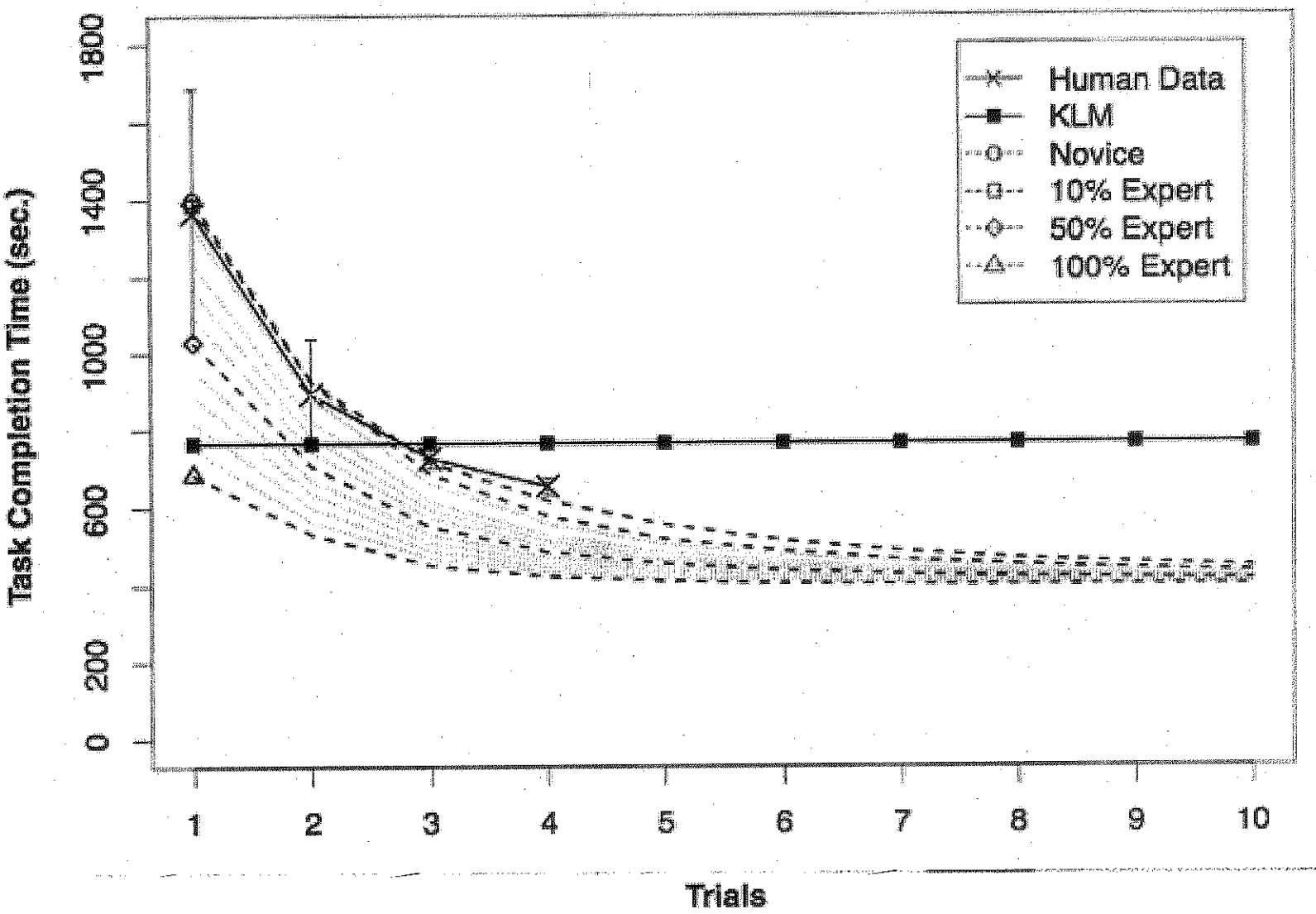


Model Performance

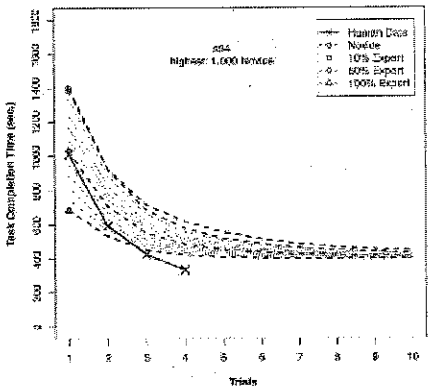
	Initial rules	Learned rules	DMs used on Trial 1
Novice	29	253	1,152
0% Expertise	617	197	1,152
10% Expertise	617	199	1,091
20% Expertise	617	197	1,030
30% Expertise	617	199	968
40% Expertise	617	199	908
50% Expertise	617	199	845
60% Expertise	617	196	784
70% Expertise	617	199	723
80% Expertise	617	199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

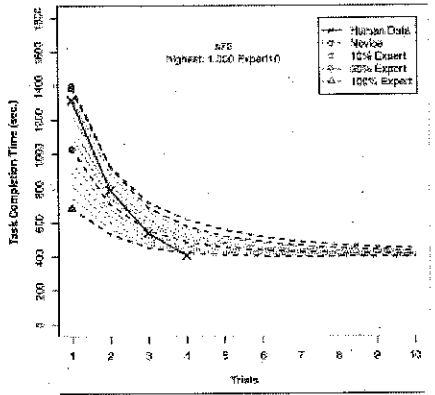


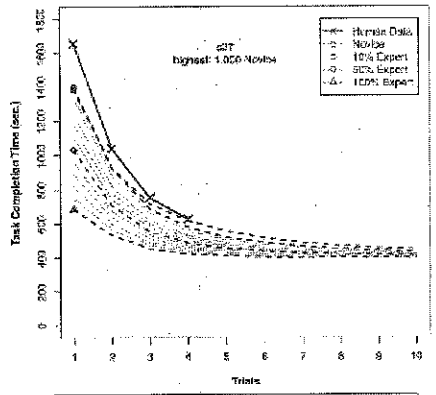
Model (predictions) to Aggregate Data



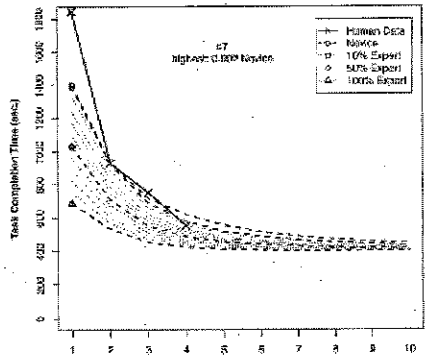
Model to individual data fits

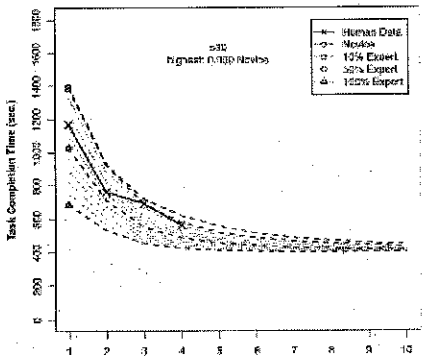


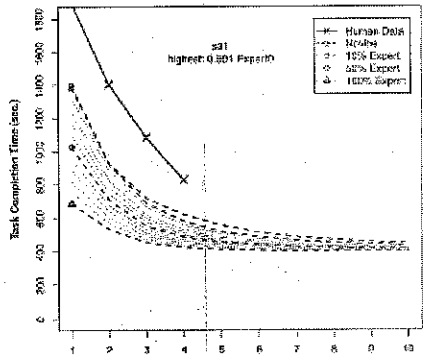




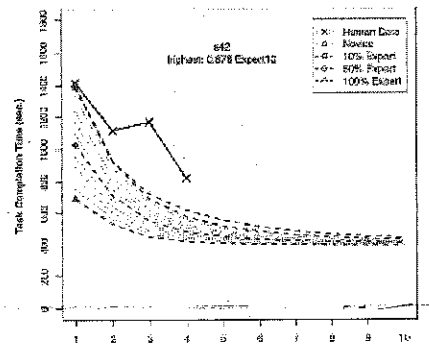
Best

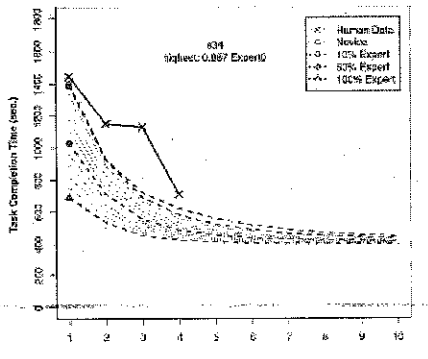


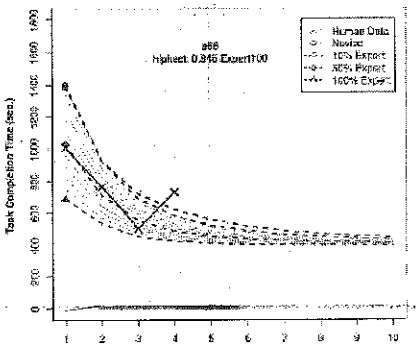




Avg







Worst

Table VII. Correlations (r) of the Model Predictions and Participant Times
(Best Fit Noted in the First Column and in Bold Italics.)

	BestFit	Nov	0	10	20	30	40	50	60	70	80	90	100
s7	Nov	***0.959***	0.949	0.952	0.952	0.955	0.954	0.955	0.954	0.953	0.953	0.952	0.951
s8	Ex0	0.938	***0.945***	0.944	0.941	0.940	0.940	0.939	0.939	0.938	0.938	0.937	0.936
s13	Nov	***0.994***	0.989	0.990	0.991	0.992	0.991	0.992	0.991	0.991	0.991	0.991	0.990
s17	Ex100	0.972	0.973	0.973	0.974	0.974	0.974	0.975	0.975	0.976	0.976	0.977	***0.978***
s18	Ex100	0.989	0.988	0.988	0.989	0.989	0.989	0.990	0.990	0.990	0.991	0.991	***0.992***
s21	Nov	***0.980***	0.973	0.975	0.976	0.978	0.977	0.978	0.977	0.977	0.977	0.977	0.977
s24	Nov	***0.979***	0.972	0.974	0.974	0.976	0.975	0.976	0.975	0.974	0.975	0.974	0.973
s29	Nov	***0.990***	0.985	0.987	0.987	0.988	0.988	0.988	0.988	0.987	0.987	0.987	0.987
s30	Nov	***0.997***	0.994	0.995	0.996	0.996	0.996	0.996	0.996	0.996	0.996	0.996	0.996
s33	Nov	***0.998***	0.996	0.997	0.997	0.997	0.997	0.997	0.997	0.997	0.997	0.997	0.996
s35	Nov	***0.980***	0.973	0.975	0.975	0.977	0.977	0.977	0.977	0.976	0.977	0.976	0.976
s37	Ex100	0.998	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	***0.999***
s38	Nov	***0.970***	0.961	0.963	0.963	0.966	0.965	0.965	0.965	0.964	0.964	0.963	0.963
s39	Ex100	0.993	0.995	0.995	0.995	0.995	0.995	0.995	0.995	0.996	0.996	0.996	***0.996***
s43	Ex100	0.976	0.981	0.980	0.980	0.979	0.979	0.979	0.980	0.981	0.981	0.981	***0.982***
s45	Nov	***0.985***	0.979	0.981	0.982	0.983	0.983	0.983	0.983	0.982	0.983	0.982	0.982
s52	Ex100	0.999	0.998	0.998	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	***0.999***
s53	Ex0	0.994	***0.997***	0.996	0.996	0.995	0.996	0.995	0.996	0.996	0.996	0.996	0.996
s55	Ex100	0.992	0.993	0.993	0.994	0.994	0.994	0.994	0.994	0.995	0.995	0.995	***0.995***
s58	Ex10	0.999	0.999	***0.999***	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999
s60	Nov	***0.950***	0.939	0.942	0.943	0.946	0.945	0.946	0.945	0.944	0.944	0.944	0.943
s62	Nov	***0.986***	0.980	0.982	0.982	0.983	0.983	0.983	0.983	0.982	0.982	0.982	0.981
s64	Ex0	0.959	***0.968***	0.965	0.965	0.963	0.964	0.963	0.964	0.965	0.964	0.965	0.966
s65	Ex0	0.998	***1.00***	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999
s69	Nov	***0.985***	0.981	0.983	0.982	0.983	0.983	0.983	0.982	0.981	0.981	0.981	0.980
s70	Nov	***0.999***	0.997	0.998	0.998	0.998	0.998	0.998	0.998	0.997	0.997	0.997	0.997
s71	Nov	***0.569***	0.562	0.565	0.559	0.561	0.561	0.559	0.558	0.554	0.553	0.551	0.547
s74	Nov	***0.987***	0.982	0.983	0.984	0.985	0.985	0.985	0.985	0.984	0.985	0.984	0.984
s78	Ex100	0.961	0.968	0.966	0.967	0.965	0.965	0.965	0.966	0.967	0.967	0.968	***0.969***
s77	Ex100	0.985	0.983	0.984	0.985	0.986	0.985	0.986	0.986	0.987	0.987	0.987	***0.988***

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

(📖 acs.ist.psu.edu/papers/)

- Cohen, M. A., Ritter, F. E., & Haynes, S. R. (2010). Applying software engineering to agent development. *AI Magazine, 31*(2), 25-44.
- Kim, J. W., & Ritter, F. E. (2015). Learning, forgetting, and relearning for keystroke- and mouse-driven tasks: Relearning is important. *Human-Computer Interaction, 30*(1), 1-33.
- Kim, J. W., Ritter, F. E., & Koubek, R. J. (2013). An integrated theory for improved skill acquisition and retention in the three stages of learning. *Theoretical Issues in Ergonomics Science, 14*(1), 22-37.
- Kukreja, U., Stevenson, W. E., & Ritter, F. E. (2006). RUI-Recording User Input from interfaces under Windows and Mac OS X. *Behavior Research Methods, 38*(4), 656–659.
- Morgan, J. H., Cheng, C.-Y., Pike, C., & Ritter, F. E. (2013). A design, tests, and considerations for improving keystroke and mouse loggers. *Interacting with Computers, 25*(3), 242-258.
- Paik, J., Kim, J. W., & Ritter, F. E. (2009). A preliminary ACT-R compiler in Herbal. In *Proceedings of ICCM - 2009- Ninth International Conference on Cognitive Modeling*, 466-467. Manchester, England.
- Paik, J., Kim, J. W., Ritter, F. E., Morgan, J. H., Haynes, S. R., & Cohen, M. A. (2010). Building user models faster with Herbal. In *Human Computer Interaction Consortium (HCIC) Workshop 2010*, 8 pages. [unrefereed workshop paper/presentation].
- Paik, J., Kim, J. W., Ritter, F. E., Morgan, J. H., Haynes, S. R., & Cohen, M. A. (2010). Building large learning models with Herbal. In D. D. Salvucci & G. Gunzelmann (Eds.), *Proceedings of ICCM - 2010- Tenth International Conference on Cognitive Modeling* (pp. 187-191).
- Paik, J., Kim, J. W., Ritter, F. E., & Reitter, D. (submitted). Implementing large user models with multiple levels of expertise.
- Ritter, F. E., & Wood, A. B. (2005). Dismal: A spreadsheet for sequential data analysis and HCI experimentation. *Behavior Research Methods, 37*(1), 71-81.
- Ritter, F. E., Yeh, K.-C., Cohen, M. A., Weyhrauch, P., Kim, J. W., & Hobbs, J. N. (2013). Declarative to procedural tutors: A family of cognitive architecture-based tutors. In *Proceedings of the 22nd Conference on Behavior Representation in Modeling and Simulation*, BRIMS2013-2127. 2108-2113. BRIMS Society: Centerville, OH.
- Ritter, F. E., Yeh, K.-C., Cohen, M. A., Weyhrauch, P., Dancy, C., & Kim, J. W. (submitted 2013). Declarative to procedural tutors: A family of cognitive architecture-based tutors and a validation study of marksmanship skill acquisition.

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:

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

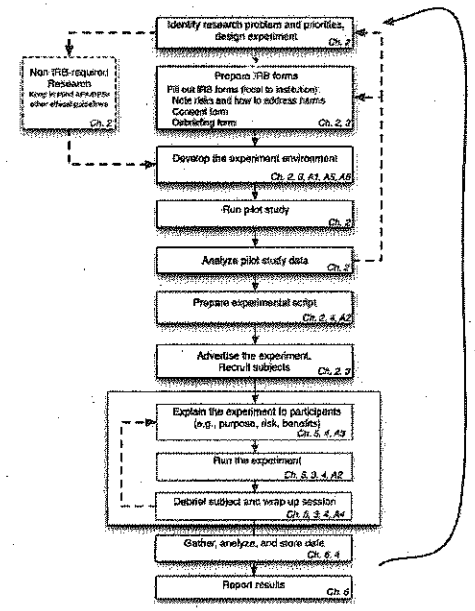
A Risk Driven Approach to Experimental Design and Practice [30]

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)

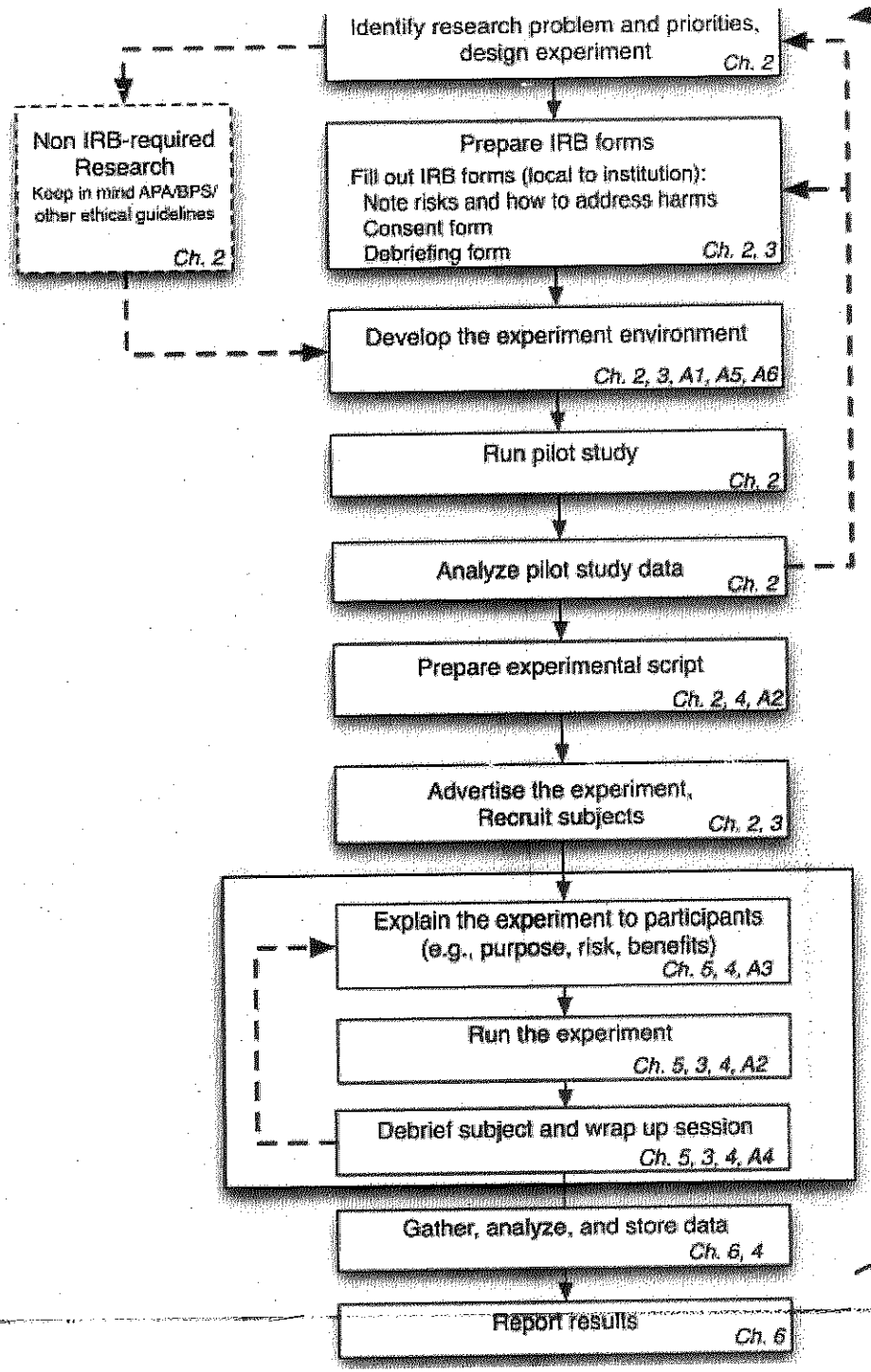
   Example problems

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 over-lapping process



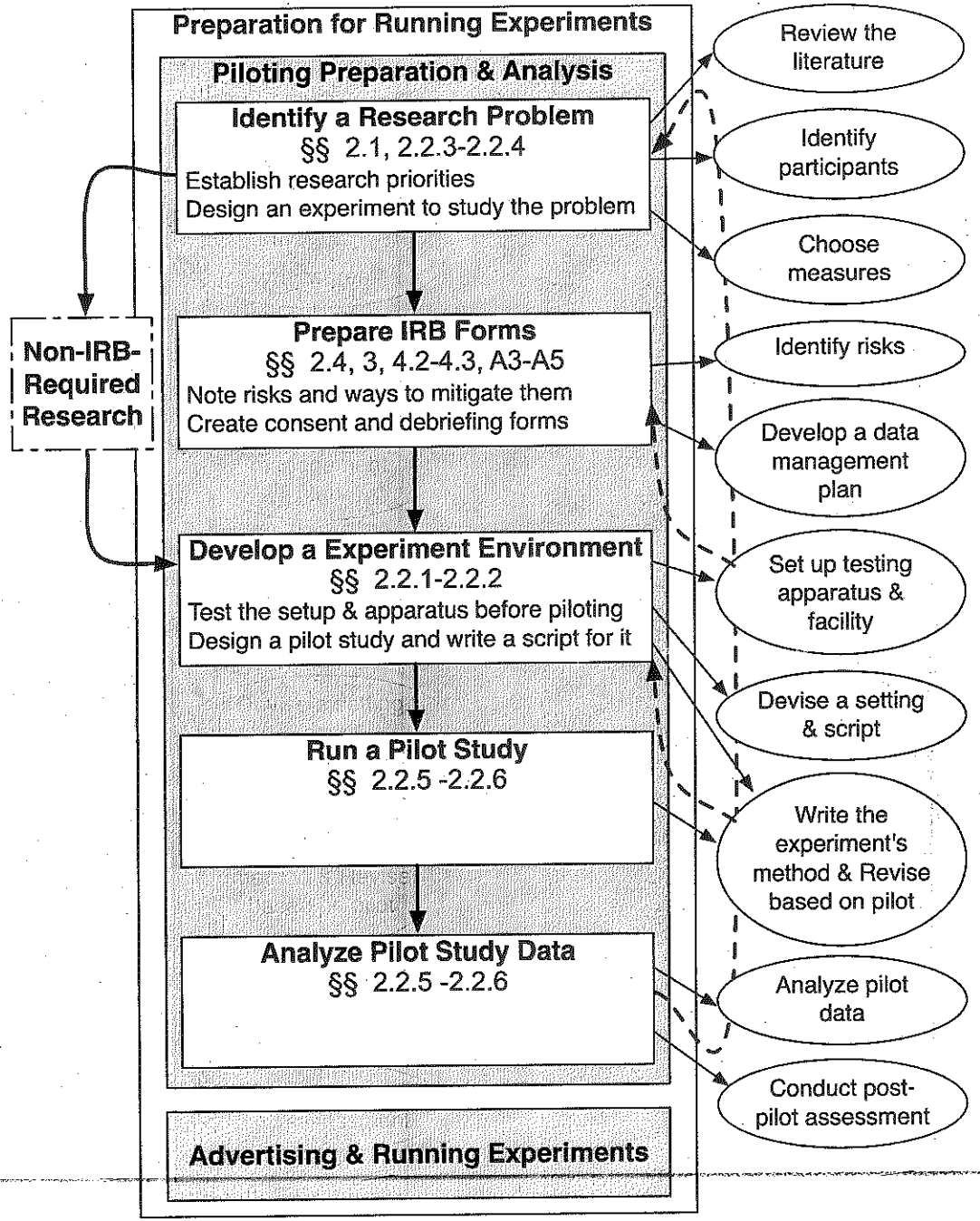
Summary: Lessons so Far

- 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 multi-disciplinary work



Summary: Piloting

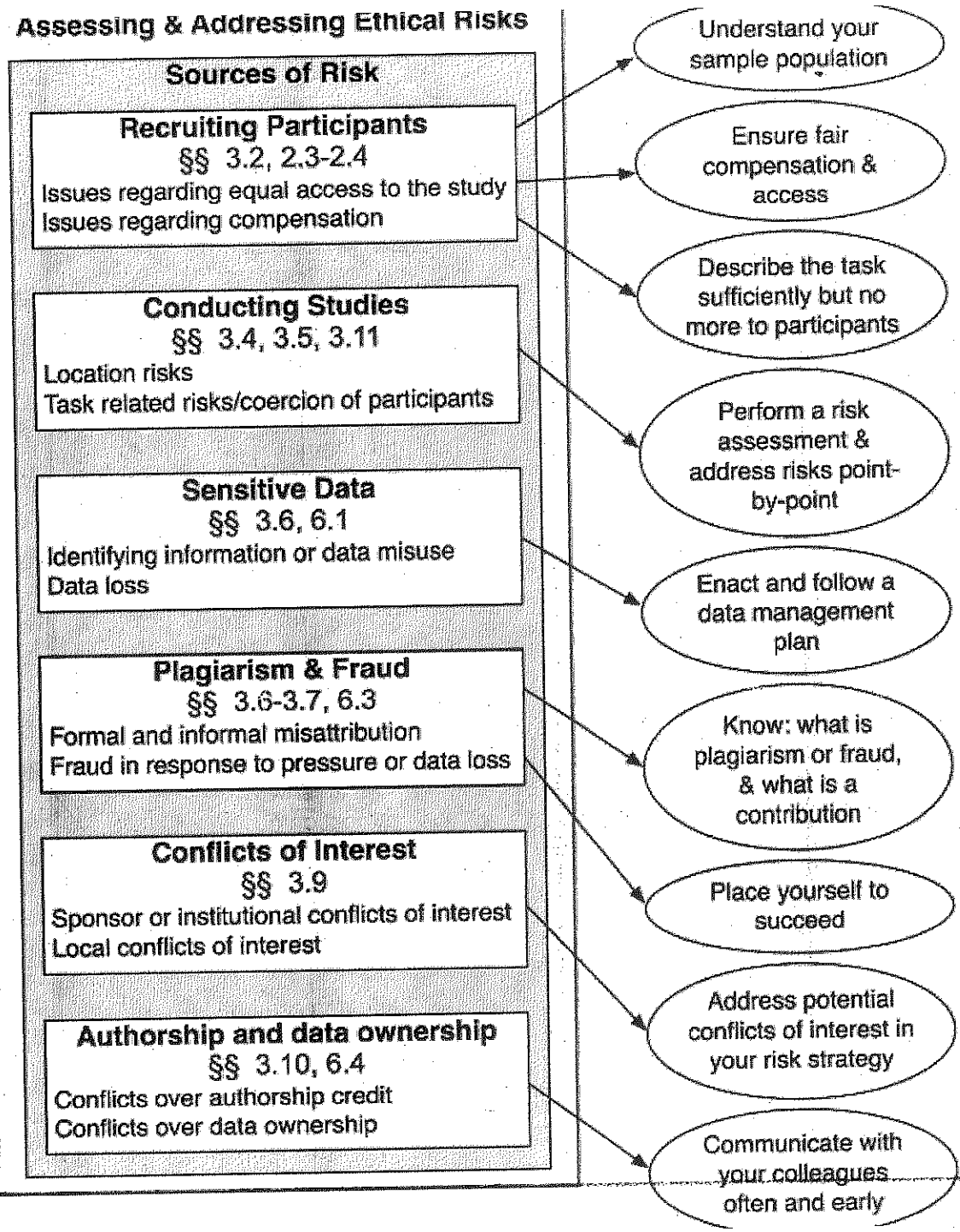
- Write out method
- Used to check your work
- Use a script,
  Step 1, start program, Step 2 "Welcome to…"
- Start local, e.g., YOU, and then officemate, and then move further and further away
- Mount a scratch monkey
- Check your apparatus and data gathering and use of data
- Consider/reconsider, number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

IST PENNSTATE

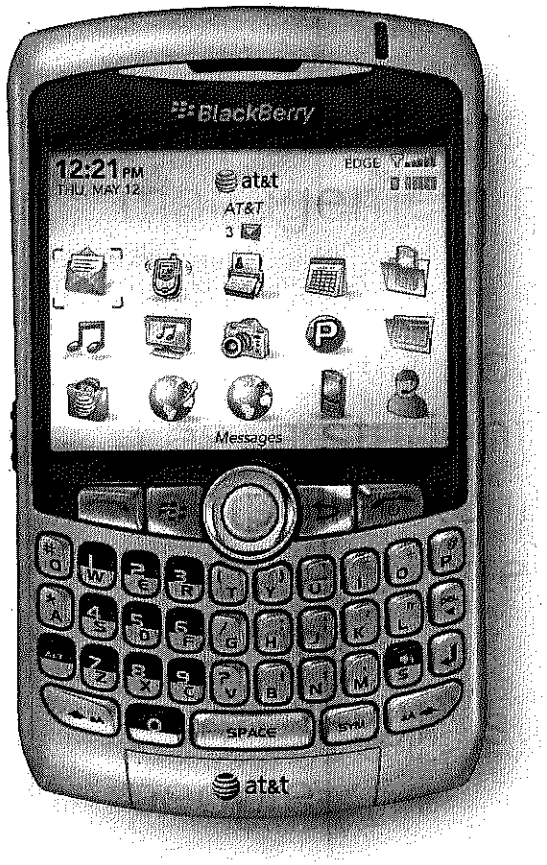
Ethical Challenges Associated with the Experimental Process

Ethical problems can be decreased by deliberate proactive action.

A couple of bad examples and then a general view



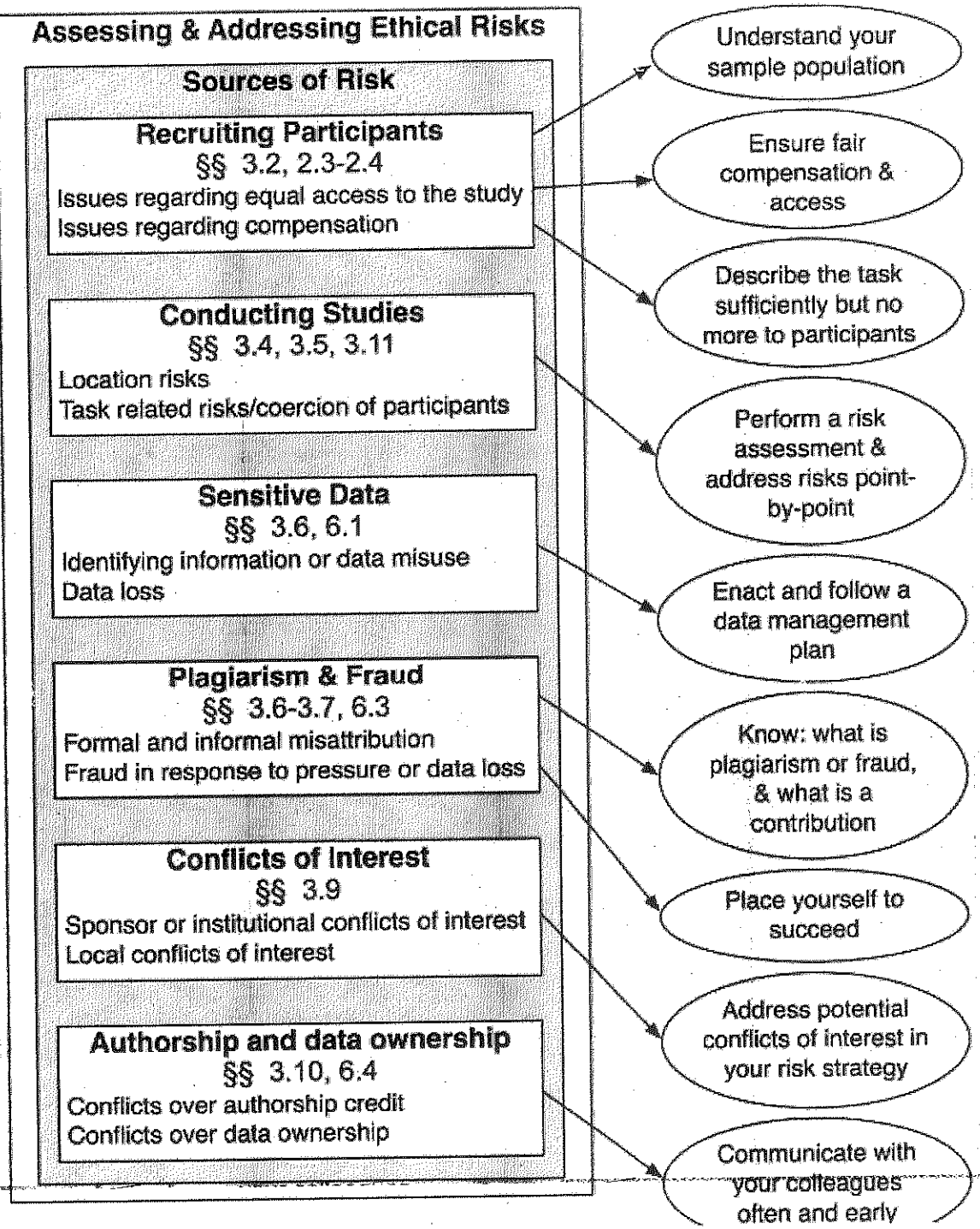
A HCI Study Gone Wrong (circa 2008)



- 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.



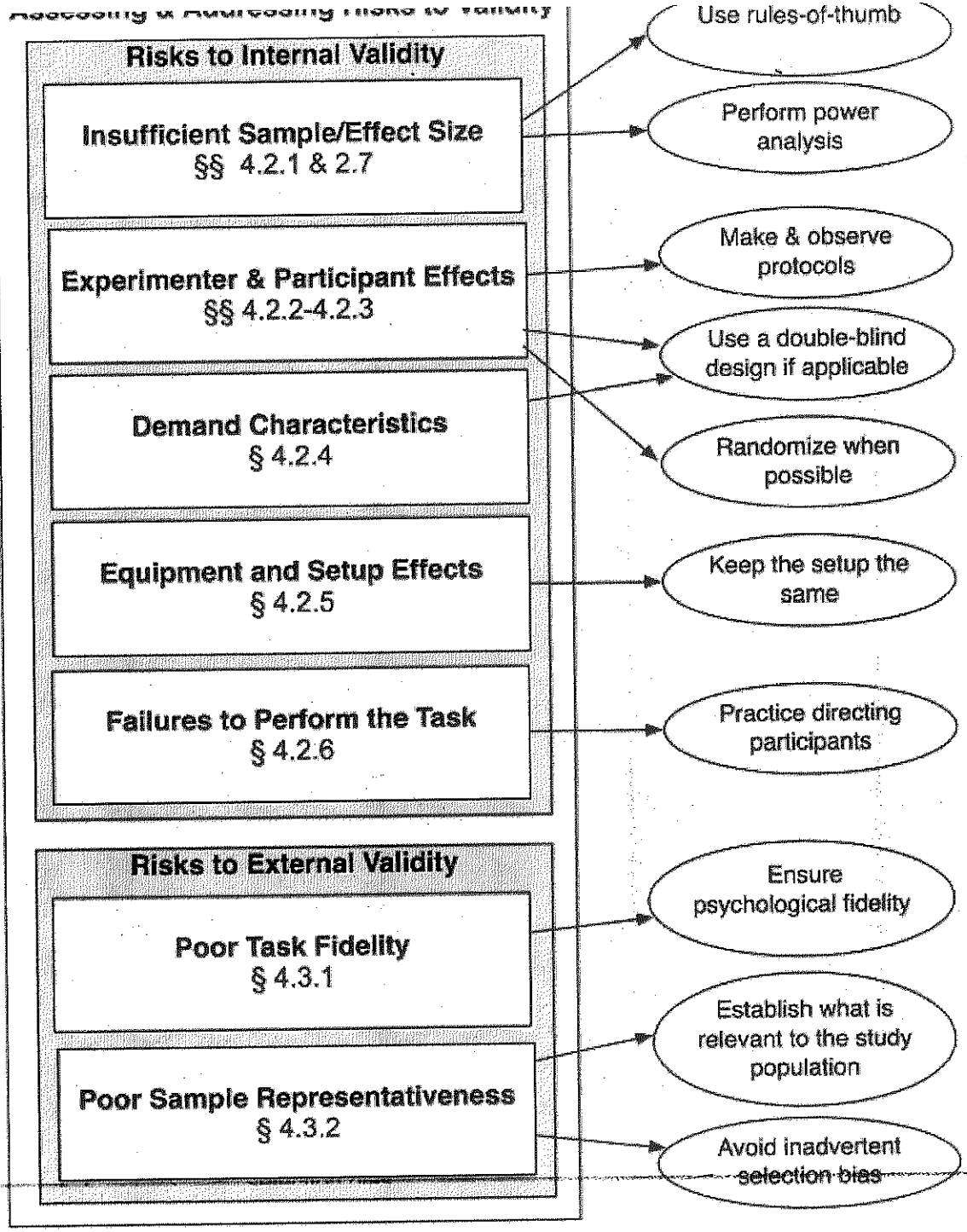
Summary: How to avoid ethical problems

- 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)

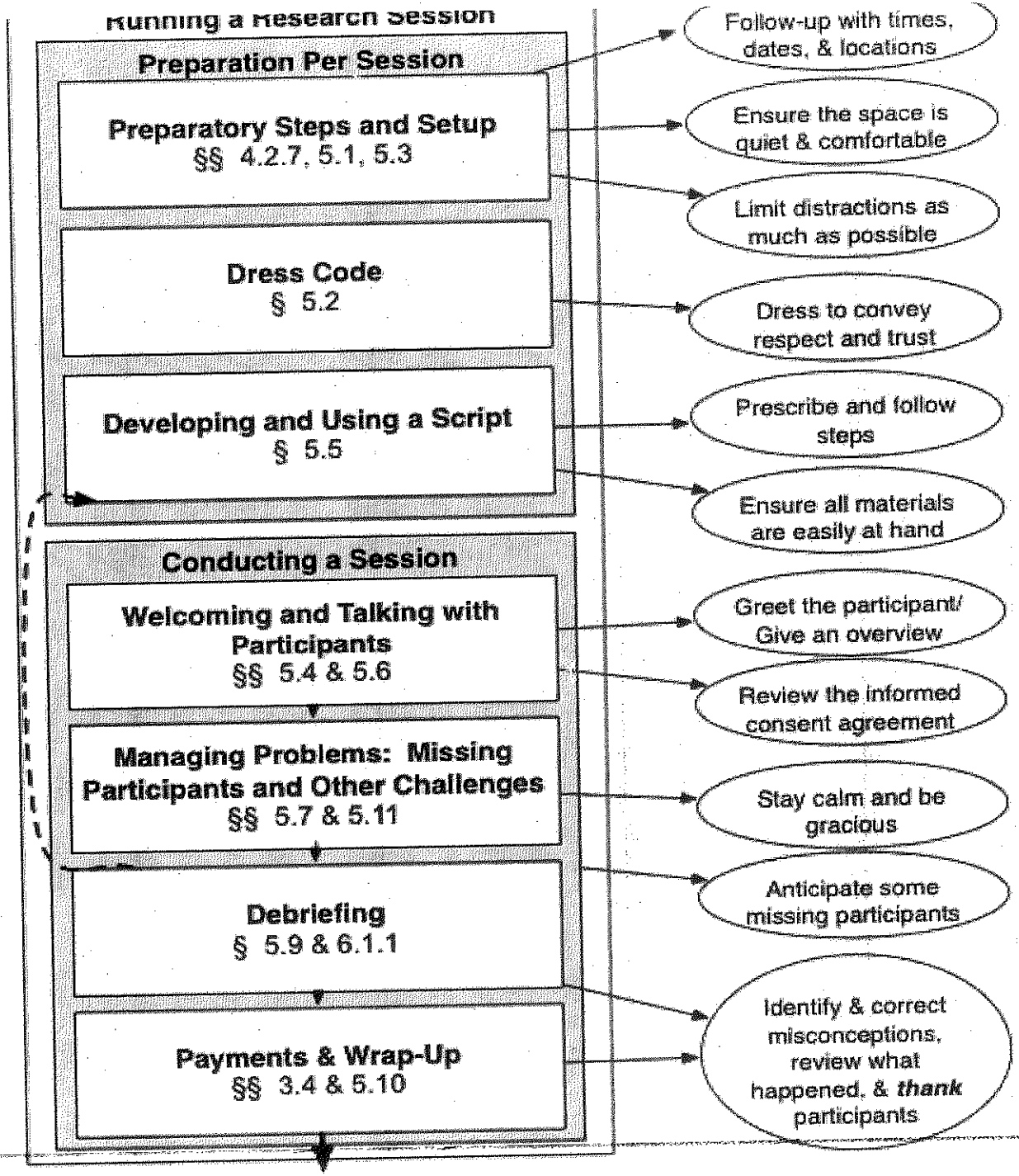
Challenges to validity can be anticipated and mitigated.



Conducting an Experiment

(TB, p. 24-25)

uccess in execution is directly correlated to careful preparation



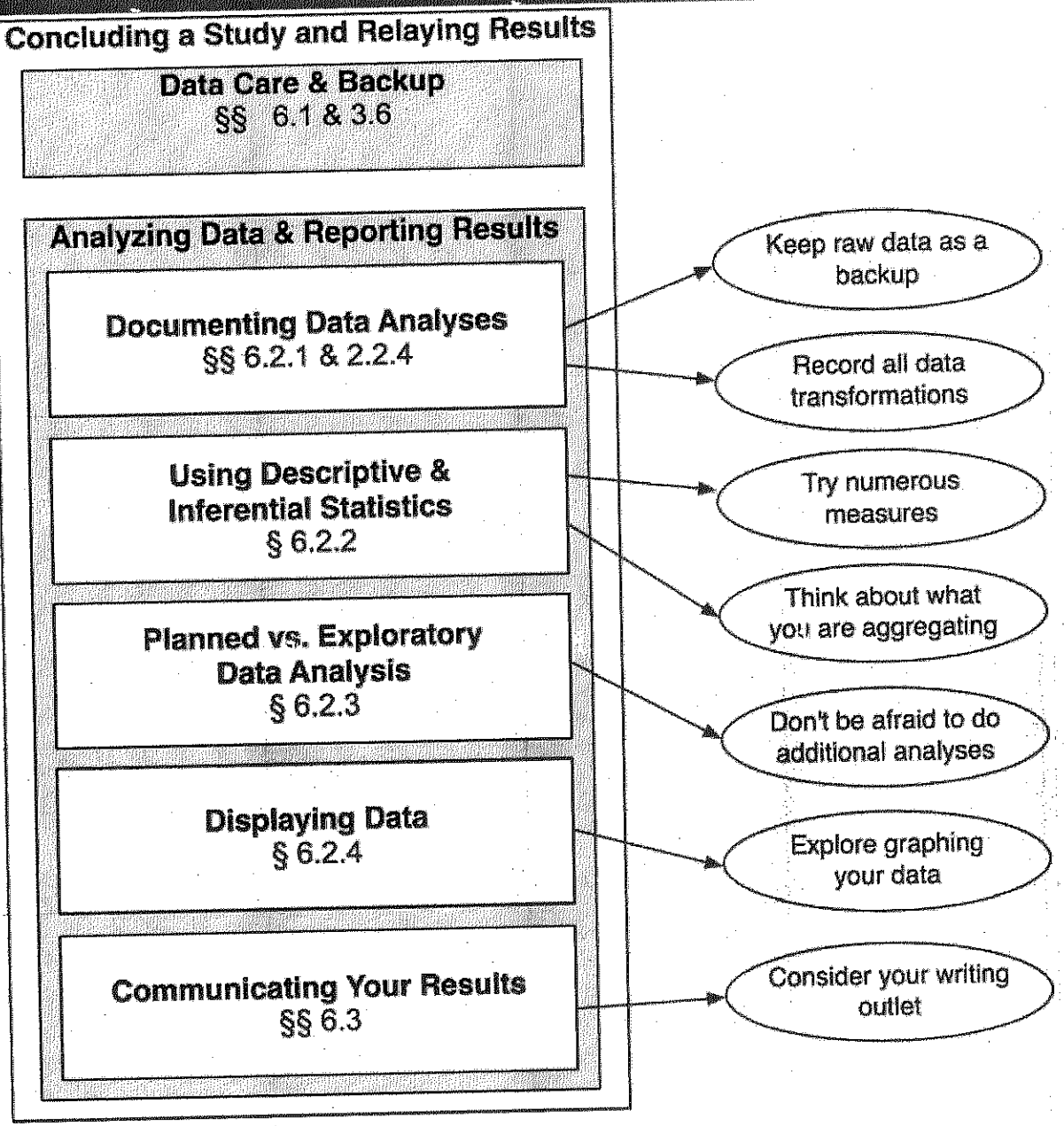
Summary: Running a session

- 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)

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 anomolies
- 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

acs.ist.psu.edu/papers

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

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

Boehm, B., & Hansen, W. (2001). The Spiral Model as a tool for evolutionary acquisition. *Crosstalk: The Journal of Defense Software Engineering, 14*(5), 4-11.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.

Delaney, P. F., Reder, L. M., Staszewski, J. J., & Ritter, F. E. (1998). The strategy specific nature of improvement: The power law applies by strategy within task. *Psychological Science, 9*(1), 1-8.

Digiusto, E. (1994). Equity in authorship: A strategy for assigning credit when publishing. *Social Science & Medicine, 38*(I), 55-58.

Kim, J. W., Koubek, R. J., & Ritter, F. E. (2007). Investigation of procedural skills degradation from different modalities. In *Proceedings of the 8th International Conference on Cognitive Modeling*, 255-260. Taylor & Francis/Psychology Press: Oxford, UK.

Pew, R. W., & Mavor, A. S. (Eds.). (2007). *Human-system integration in the system development process: A new look.* Washington, DC: National Academy Press. http://books.nap.edu/catalog.php?record_id=11893, checked March 2012.

Reder, L. M., & Ritter, F. E. (1992). What determines initial feeling of knowing? Familiarity with question terms, not the answer. *Journal of Experimental Psychology : Learning, Memory & Cognition, 18*(3), 435-451.

Ritter, F. E., Kim, J. W., Morgan, J. H., & Carlson, R. A. (in press, 2012). *How to run experiments: A practical guide to research with human participants.* Currently 150 pages. Thousand Oaks, CA: Sage.

Ritter, F. E., Schoelles, M. J., Quigley, K. S., & Klein, L. C. (2011). Determining the number of model runs: Treating cognitive models as theories by not sampling their behavior. In L. Rothrock & S. Narayanan (Eds.), *Human-in-the-loop simulations: Methods and practice* (pp. 97-116). London: Springer-Verlag.

Roediger, R. (2004). What should they be called? *APS Observer, 17*(4), 5, 46-48. Online: www.psychologicalscience.org/observer/getArticle.cfm?id=1549.

Winston, A. S. (1990). Robert Sessions Woodworth and the "Columbia Bible": How the psychological experiment was redefined. *The American Journal of Psychology, 103*(3), 391-401