The 30th Annual Conference of the Cognitive Science Society

Computational Modeling of Spoken Language Processing: A hands-on tutorial

Computational Modeling of Spoken Language Processing: A hands-on tutorial

Ted Strauss

New School University

Dan Mirman

Jim Magnuson

University of Connecticut Department of Psychology

and

Haskins Laboratories

Plan

- Module 1: Introduction, About TRACE
- Module 2: Tour of jTRACE
- Module 3: Classic simulations
- Module 4: Scripting
- Module 5: Linking hypotheses
- Module 6: Lab time, Q&A, one-on-one

An aside

- Why did we develop jTRACE?
  - To facilitate large-scale modeling
  - To promote active testing of TRACE predictions and wider use of modeling
  - To facilitate replication and sharing of simulations
- How did we develop jTRACE?
  - With a budget supplement to an NIDCD R01 and a couple Ted years
- Why are we doing this tutorial?

Module 5: Advanced topics

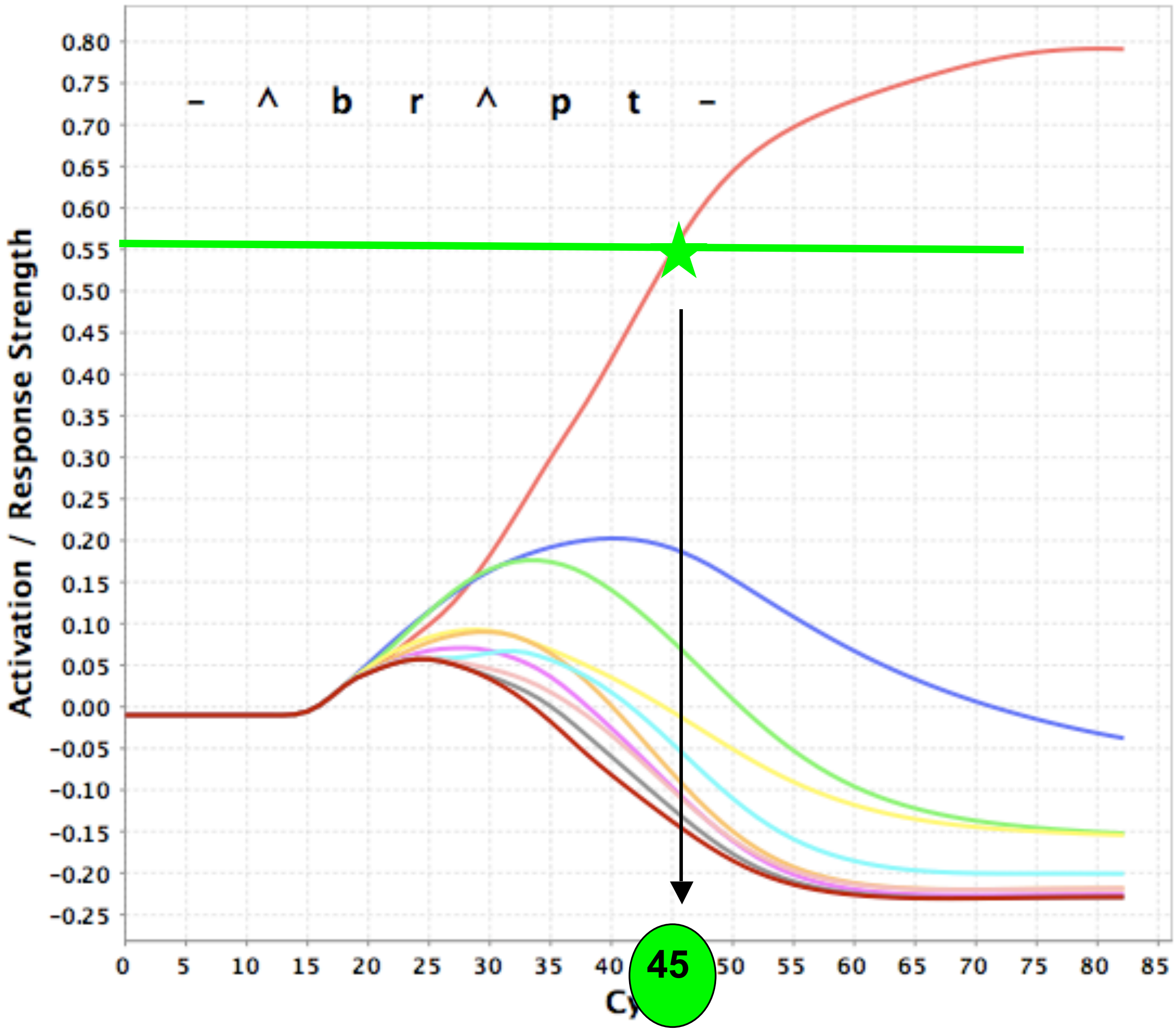
- How do you decide whether a model has succeeded or failed?
  - *Connecting model to human behavior*
- Pitfalls: simulations can fail at multiple levels
  - Theory -- *most interesting/informative*
  - Implementational details/parameters
  - Linking hypotheses -- *not a **model** failure -- equivalent to flawed operational definitions in an experiment!*
- Before assuming a failure has theoretical implications, other levels must be excluded

Linking hypotheses

- Informal: does model capture basic trends?
- Formal: linking hypothesis
  - Link **model** to **data** by constructing task constraints for the model analogous to those faced by human subjects
  - Model: Activations over time
  - Data: Reaction times/accuracy for **specific** decisions or behaviors (lexical decision, eye tracking, ERP)

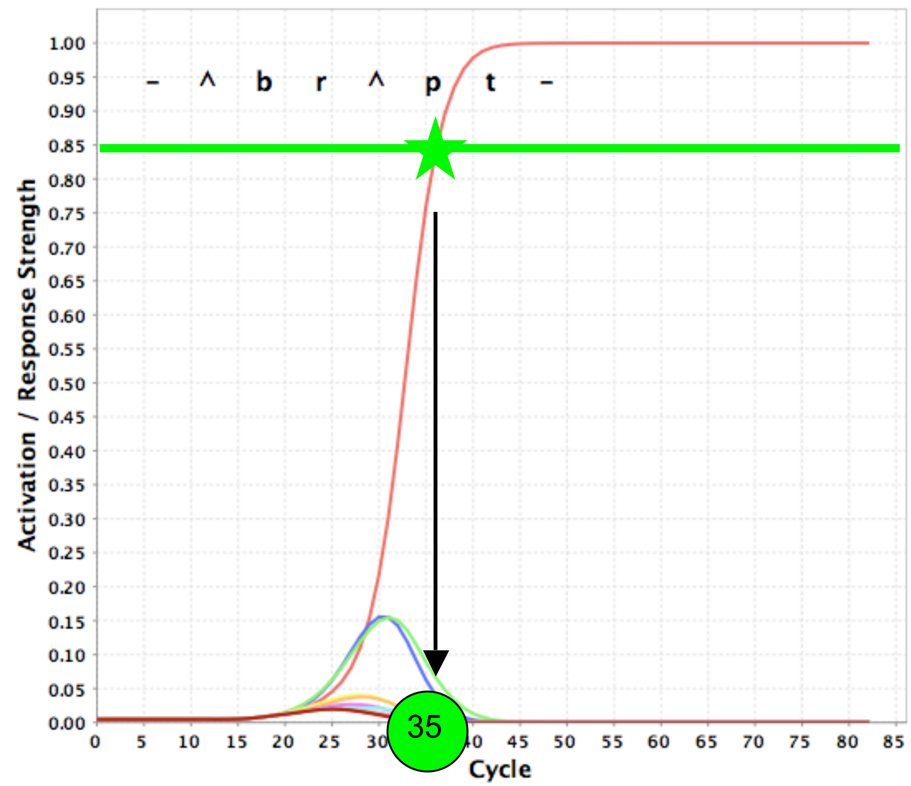
Simple example 1: Threshold

- Recognition = word unit activation exceeds threshold
- RT ≈ number of processing cycles from word onset
- Activation ~ internal state; what about choice behavior?



Simple example 2: Response probability

- RT ≈ number of processing cycles from word onset
- Additional competition analogous to human choice behavior in many domains
- Formalization of overt choice based on internal states
- When to use: many choice situations, but especially AFC



$$L_i = \frac{S_i}{\Sigma S_j} \qquad S_i = e^{ka_i}$$

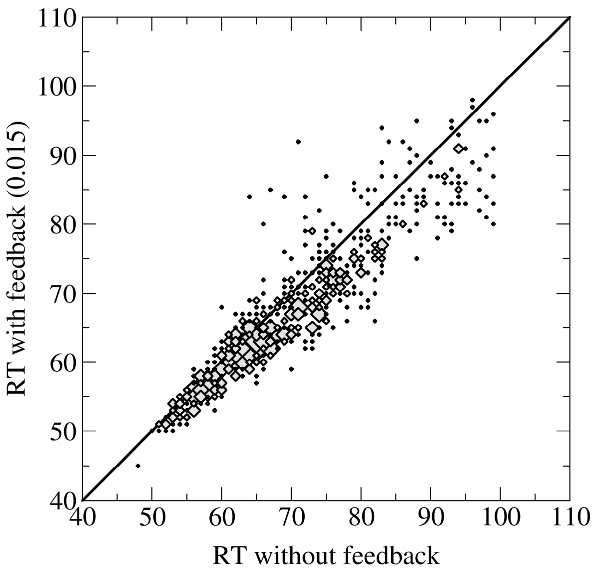
Pitfalls in modeling

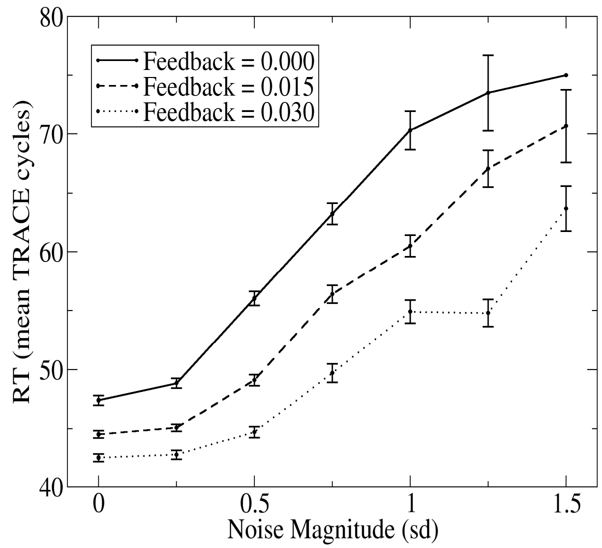
1. Material selection
2. Material manipulation
3. Linking hypotheses
4. Logic

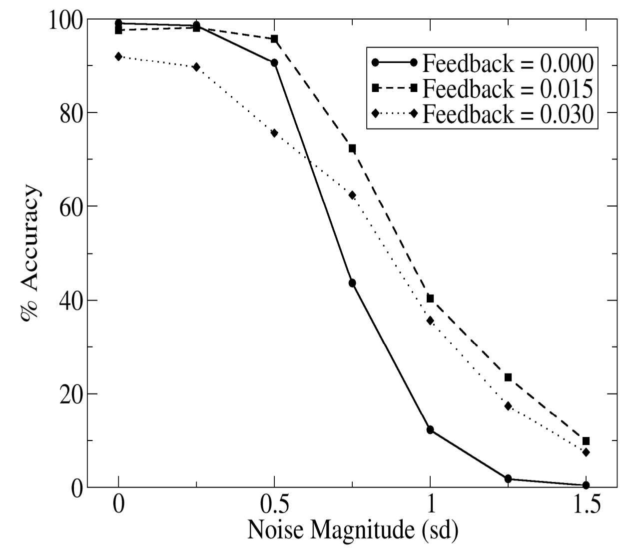
Pitfall 1: Material selection

- Frauenfelder & Peeters (1998)
- Feedback doesn't help
  - Half their items were recognized more quickly with feedback off
  - Feedback only allows TRACE to account for top-down effects?
- 21 items: 7 phones long, UP at phone 4

- Magnuson, Strauss, & Harris (2005)
- Tested 900 words
- With/without feedback
- Increasing levels of noise
- 73% of words recognized more quickly w/feedback

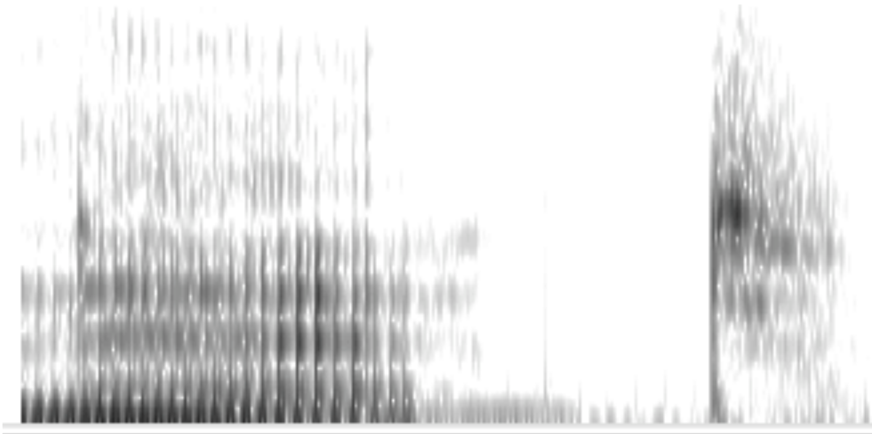






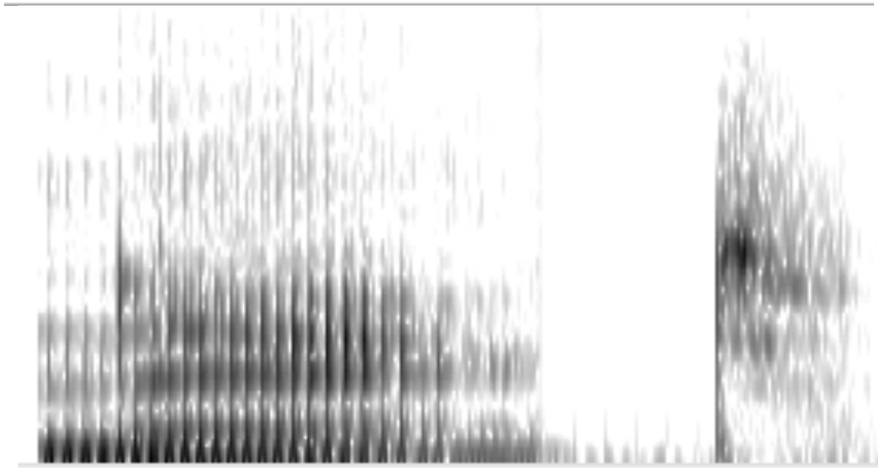
Pitfall 2: Material manipulation

- Model materials must be held to same standard as behavioral materials
- Material manipulations should have analogous effects
- Marslen-Wilson & Warren, 1994 (subcategorical mismatch)
- Hypothesis:
  - If there is lateral inhibition between words,
  - Then, if we provide misleading coarticulation consistent with a
    - **Word:** substantial lexical competition
    - **Nonword:** less competition



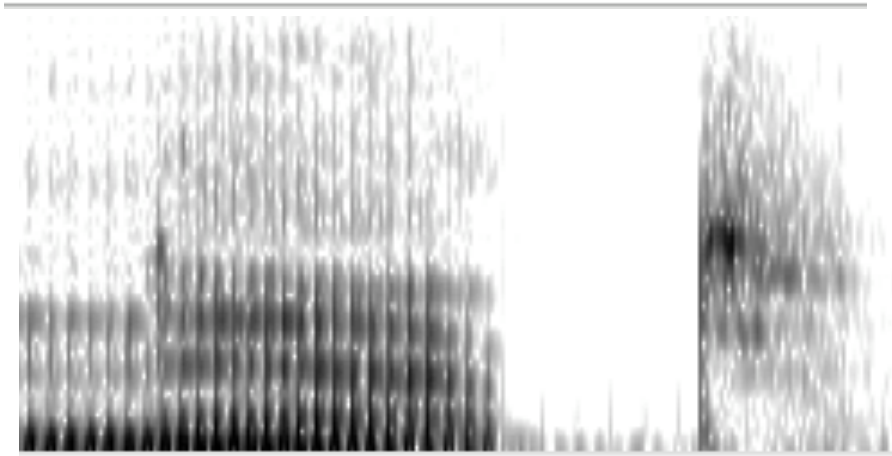
ne^{t}t

net spliced with **net**



ne^{k}t

neck spliced with **net**



ne^{p}t

nep* spliced with **net**

Pitfall 2: Material manipulation

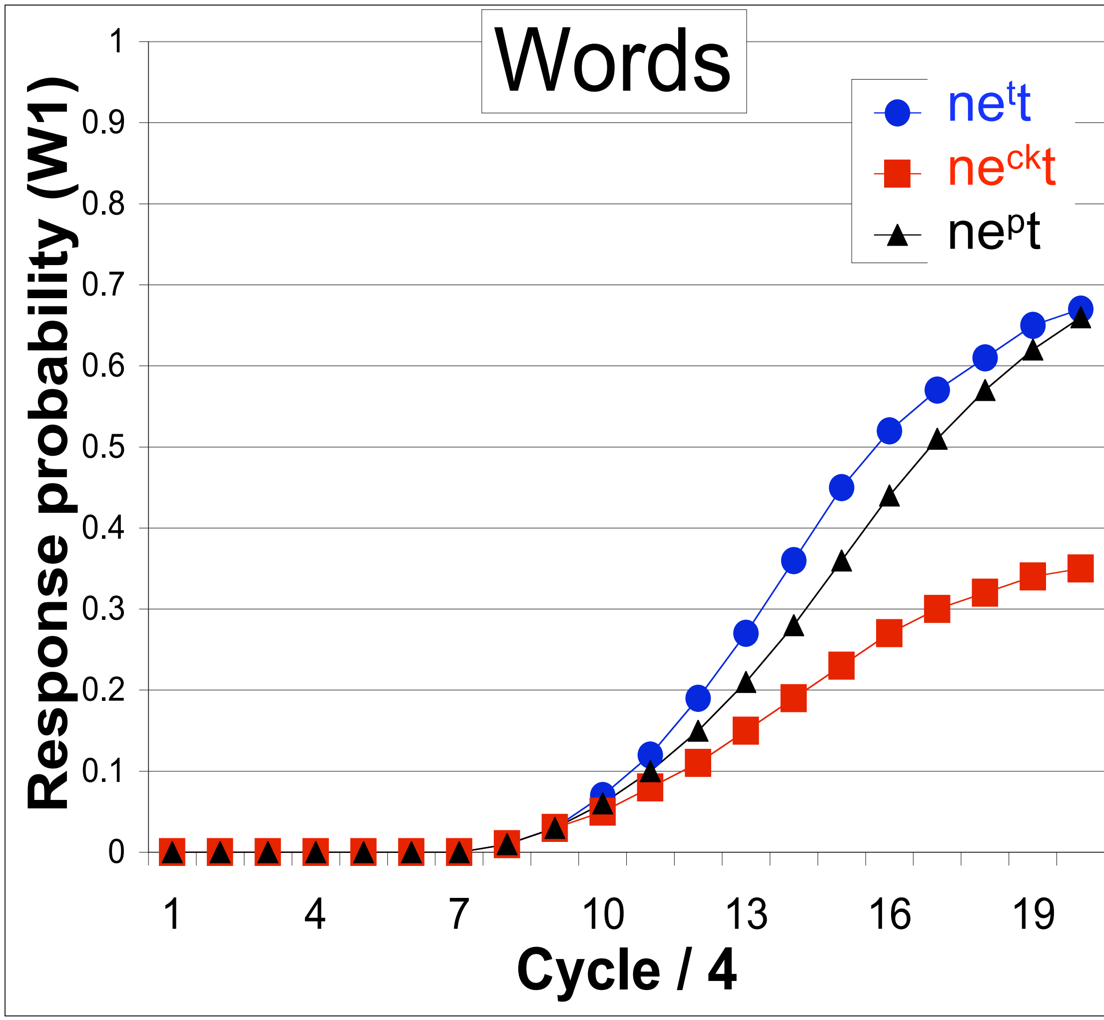
- Marslen-Wilson & Warren (1994): TRACE simulations

Lexical decision

ne^{t}t 487

ne^{p}t 609

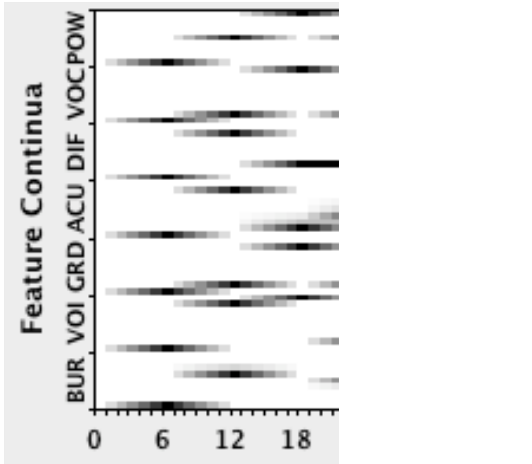
ne^{ck}t 610

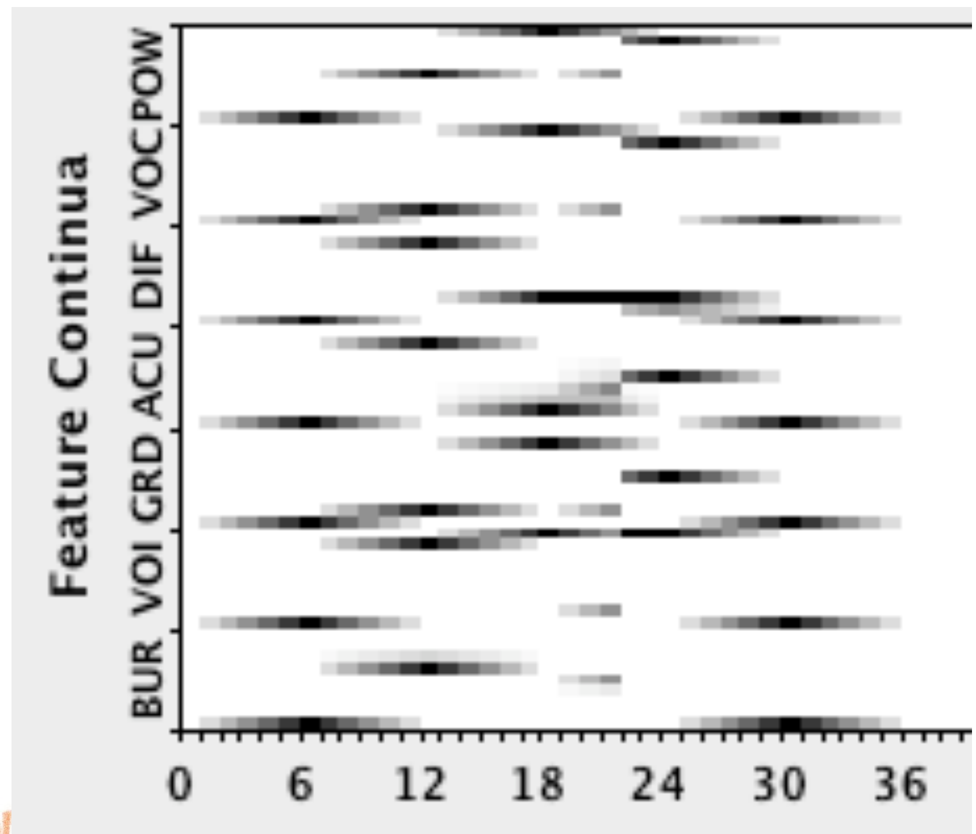


THE NEW SCHOOL

Haskins Laboratories

UNIVERSITY OF CONNECTICUT 1881



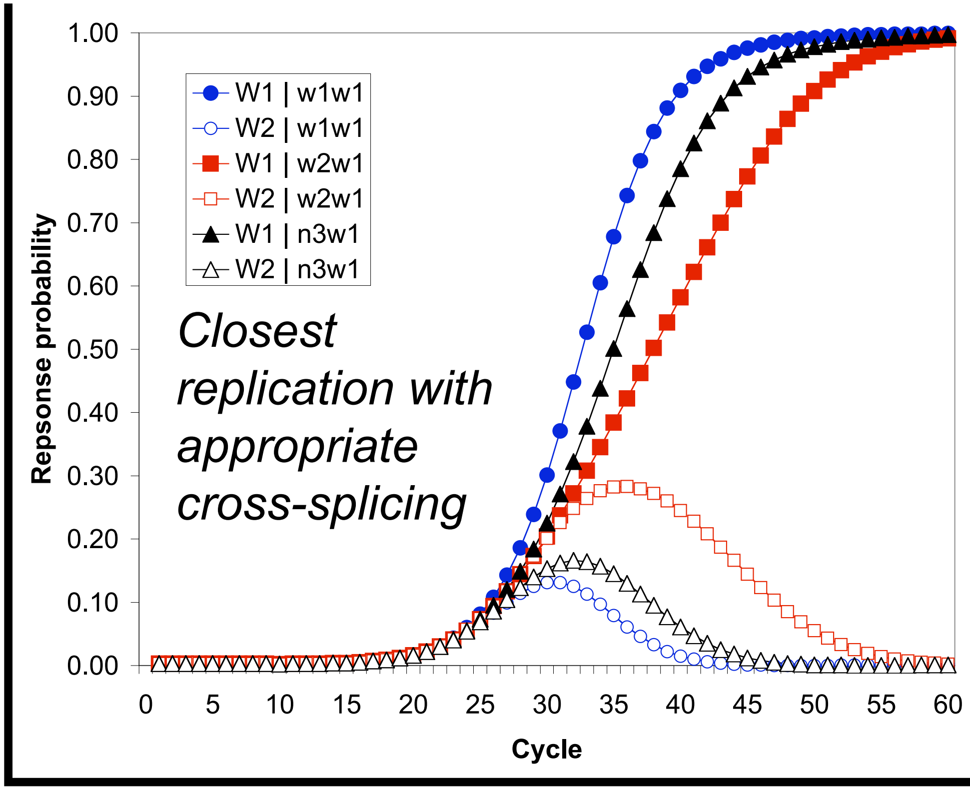


THE NEW SCHOOL

Haskins Laboratories

Pitfall 2: Material manipulation

- Original splicing was so late that “neck” was recognized instead of “net”!
- We cross-spliced at the latest slice where “net” was still ultimately recognized

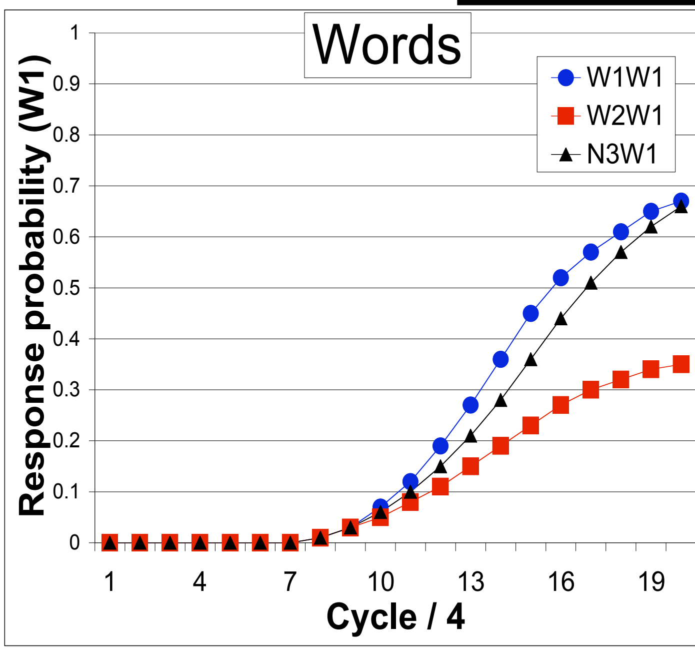


Lexical decision

ne^{t}t 487

ne^{p}t 609

ne^{ck}t 610



THE NEW SCHOOL

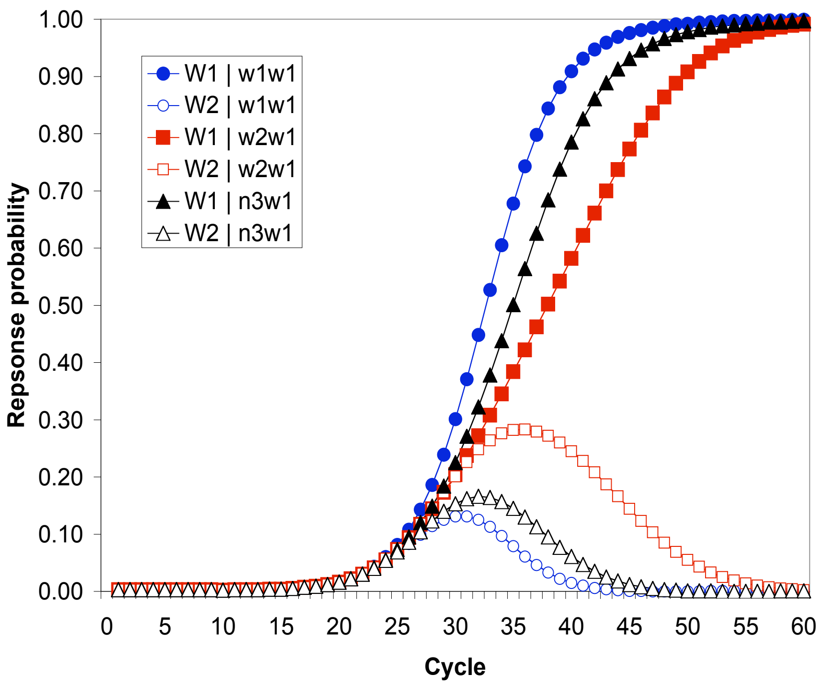
Haskins Laboratories

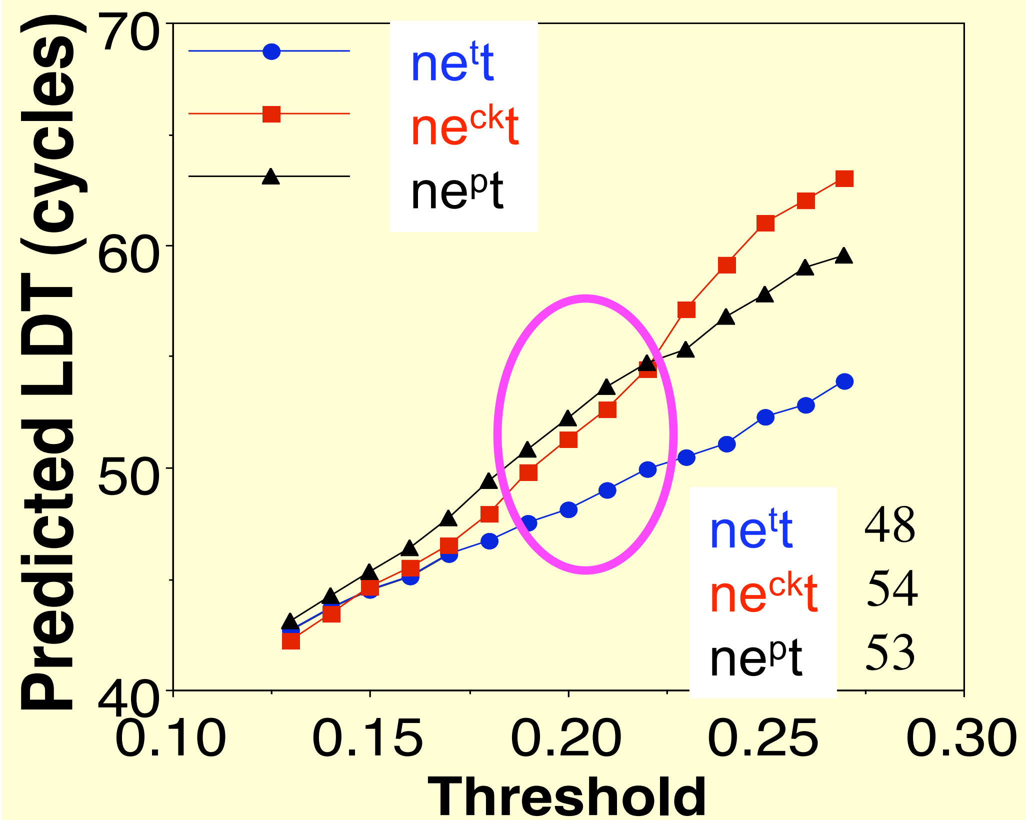
Pitfall 3: Linking hypotheses

- How can we reconcile TRACE activations and human LD RTs?
- Lexical decision need not be based on target
- What if we set threshold so that W2's activation sometimes passes?

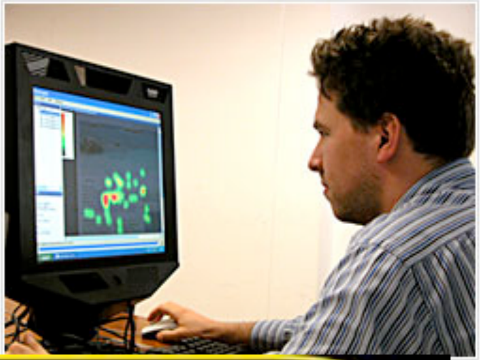
Lexical decision

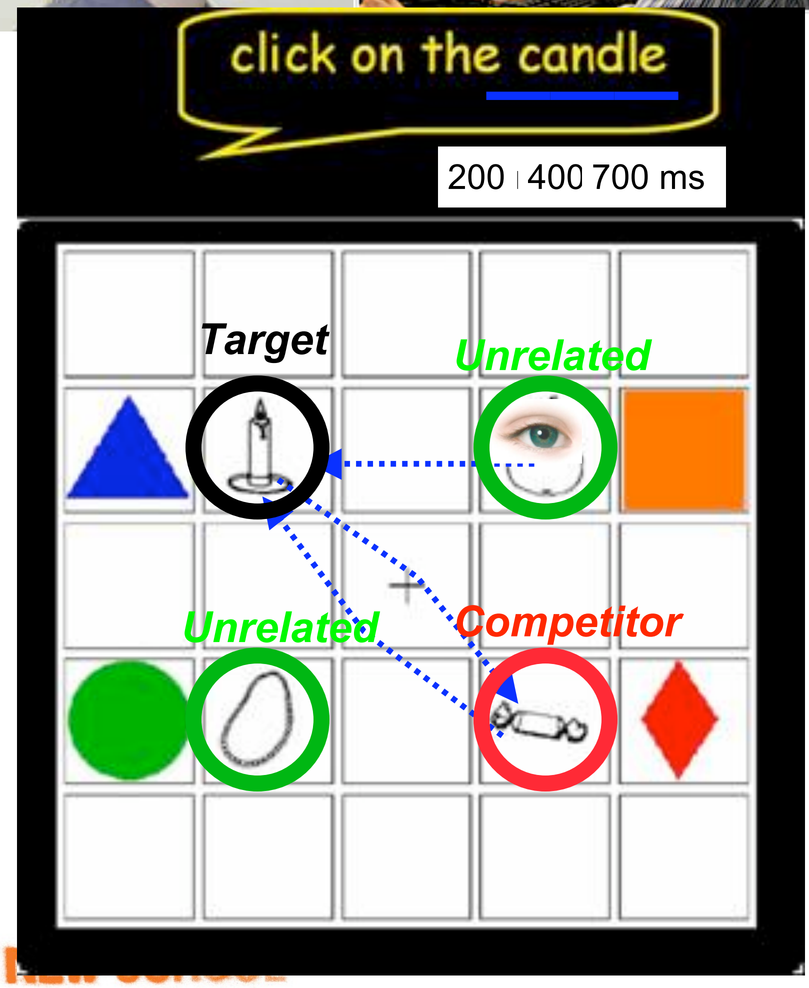
ne^{t}t	487
ne^{p}t	609
ne^{ck}t	610

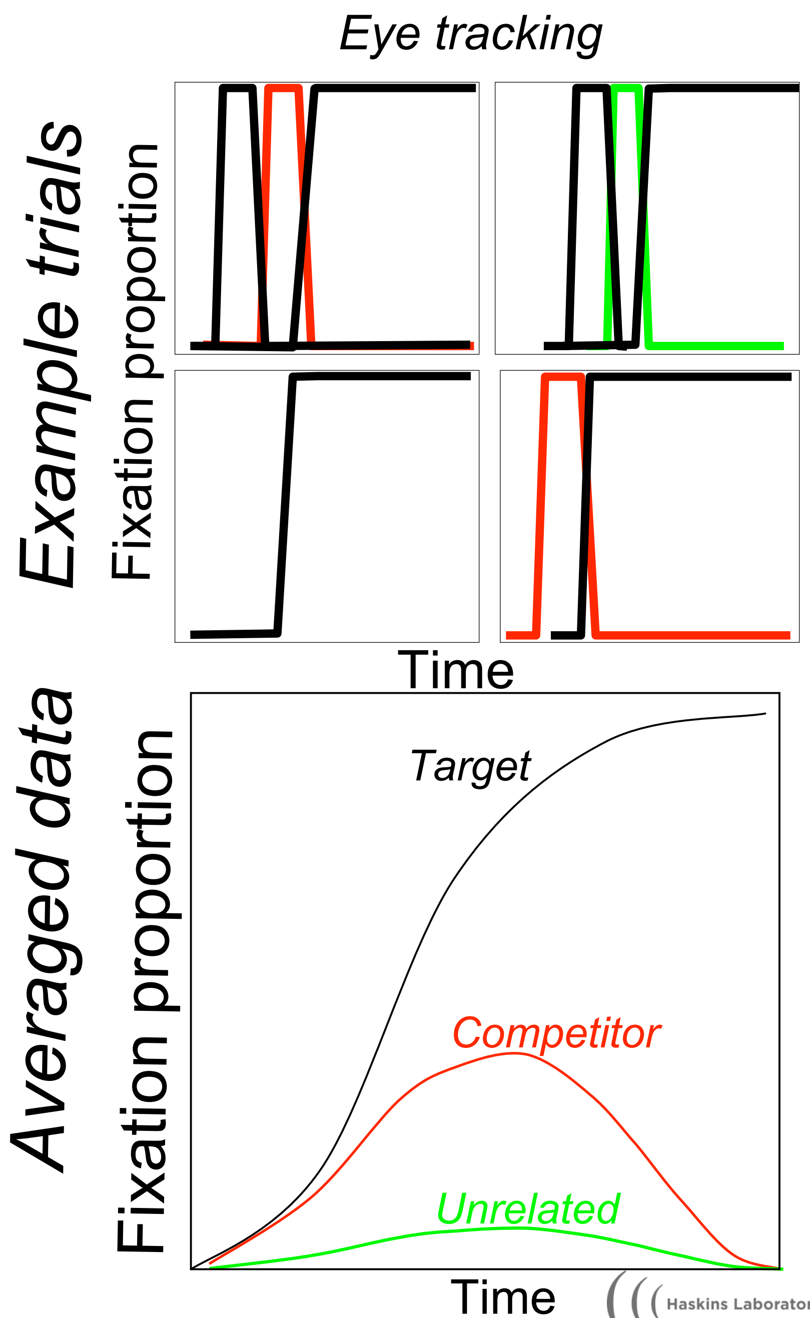




The “visual world” paradigm







Linking hypothesis

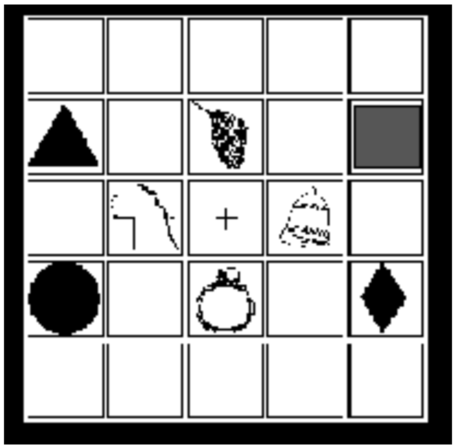
- Link fixation data to TRACE word activations
- Luce Choice rule: Activations and strengths based on entire lexicon; **choice** only includes the 4 onscreen items -- analogous to choice faced by subjects

$$S_i = e^{ka_i} \quad L_i = \frac{S_i}{\Sigma S_j} \quad \Longrightarrow \quad L_{target} = \frac{S_{target}}{(S_{target} + S_{comp} + S_{d1} + S_{d2})}$$

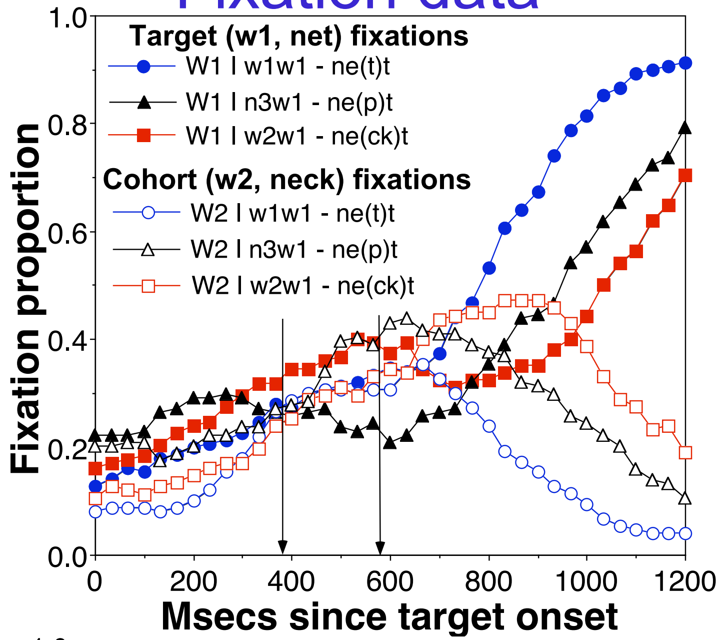
- Provides clear and testable predictions about number and nature of the items in the display
- See Dahan et al. (2001) for examples where this simple linking hypothesis accurately predicts changes in fixation proportions depending on display

Dahan, Magnuson, Hogan & Tanenhaus, 2001

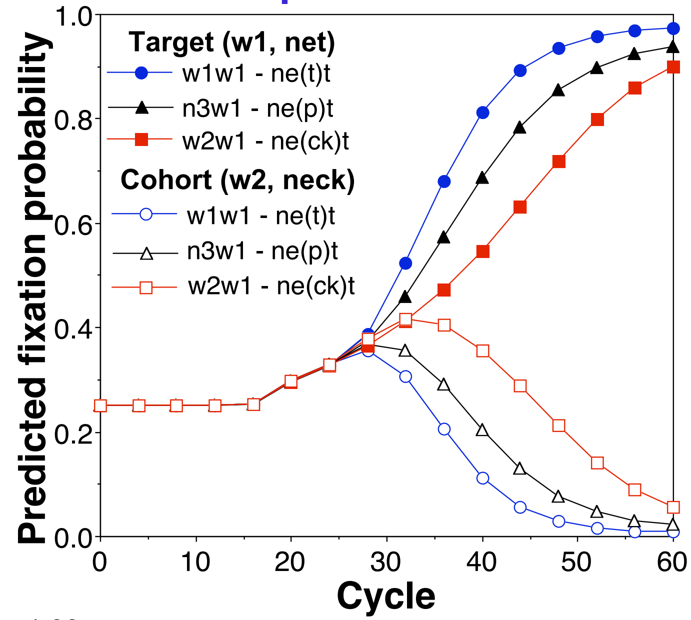
Cohort present



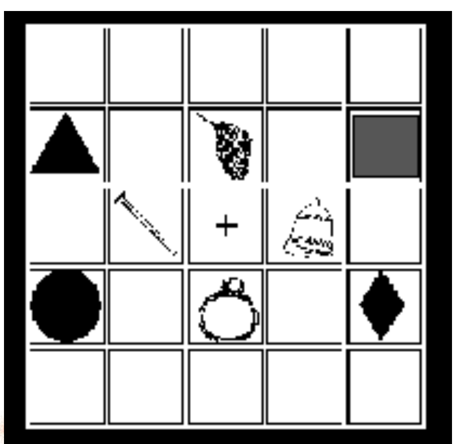
Fixation data

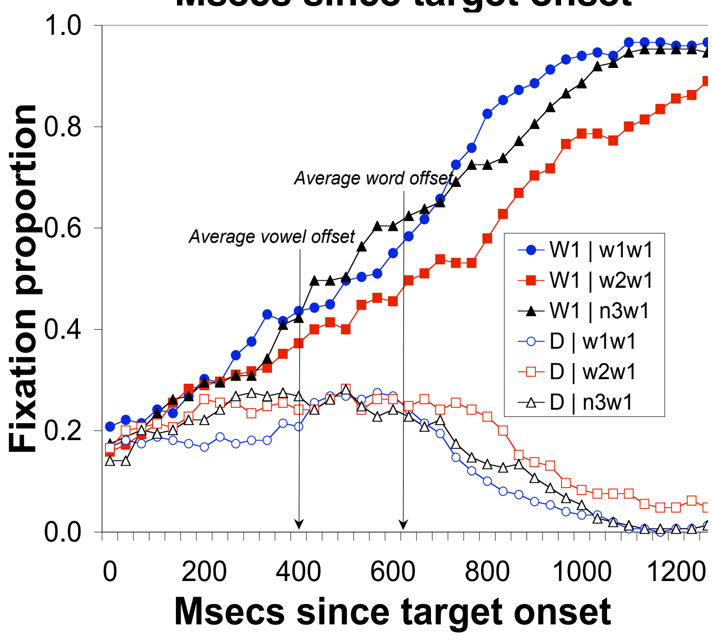


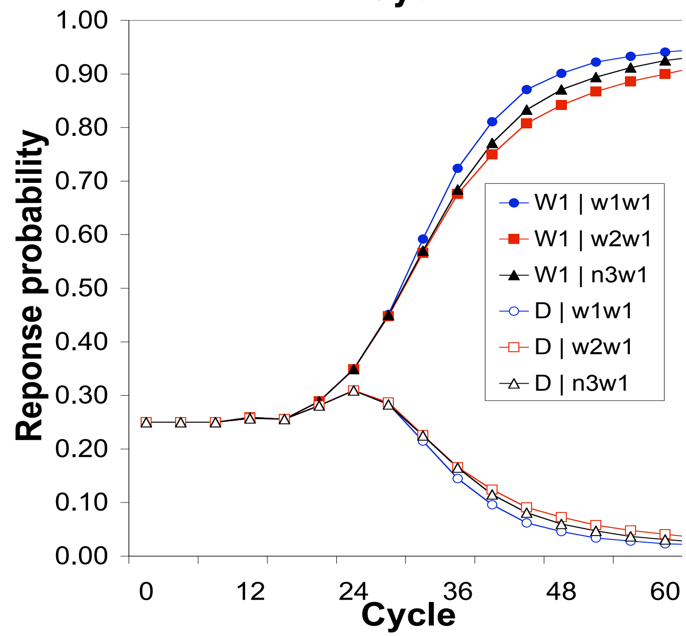
Model predictions



Cohort absent







Advantages of linking hypotheses

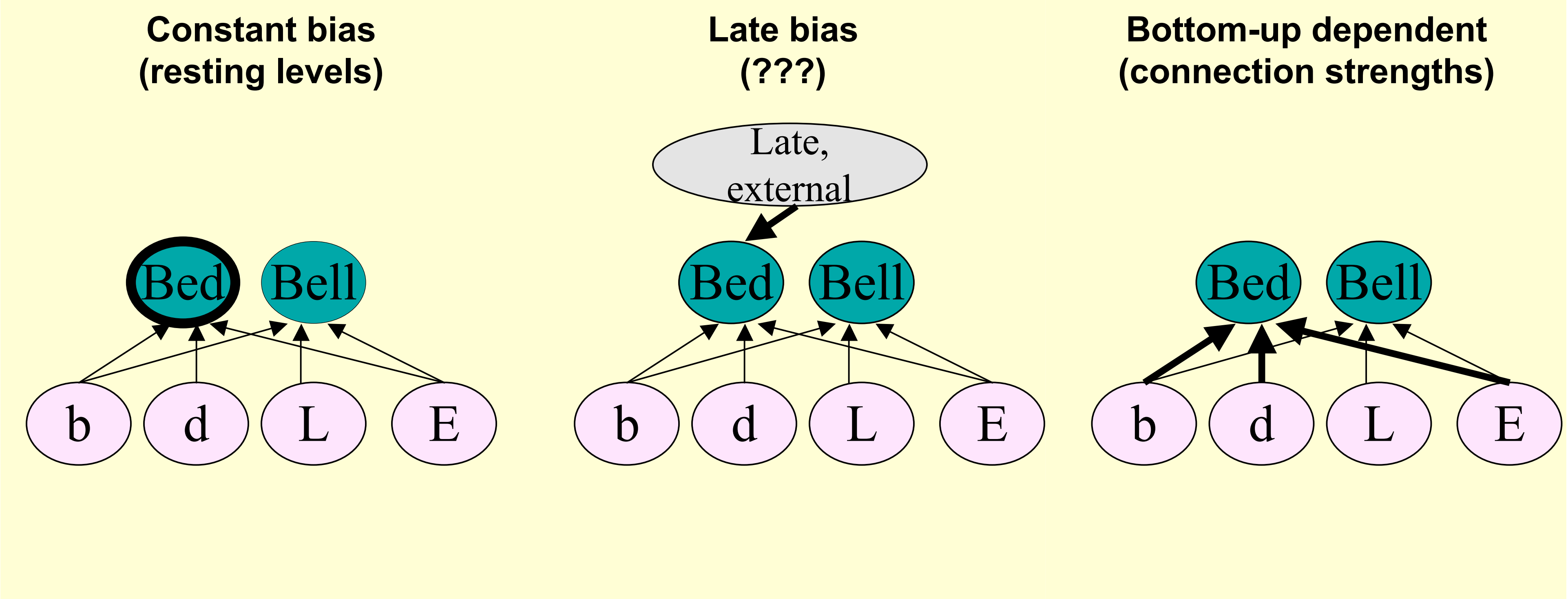
- Formalizes model of both internal state **and** overt behavior for specific tasks
- Also links cognitive systems -- e.g., visual attention and lexical activation
- Facilitates the interpretation of behavioral data from complex experimental paradigms
- Without careful, explicit modeling of task constraints via linking hypotheses, simple model activations can be misleading

Pitfall 4: Logic

- Sometimes, the most reasonable predictions turn out to be **wrong** <u>when you test them by simulation</u>
- Case in point: word frequency
- Assumptions about modeling: resting activation ≈ bottom-up connection strength
- Assumptions about empirical results:
  - – Absence of frequency effects in early responses
  - – Therefore, frequency is a late, top-down bias

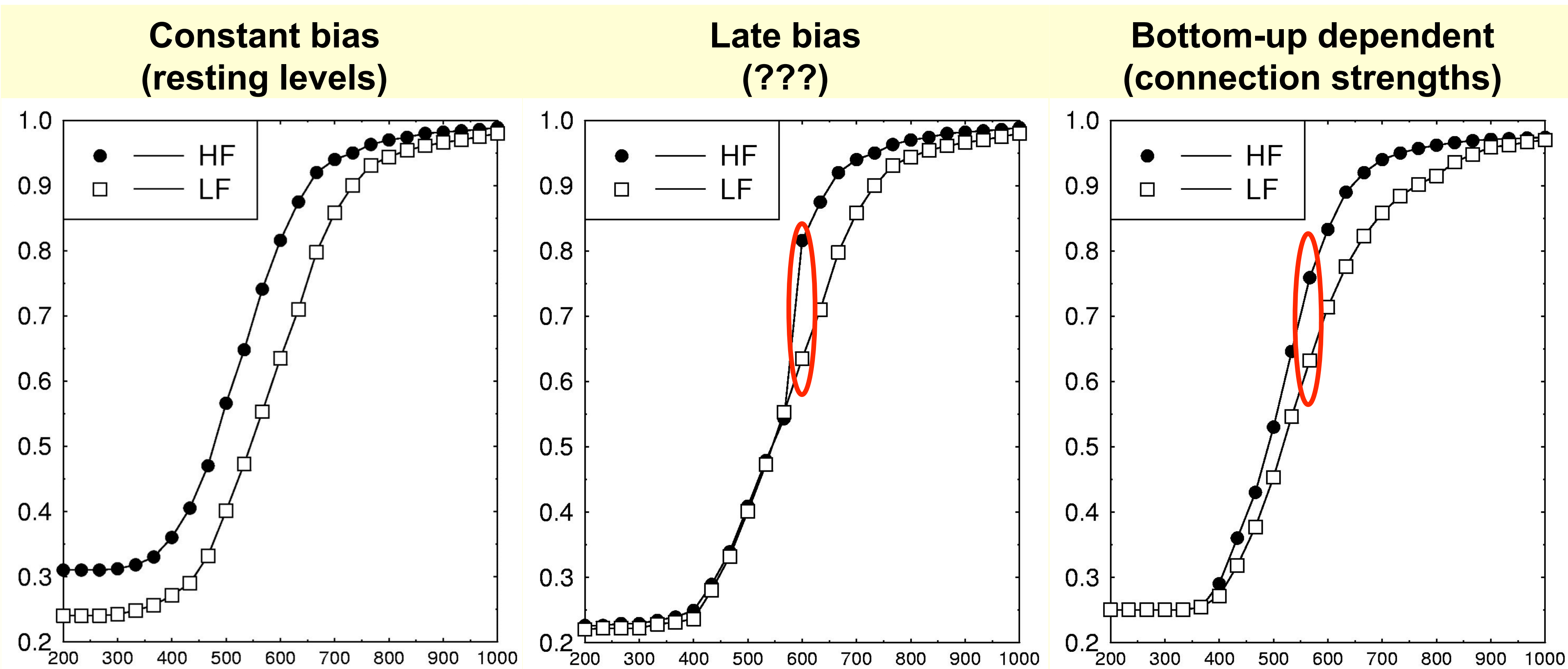
Proposed loci of frequency

- HF words recognized more quickly than LF words
- Early sampling (e.g., fast reactions) sometimes fails to detect frequency effects (Connine, Titone, & Wang, 1993)
- Conclusion: frequency is a late/2nd stage bias?



Time course implications

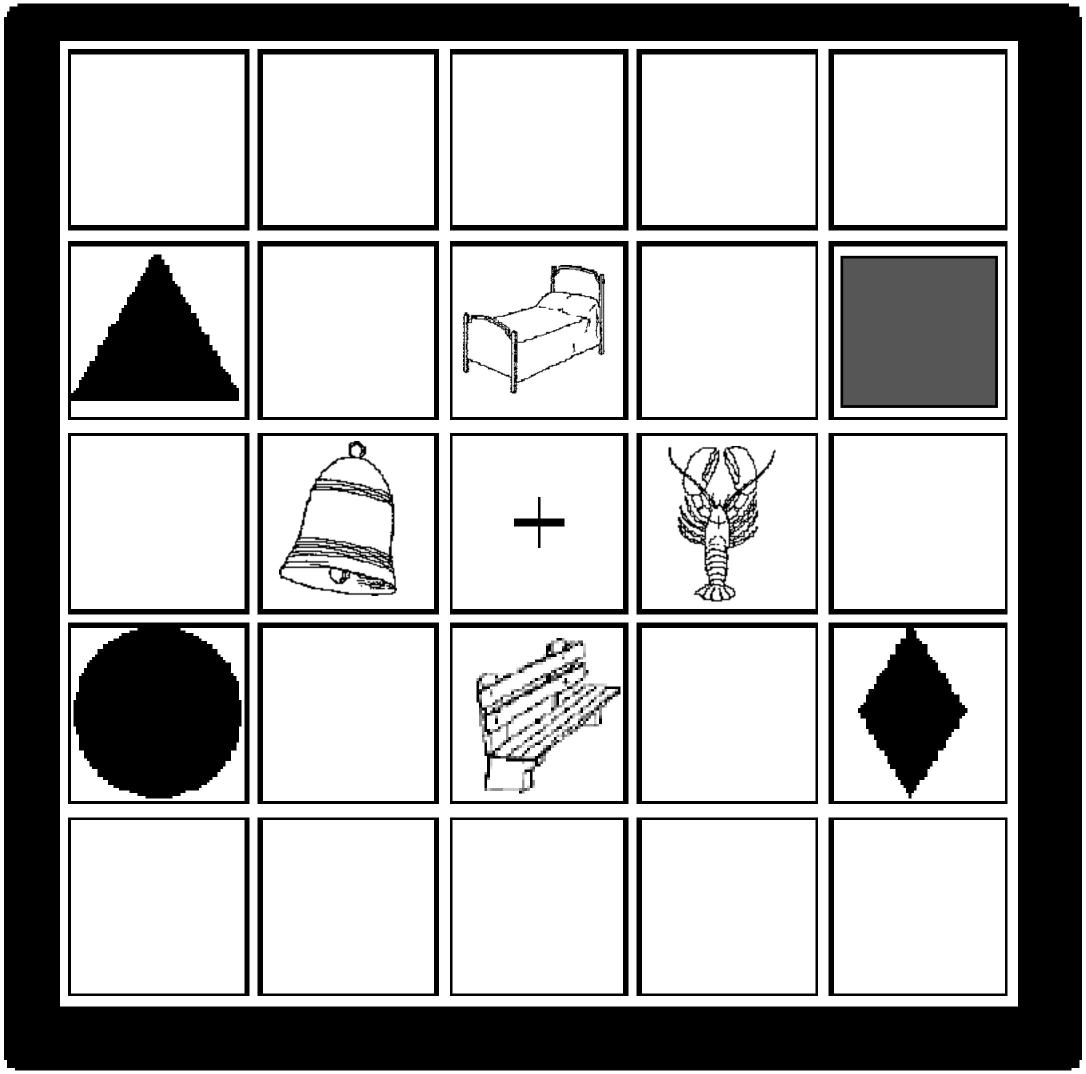
- Is a late frequency effect evidence for 2nd stage?
- No; consistent with bottom-up **dependency**



Cohort frequency

Sample critical display

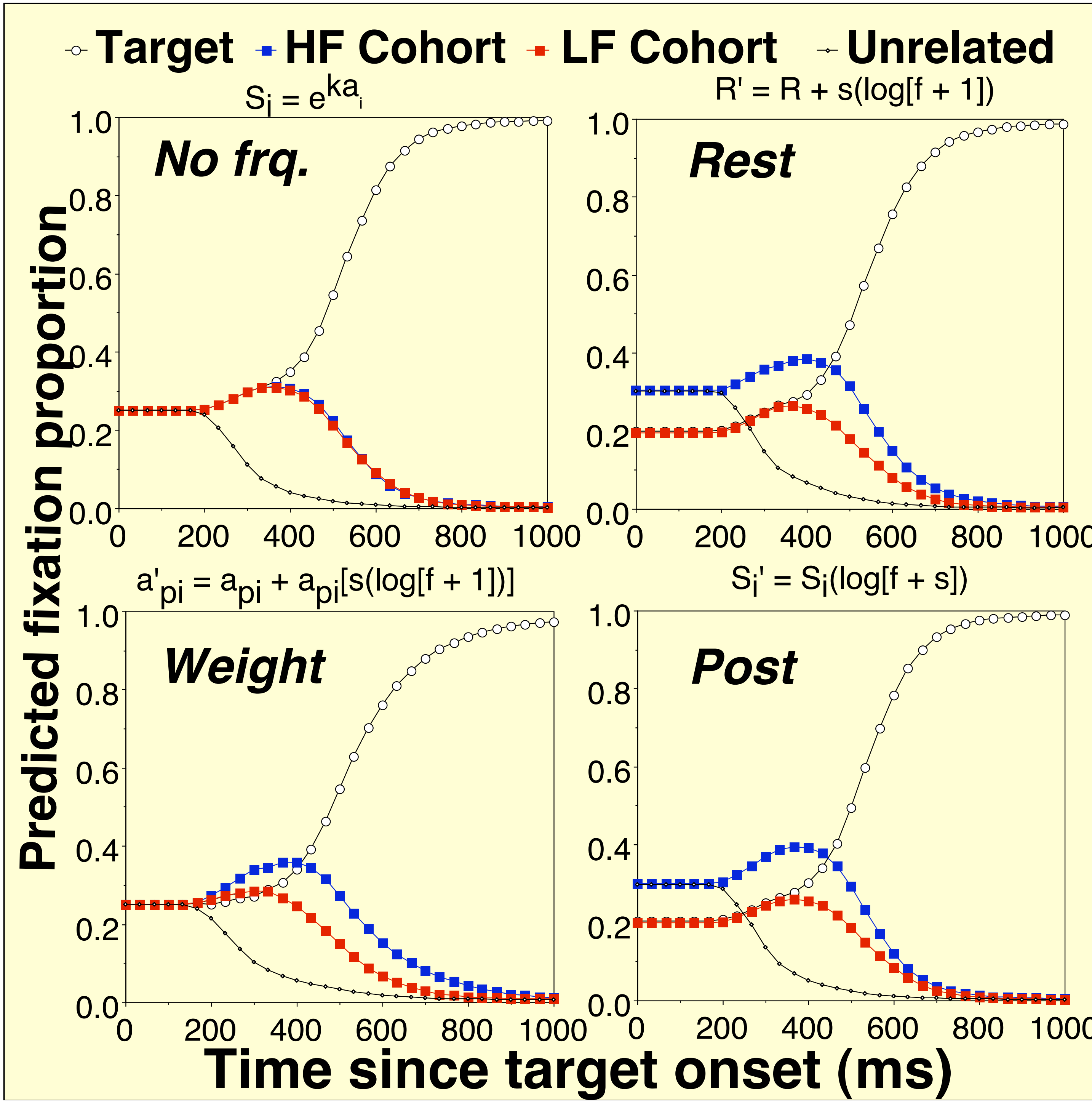
- LF target (bell)
- LF cohort (bench)
- **HF cohort (bed)**



THE NEW SCHOOL

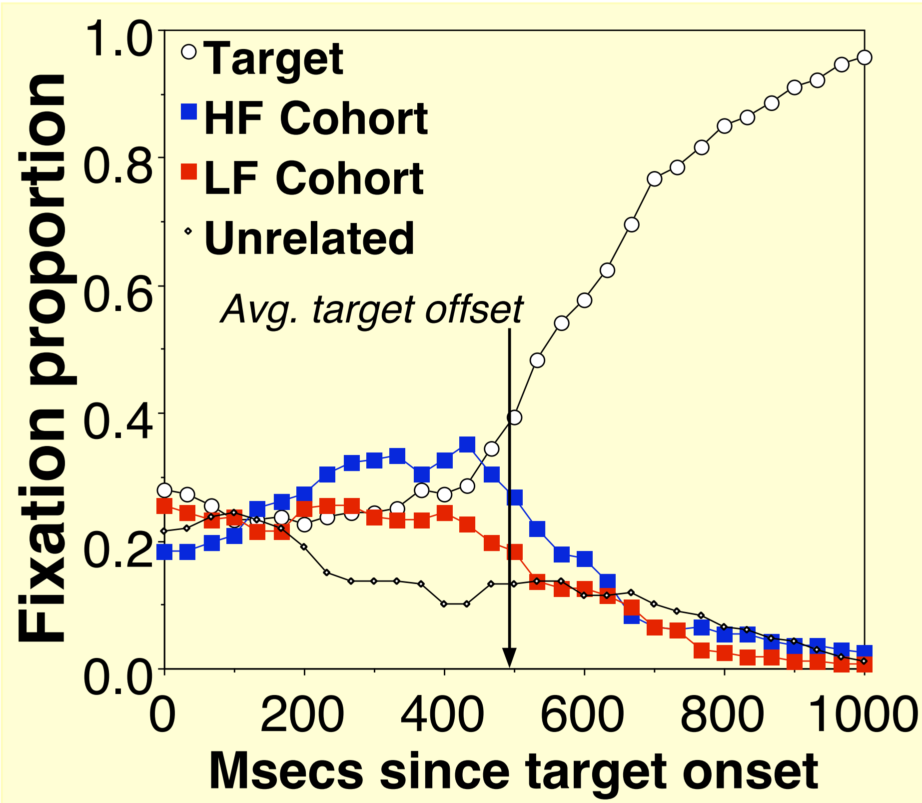
Haskins Laboratories

Model predictions

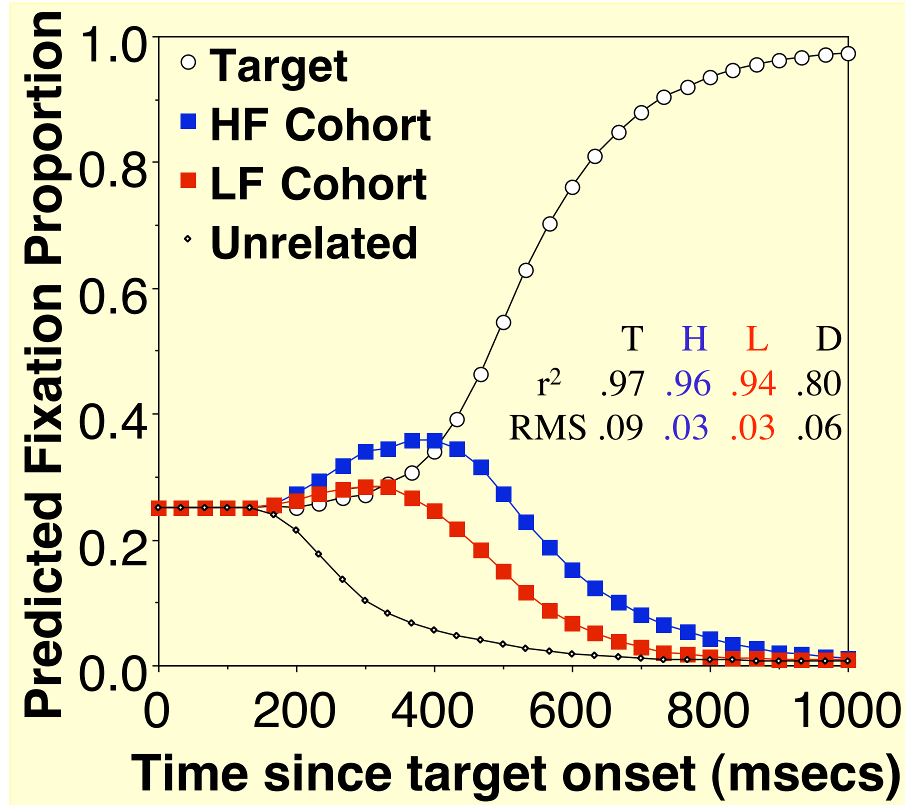


Model-data comparisons

Data



Model (weight version)



THE NEW SCHOOL

Haskins Laboratories

Gauging success and **failure**

- Given an apparent mismatch between model and data, you must avoid the pitfalls -- you must exclude:
  - Poor analogs to materials (representativeness, manipulations)
  - Insufficient linking hypothesis
  - Worst case: mismatch between *expectations* and data -- jTRACE was created to encourage more testing of logical expectations
- Then, if you still have a mismatch between model and data, you must determine the *level* of the failure
  - Parameter value?
  - Specific aspects of model mechanisms
  - Theoretical assumptions

Gauging **success** and failure

- Model flexibility, model fit
  - Fit measures: r^2, RMS error
  - Debates over whether fit is sufficiently constraining
  - Parameter space partitioning (Pitt et al., 2005)
- Comparing two models
  - Occam's razor
- More important?
  - Breadth

Using jTRACE

- Now what?
  - Work through the examples in the gallery (see documentation in ‘help’)
  - Make a plan for doing your own simulations
  - Email us if you need help (or find bugs, or have feature requests*)
  - Save your simulations! This is a great way to save yourself work and facilitate replication
- Right now: ‘lab time’