Individual Differences

Marsha C. Lovett
Carnegie Mellon University

Differences Beyond Mere Noise

Between-subject differences: systematic, reliable

- Architectural (processing) differences

  e.g., processing speed, working memory capacity, decay

- Knowledge-based differences
  - Knowledge contents (e.g., facts, strategies, etc.)
  - Same content, but differences in experience/practice

    e.g., different trial sequences, different real-world experiences

- Representational differences
  - Features represented, knowledge structures

Differences Beyond Mere Noise

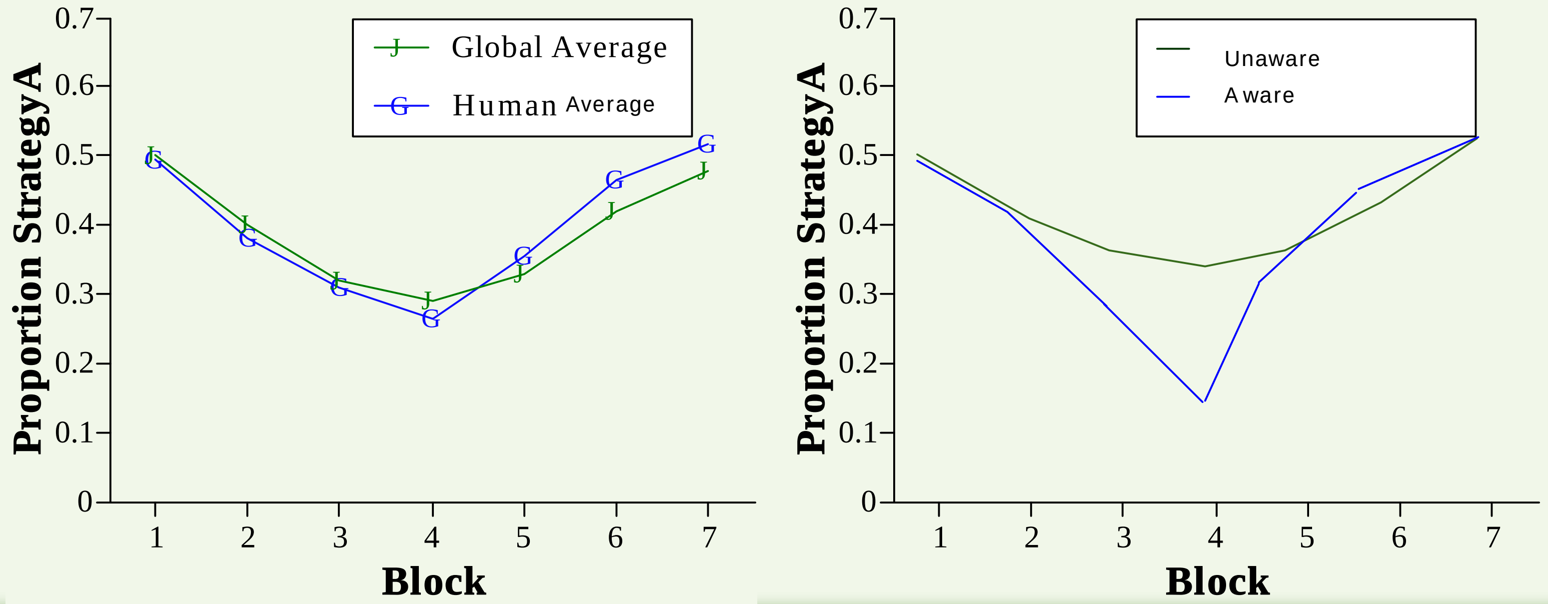
Within-subject differences: temporal, subtle

- Knowledge/experience grows (learning)
- Processing parameters change (e.g., fatigue)
- Representation changes (insight)

Why model Ind Diffs in ACT-R?

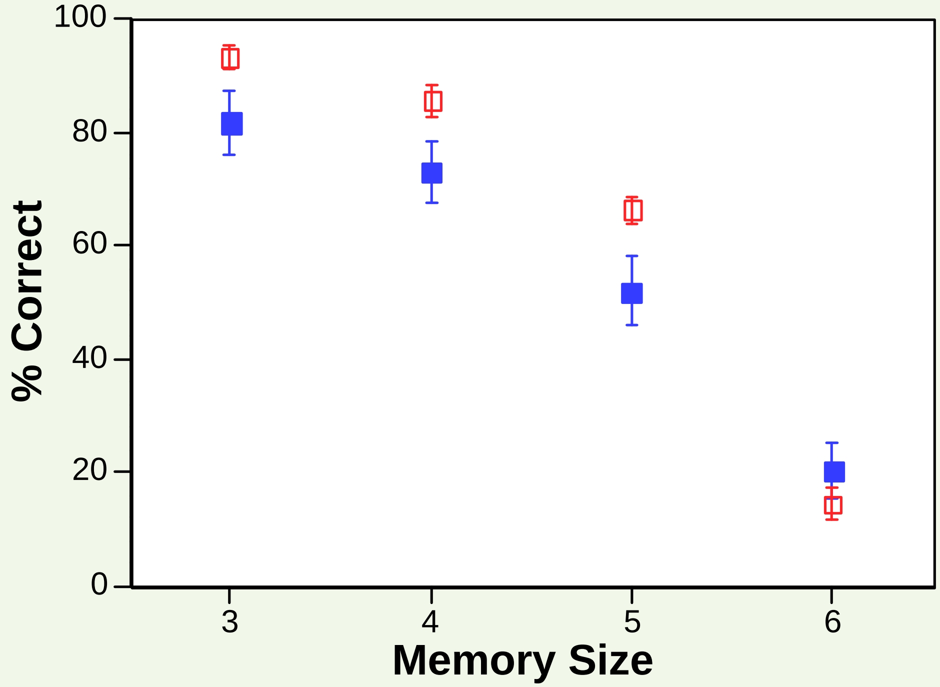
Additional constraints on model/theory

– Models can fit the averaged data and yet fail to fit individuals' (or subbroups') data (cf. Siegler, 1987)



Why model Ind Diffs in ACT-R?

Predicted variability often lower than observed



(from Lovett, Reder, & Lebiere, 1997)

Why model Ind Diffs in ACT-R?

ACT-R is a nonlinear system $\Rightarrow$ variation in input can have surprising consequences

FIXED CASE

fixed input		fixed output
$W \equiv 1$ $\rightarrow$	Linear System $f(W) = aW$	$\rightarrow$ $a \cdot 1 = a$
fixed input		fixed output
$W \equiv 1$ $\rightarrow$	Nonlinear System $g(W) = e^W$	$\rightarrow$ $e^1 = e$

Why model Ind Diffs in ACT-R?

VARYING CASE

variation in input

$W \sim \text{Normal}(1,\sigma^2) \rightarrow$

Linear System
$f(W) = aW$

$\rightarrow \text{Normal}(a, a\sigma^2)$

variability in output

mean same as fixed

variation in input

$W \sim \text{Normal}(1,\sigma^2) \rightarrow$

Nonlinear System
$g(W) = e^W$

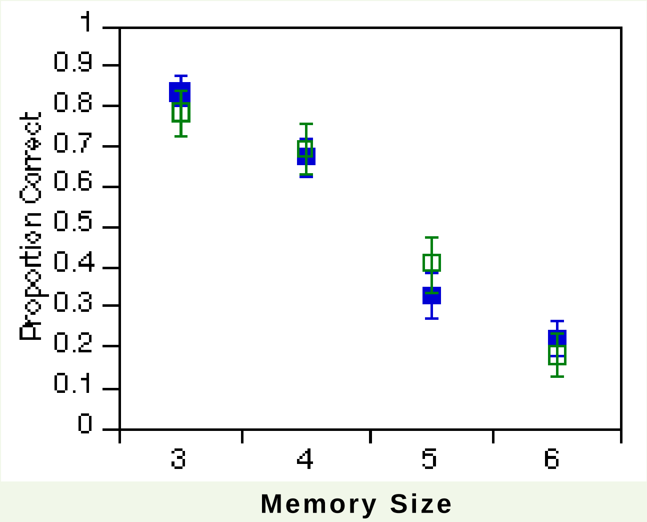
$\rightarrow \text{Normal}(e^{1+.5\sigma^2}, ?\#@!)$

variability in output

change in mean!

Why model Ind Diffs in ACT-R?

- Adding Ind Diffs changes variability *and mean*



- This affects fitting of other global parameters

Especially within unified theories

- Unified theories like ACT-R use single set of mechanisms to capture data *across tasks*
- Modeling Ind Diffs within ACT-R
  – Allows modeling of an individual across tasks
  – Allows testing of individual difference theories across tasks

How to model Ind Diffs in ACT-R?

- Architectural (processing) differences
  - Global parameters: G (motivation), W (WM), P/M …
- Knowledge differences — content vs experience
  - Symbolic: Different sets of productions, chunks
  - Subsymbolic: Production utilities, chunk activations
- Representational differences (cf. Lovett & Schunn, 1999)
  - Chunk types, production conditions, proc vs. decl

ACT-R Ind Diff Models

Architectural Parameters Varied

	Param	Task (Modeler)
Ind'l Subject	G egs d	Strat choice in BST (Schunn)
	W	KA-ATC (Taatgen)
	W*	Scheduling (Taatgen, Jongman)
	W	WM tasks (Lovett, Reder,Lebiere)
		List memory(Reder, Schunn)
		Exp'tal Design (Schunn)
		Strat choice in memory (Reder, Schunn)
		Scaling (Petrov)
Sub-group		Device operation (Byrne)
		Digit symbol (Byrne)

ACT-R Ind Diff Models

Symbolic Knowledge Varied

	Knowledge	Task (Modeler)
Ind'l Subject	productions	User interface (Gray)
	productions	2-col subtraction (Young)
	productions	Seriation length (Young)
Sub-group	productions chunks	Exp'tal Design (Schunn)

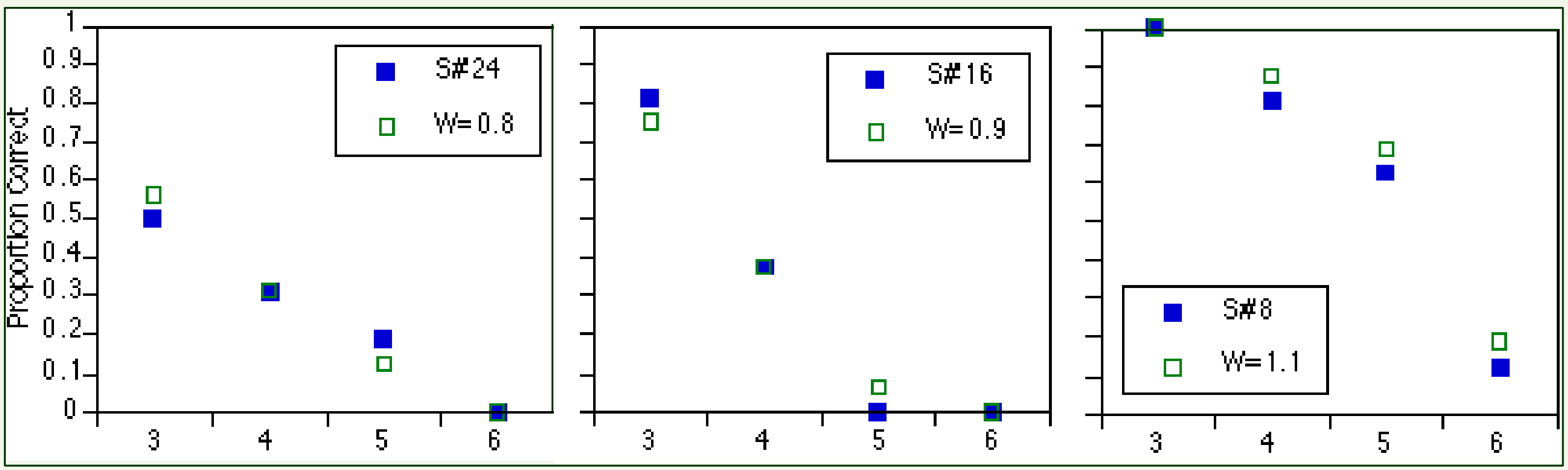
ACT-R Ind Diff Models

Other combinations varied

	? varied	Task (Modeler)
Ind'l Subject	egs p-utils	Analogy in prob solving (Salvucci)
	knwldg wm	Unix tutor, piloting (Doane)
Sub-group	p-utils	Early algebra (Koedinger)
	rt egs p-conds	TON (Jones, Ritter)

Example: Arch'l Param Varied

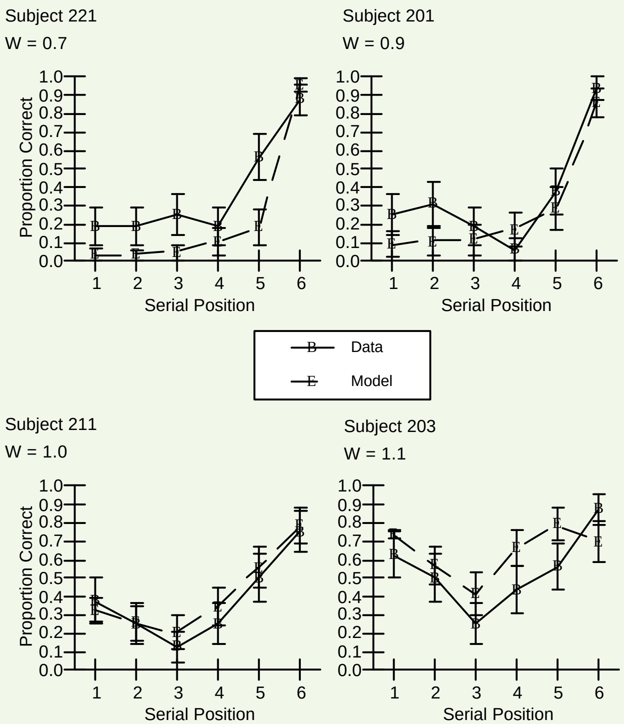
- WM capacity's effect on performance
  - Model individuals in a WM task called MODS
  - Take MODS model and vary W parameter



(Lovett, Reder, & Lebiere, 1999)

Example cont'd

- Can fit individual data at even finer level
  - Serial position effect (w/ W fit from set size effect)

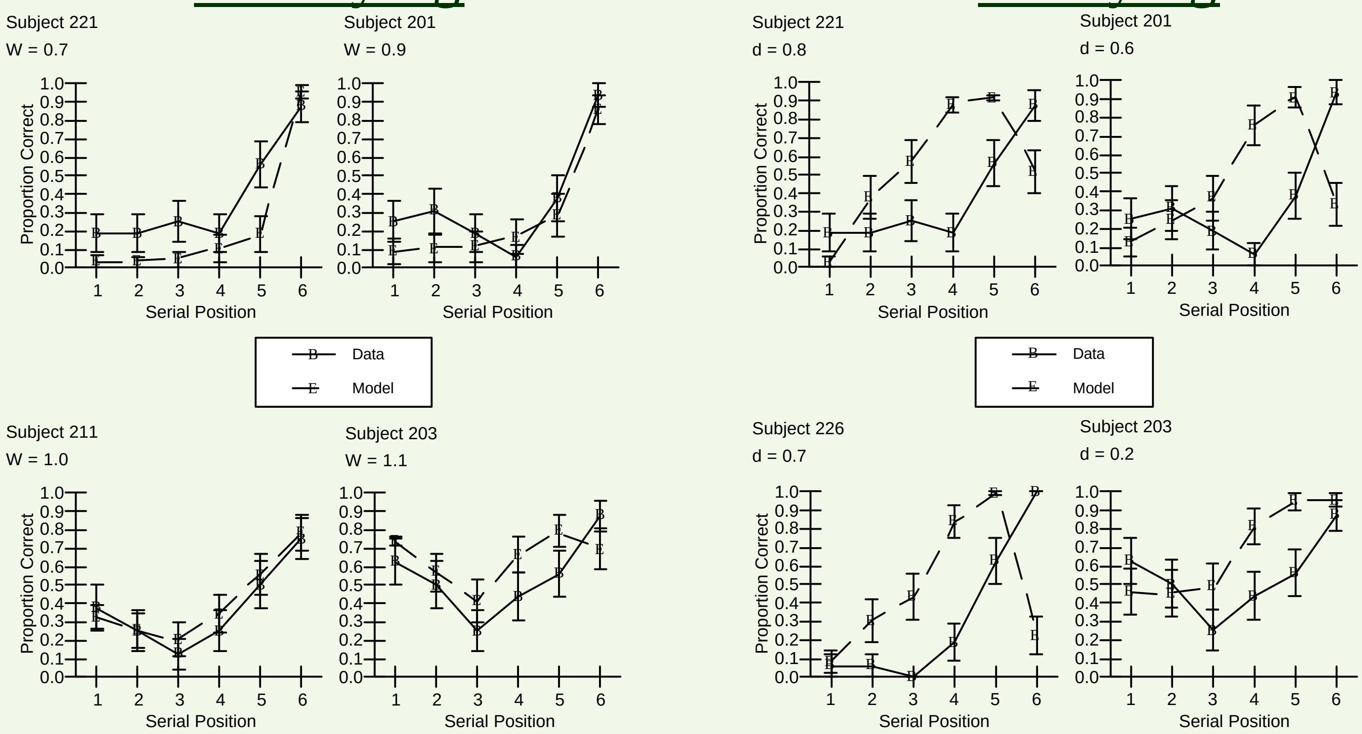


Example cont'd

- Can other params account for ind'l patterns?

W varying

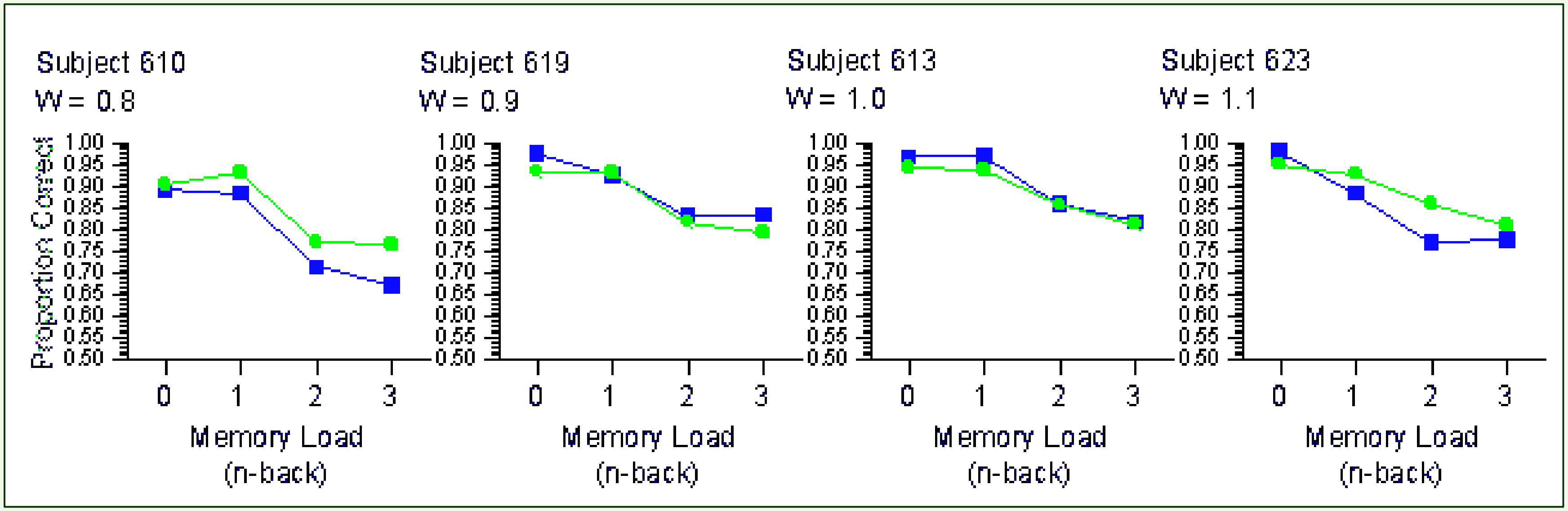
d varying



(Daily, Lovett, & Reder, 2001)

Example cont'd

- Cross-task predictions w/ no new parameters!
  - Subjects performed MODS & N-back tasks
  - W's estimated from MODS to predict N-back



(Lovett, Daily, & Reder, 2000)

Summary: Architectural Ind Diffs

- Vary global parameter(s) to represent Ind Diffs
- Parameter values take on meaning
  - Predict other measures within task
  - Predict performance on other tasks
  - Relate to other empirical measures
  - Change across time
- Can compare different theories of Ind Diffs

Ex: Symbolic Knowledge Varied

- Scientific Discovery in Ψ Microworld
  (Schunn & Anderson, 1998)
  - Task: reveal "truth" behind data by conducting experiments and interpreting data tables
  - Large performance differences in experiment designs and interpretation
  - Subgroups modeled with different sets of procedural & declarative knowledge

Issues in Symbolic Knowledge Diffs

- Varying procedural/declarative knowledge
  - Consider elements as qualitative parameters
    - Model is set of elements drawn/not from fixed set
  - Use other constraints to winnow possible model versions from the power set
    - Developmental progression
    - Learnable via symbolic learning mechanisms

Ex: Subsymbolic Knowledge Diffs

- Interface use (Gray et al., 2001)
  - Model runs get same experience as subjects
  - Use same priors for chunk activations and production utilities, but different experience across model runs leads outputs to diverge
- Early algebra (Koedinger & Maclaren, 2001)
  - Vary priors for production utilities to account for subjects' different pre-experimental experience

Ex: Representational Ind Diffs

- Tower of Nottingham (Jones, Ritter, Wood, 2000)
  - Goal: Capture developmental differences by implementing developmental theories in ACT-R
  - Besides global parameters (rt, egn), vary # of conditions in productions' left hand sides
- Analogical Problem Solving (Salvucci, 1998)
  - Goal: Capture wide variation in eye-fixation strategies when subjects refer to source problem
  - Besides varying parameters (egn, prod utilities), built decl-based and proc-based models

Ind Diffs Model Fitting

- Many Ind Diff parameters
  - Fit IndDiff parameters for each subject $\Rightarrow$ NP_{big}
- Few Ind Diff (esp 1) parameters
  - Fit global params while IndDiff param(s) are drawn from distribution (i.e., not fixed)
  - Fit IndDiff param(s) for each subject $\Rightarrow$ $G+NP_{small}$
- Hierarchical modeling: best of both!
  - Fit global and IndDiff params together
  - IndDiff param-values from distribution $\Rightarrow$ $N_{small}P$

Concluding Bold Statement

All ACT-R models should be Ind Diffs Models

- You have the data (simply omit averaging step)
- It's just a few more fitting cycles (see prev slide)
- Avoids perils of averaging over subjects
- Increases model variability (closer to observed)

- Default parameter settings would become distributions, not fixed values

More on Why: Guess the Authors

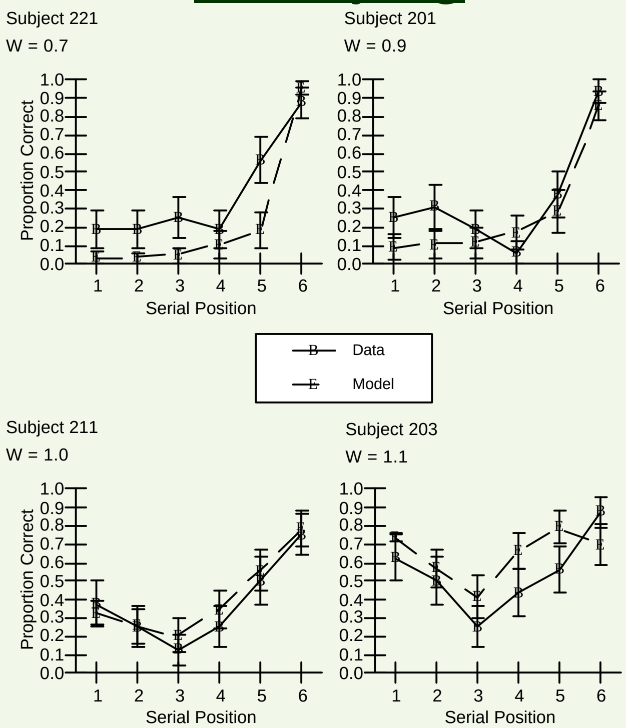
Computational models need to be able to account for both the commonalty across individuals' processing as well as the variation between individuals' performance. Cognitive models should be developed to predict the performance of individual participants across tasks and along multiple dimensions. Ideally, such a modeling effort would be able to predict individuals' performance in a new task with no new free parameters, presumably after deriving an estimate of each individual's processing parameters from previous modeling of other tasks. (Lovett, Daily, & Reder, 2001)

A way to keep the multiple-constraint advantage offered by unified theories of cognition while making their development tractable is to do Individual Data Modelling (IDM). That is, to gather a large number or empircal/experimental observations on a single subject (or a few subjects analysed individually) using a variety of tasks that exercise multiple abilities (e.g., perception, memory, problem solving), and then to use these data to develop a detailed computational model of the subject that is able to learn while performing the tasks. (Gobet & Ritter, 2000)

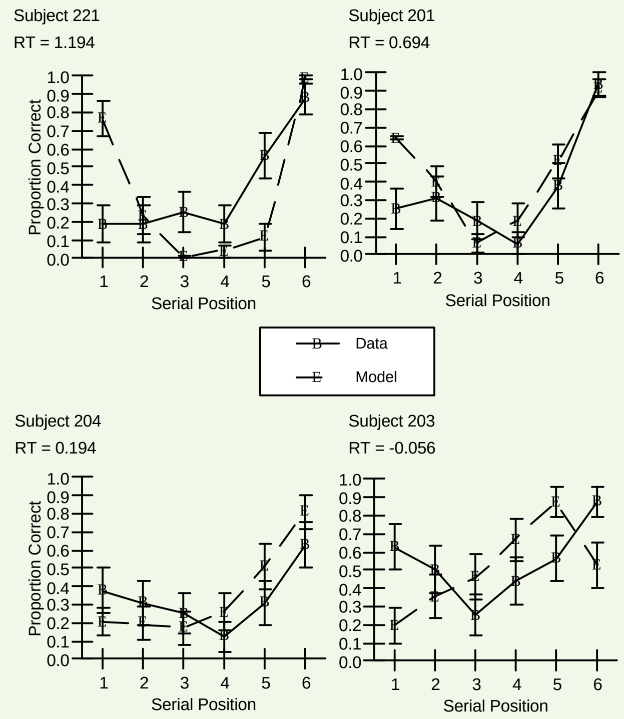
Example cont'd

- Sensitivity analysis: do other params manage?

W varying



rt varying



(Daily, Lovett, & Reder, 2001)