Fitting ACT-R Models with Trial-by-Trial Maximum Likelihood

Andrea Stocco, University of Washington
stocco@uw.edu

Fitting models

- We all do. That’s part for the job!

Fitting models

- We all do. That's part for the job!
- But what do we **fit for**?
- In most cases, we **minimize** *RMSE* or R^2
- Suggestion: We should use Maximum Likelihood (MLE)
- In linear models, minimizing RMSE and maximizing log-likelihood are the same
  - … and they both maximize R^2
- When we use **non-linear** models, however, things are different
- And ACT-R has several non-linear equations

What is MLE?

Find parameters θ of a model that maximize likelihood $\mathcal{L}$

$$\mathcal{L}(\underbrace{m}_{\text{model}}, \underbrace{\theta}_{\text{parameters}} \mid \underbrace{x}_{\text{data}}) = P(x \mid m, \theta)$$

In practice, you use **log**-likelihood, because probs become vanishingly small when there are series of products

$$\log \mathcal{L}(m, \theta \mid x) = \log P(x \mid m, \theta)$$

Why would you use log-likelihood?

Intuitively, that is what you are trying to do: Finding the most probable model. But, also:

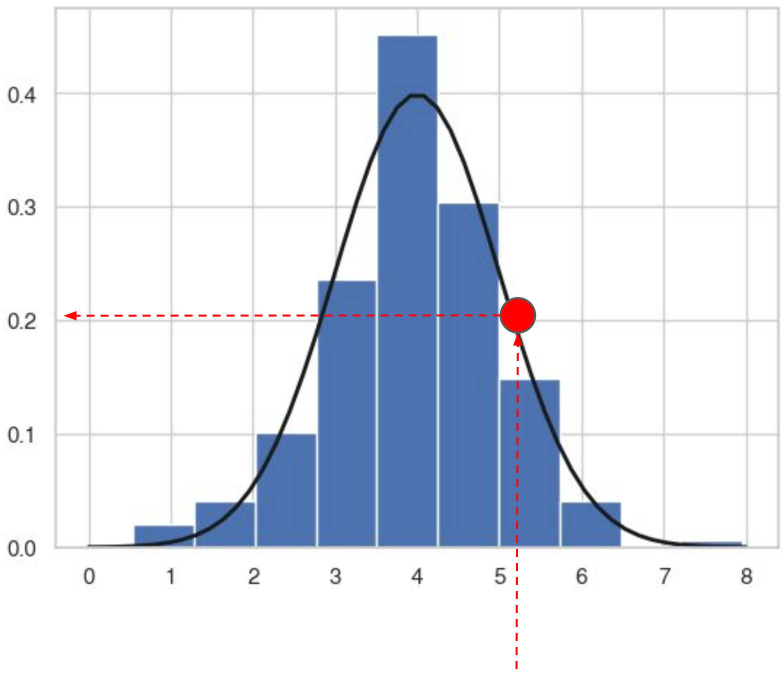
- It allows **comparison** across models with different complexity
  - BIC and AIC are expressed as a function of log likelihood:

$$\textbf{BIC} = k \log(n) - 2 \log \mathscr{L} \qquad \textbf{AIC} = 2k - 2 \log \mathscr{L}$$

- It allows fitting to **individuals** as well as **group** data
  - Group-level log likelihoods are the **sum** of individual log likelihoods!

How to do it – easy way

- Set your values for *m* and θ
- Run many simulations
- Calculate mean and standard deviation
- Compare to subject data point **x**



Participant data

Limits

ACT-R models often takes a long time to run

Necessary to run model many times to get stable data

Aggregated data often contains very few data points

Trial by trial

Trial by trial likelihood

$$\mathcal{L}(m, \theta \mid \mathbf{x}) = P\,(\mathbf{x} \mid m, \theta); \qquad \mathbf{x} = \{x_1, x_2, \ldots\, x_N\}$$

$$P(\mathbf{x}|m, \theta) = P(x_1|m, \theta) \cdot P(x_2|m, \theta, x_1) \cdot \ldots \cdot P(x_N|m, \theta, x_1, x_2 \ldots x_N)$$

ACT-R is a Markov model, and every choice is determined only by the current state.

So, if we force the model to follow the choices:

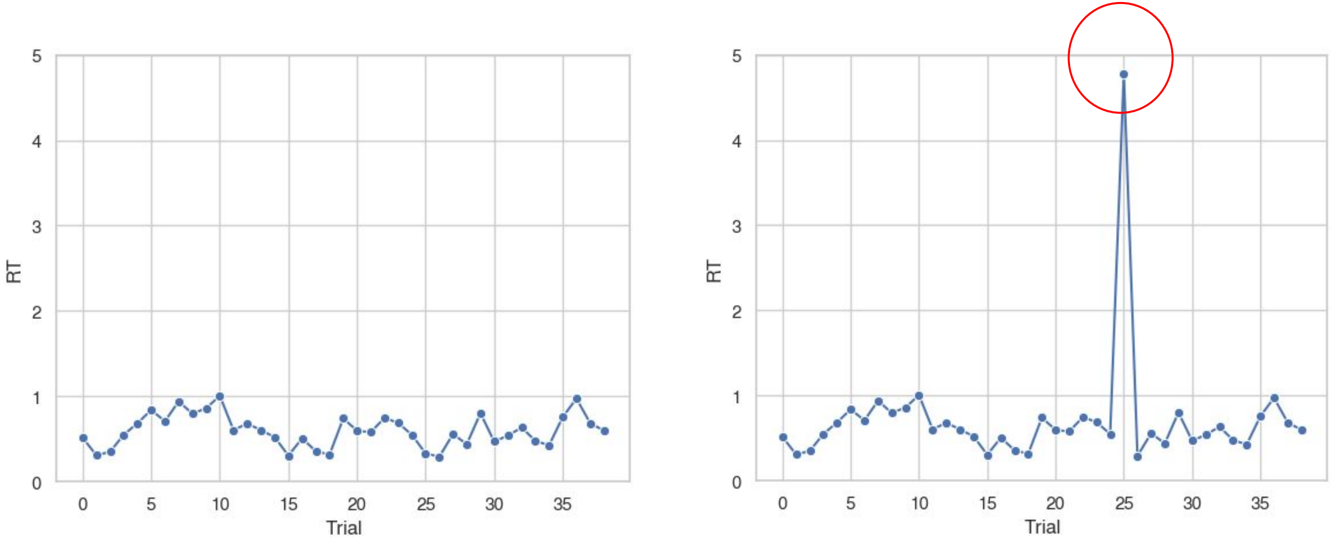
$$P(\mathbf{x}|m, \theta) = P(x_1|m, \theta) \cdot P(x_2|m, \theta) \cdot \ldots \cdot P(x_N|m, \theta)$$

$$\log \mathcal{L} = \textstyle\sum_i \log P(x_i|m, \theta)$$

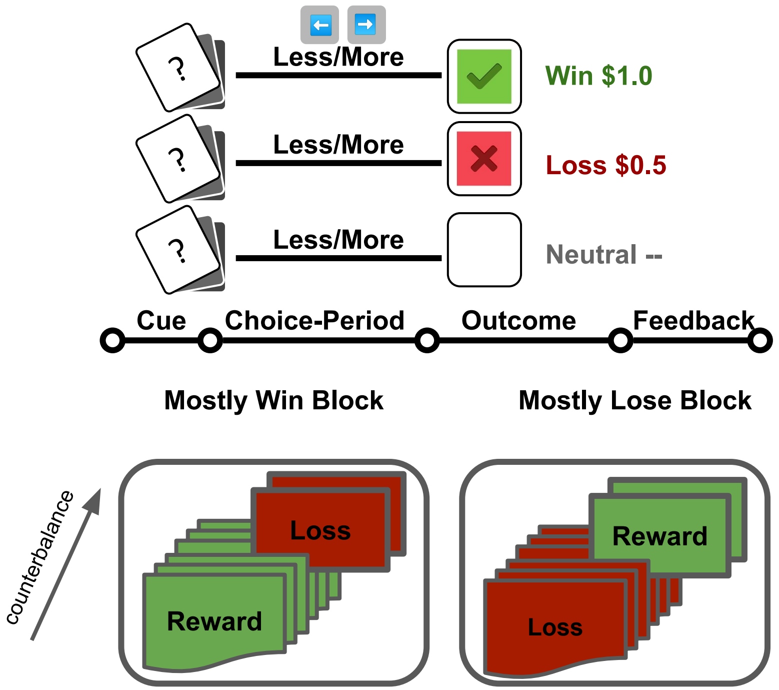
This is just **model tracing**! (Koediger & Anderson, 1993)

Advantages of trial by trial data

- You get **more data points** for every individual
- Aggregated data can be deceiving:



Example: Incentive Processing Task

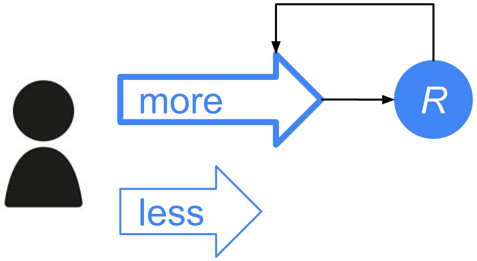


CONNECTOME COORDINATION FACILITY

- *N* = 199 participants
- **2 runs** for each participant
- **4 blocks** per run (2 Win, 2 Loss)
- **8 choices** per block
- **64 trials** total

Two ways to approach the task

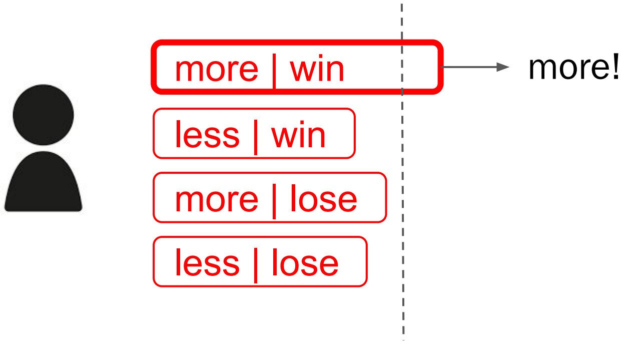
Procedural Memory



$$U_t(p) = U_{t\text{-}1}(p) + \alpha[R - U_{t\text{-}1}(p)]$$

$$P(more) = e^{U(more)/T} \;/\; e^{U(more)/T} + e^{U(less)/T}$$

Declarative Memory



$$A_t(c) = \Sigma_i\,(t - t_i)^{-d}$$

$$P(more) = e^{A(more)/s} \;/\; \Sigma_{chunk} e^{A(chunk)}$$

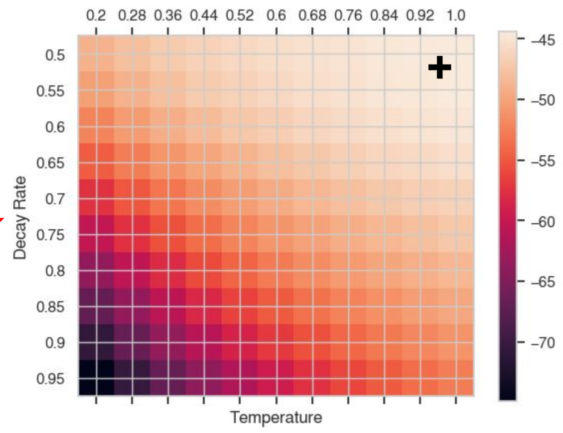
Model-Based Group Assignment

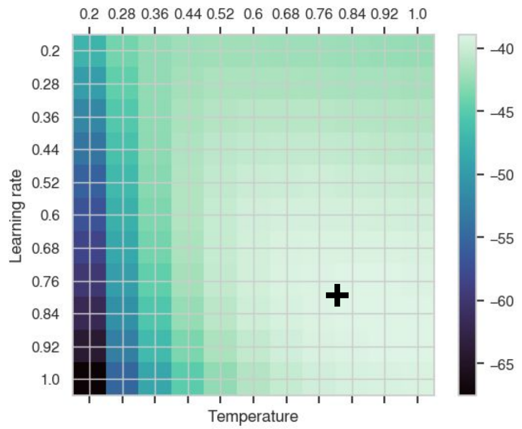
Implemented equations in Python

For every participant:

- Use Powell's method to maximize **Declarative** log-likelihood across parameters (θ)
- Repeat for **Procedural**
- Assign participant to the model with greater likelihood

Total runtime: ~ 20 mins!

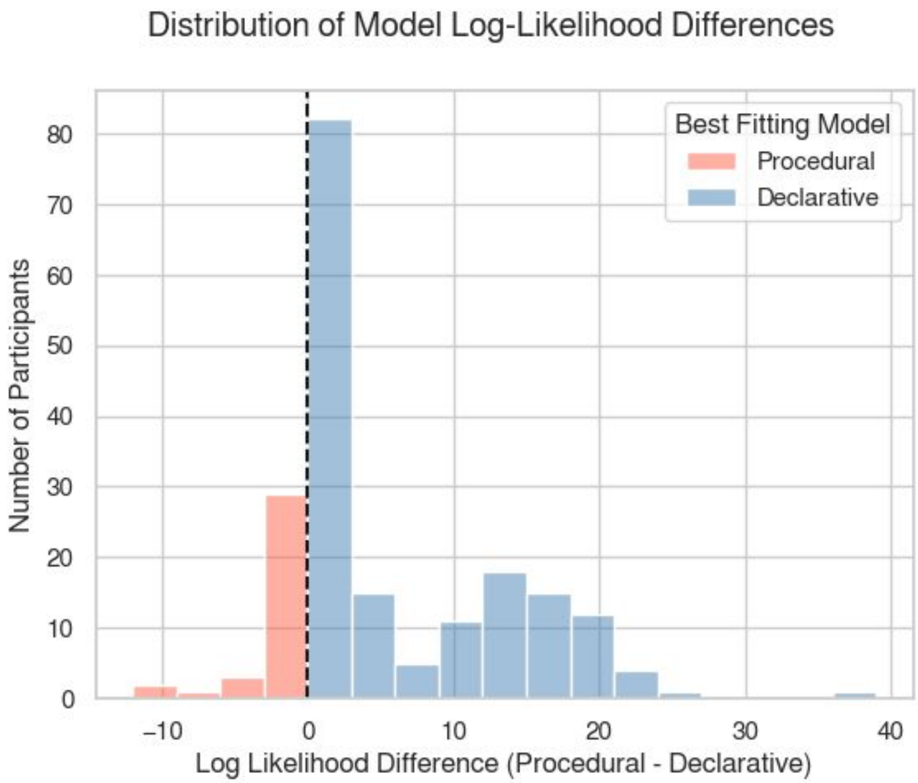




https://github.com/UWCCDL/ProcVsDecl/

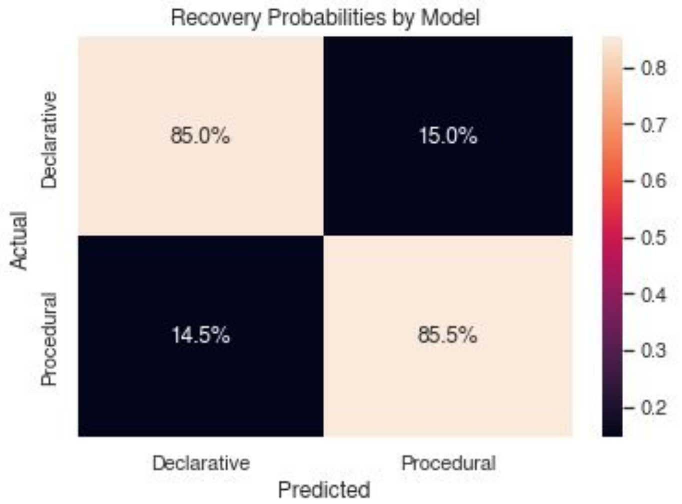
Yang et al, *BioArXiv*

Participant assignments

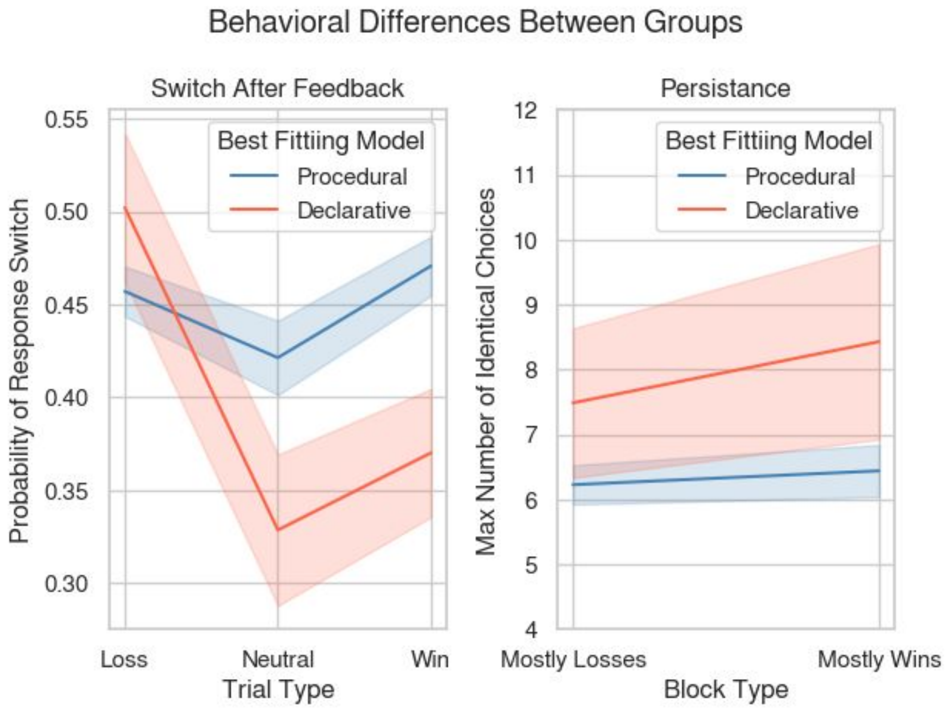


How Reliable are our Models?

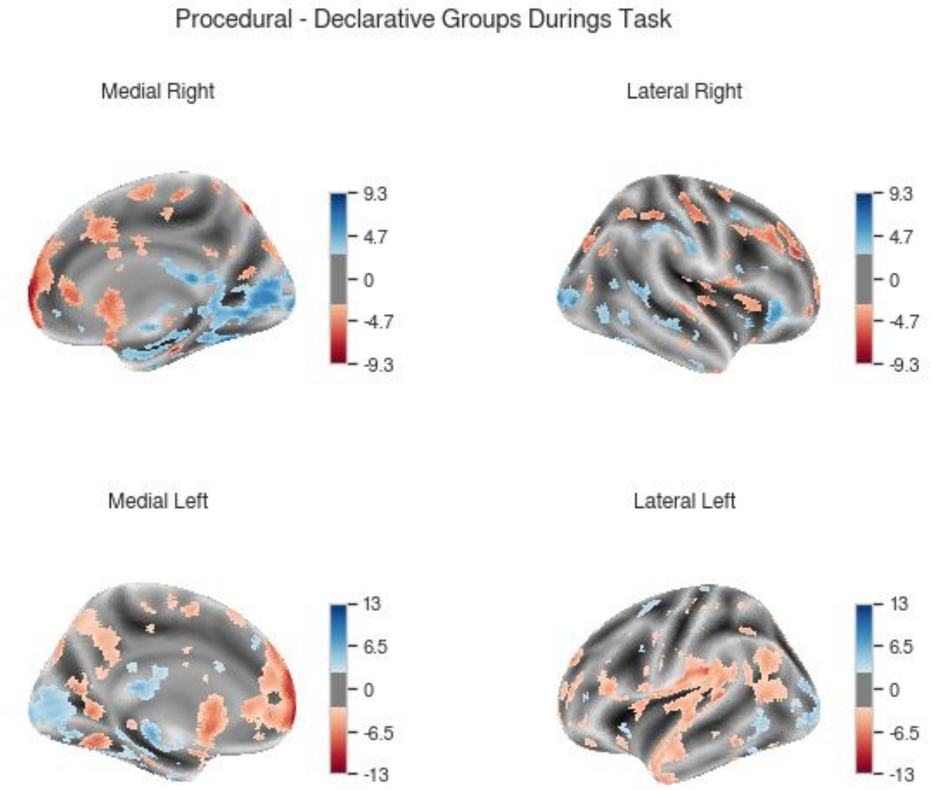
- Generated 20,000 simulated runs for each model
  - with random initial params
- Applied trial-by trial MLE to recover model



Differences btw Declarative and Procedural groups



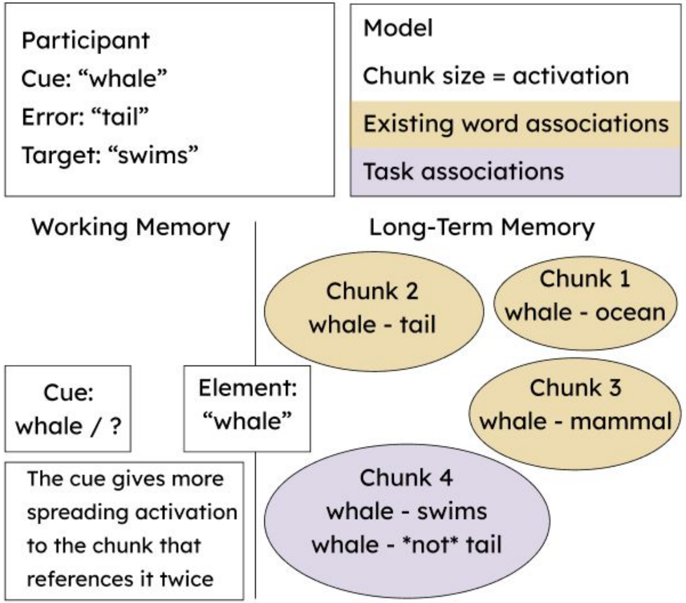
Differences btw Declarative and Procedural groups



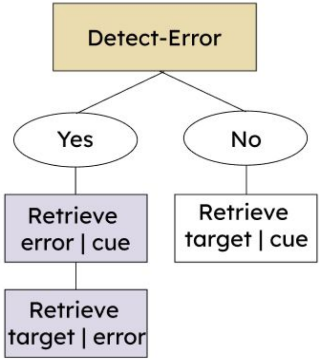
Mixing different measures

Better memory after errors

Elaborative Hypothesis



Mediator Hypothesis



Mediator predicts longer RTs!

Leonard et al., 2022 *CogSci*

Mixing different measures

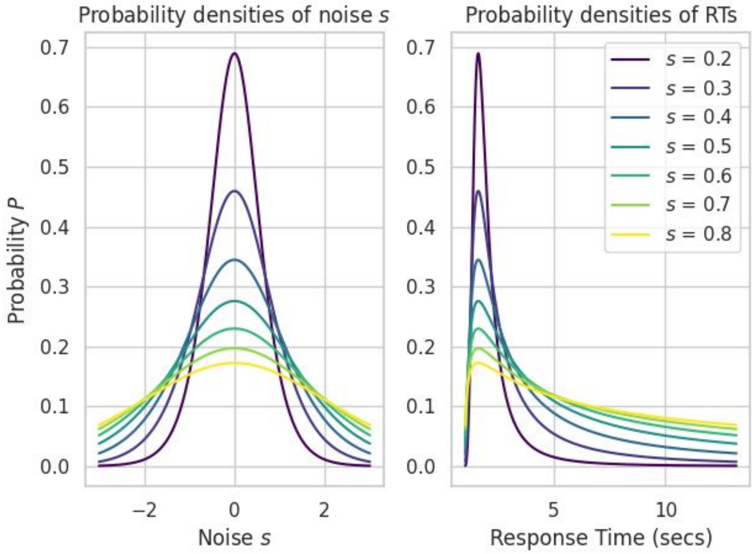
$$A_t(c) = \sum_i (t - t_i)^{-d} + s$$

$$P(c) = e^{A(c)/s} \,/\, \sum_i e^{A(i)/s}$$

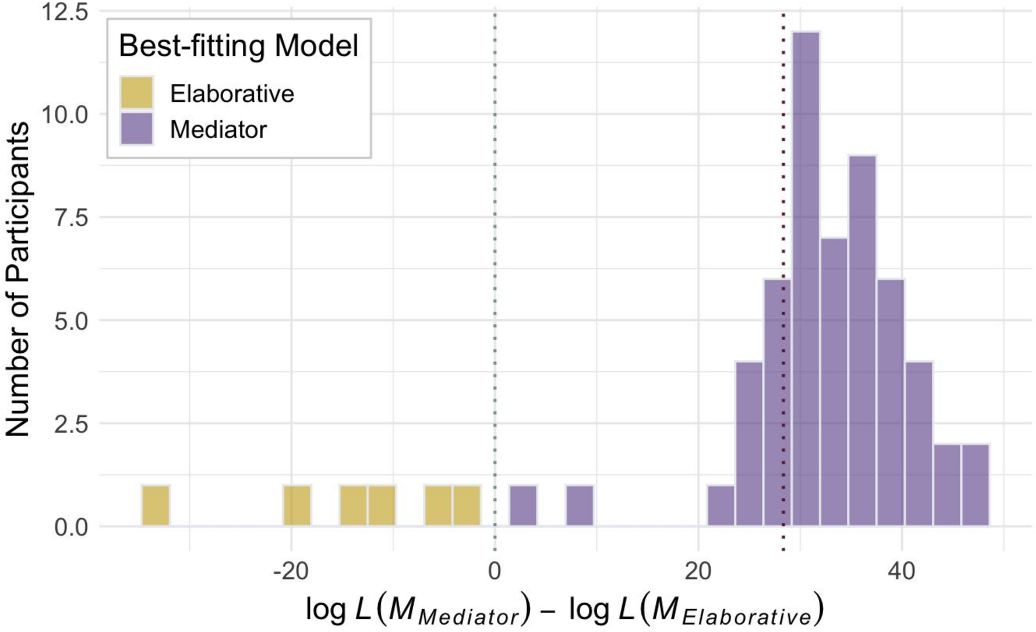
But ACT-R also makes **predictions about RTs:**

$$rt = t_0 + F\,\mathrm{e}^{A(c)}$$

$$P(\mathbf{x}_i|m, \theta) = P(c_i|m, \theta) \cdot P(rt_i|m, \theta)$$



Mediator vs Elaborative



Evidence for Mediator = sum of individual ΔLL = 1,728. **Mediator is $e^{1{,}728}$ more likely**

Accurate Parameter Recovery

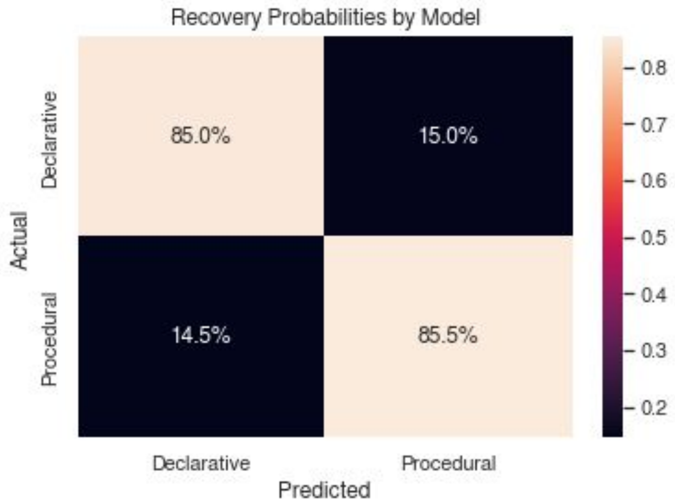
Accurate parameter recovery

In my lab, we make a big deal about understanding individuals using parameters

But to make sense, these parameters need to be **accurate**

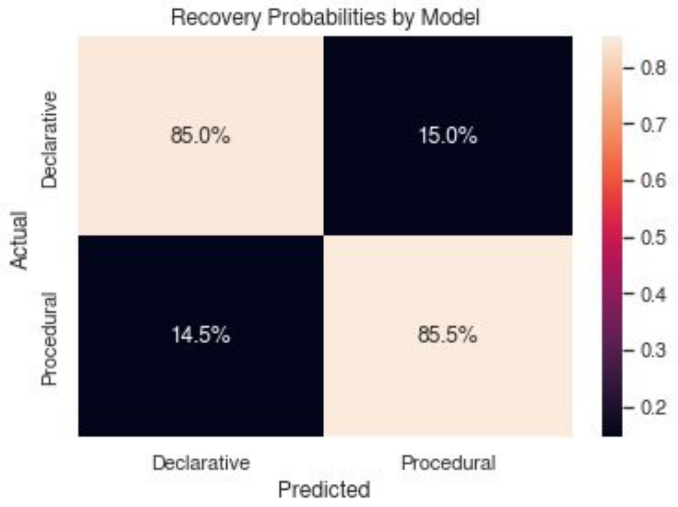
How Reliable are our Models?

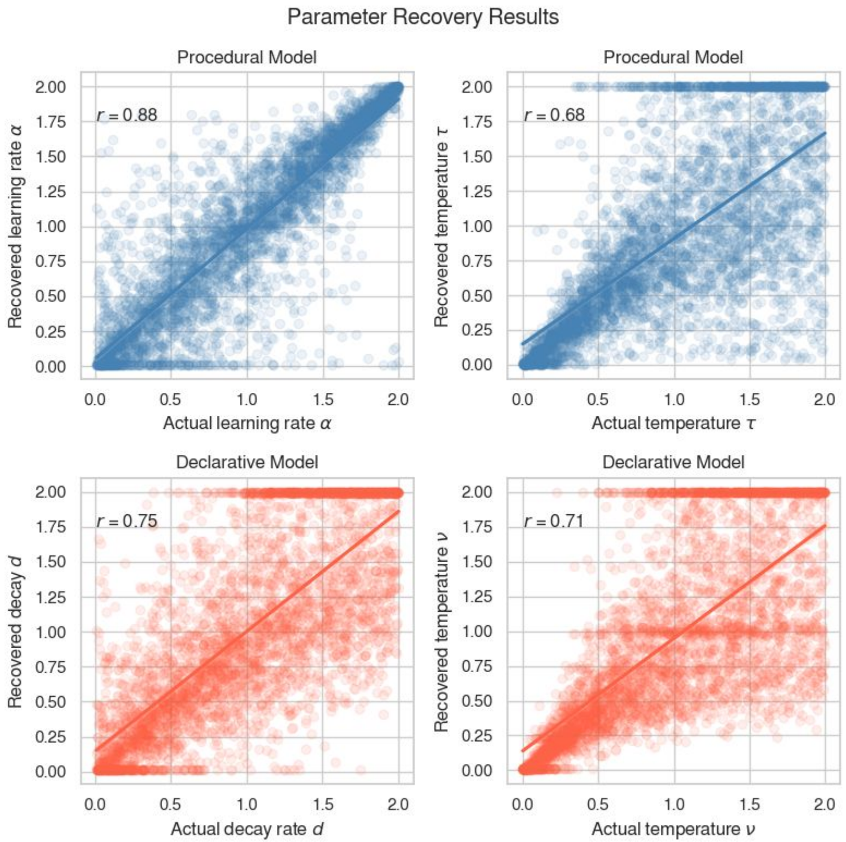
- Generated 20,000 simulated runs for each model
  - with random initial params
- Applied trial-by trial MLE to recover model



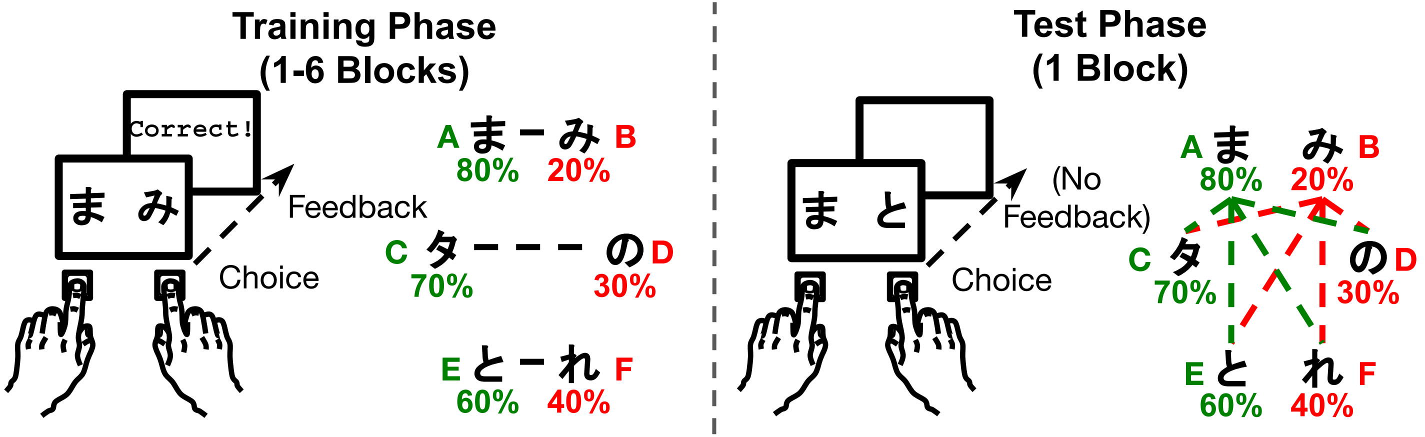
How Reliable are our Models?

- Generated 20,000 simulated runs for each model
  - with random initial params
- Applied trial-by trial MLE to recover model



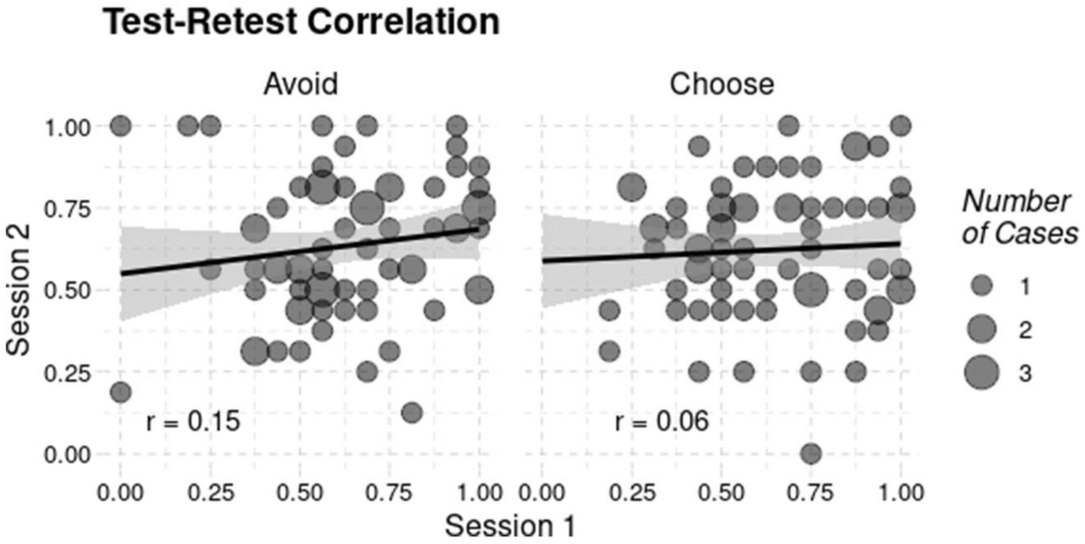


PSS Task (Frank et al., 2004)



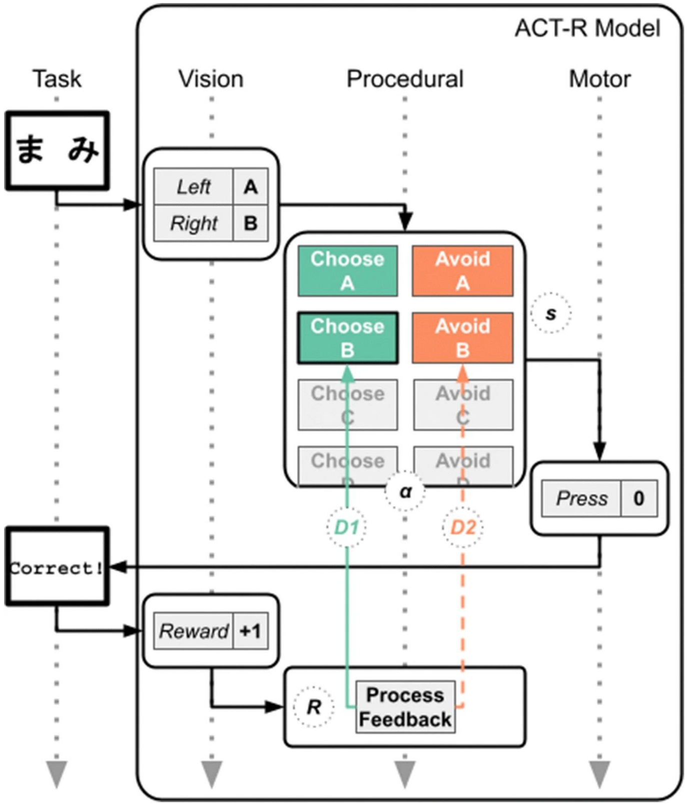
ChooseA and AvoidB are proxies for D1/D2 dopamine receptors

Poor reliability

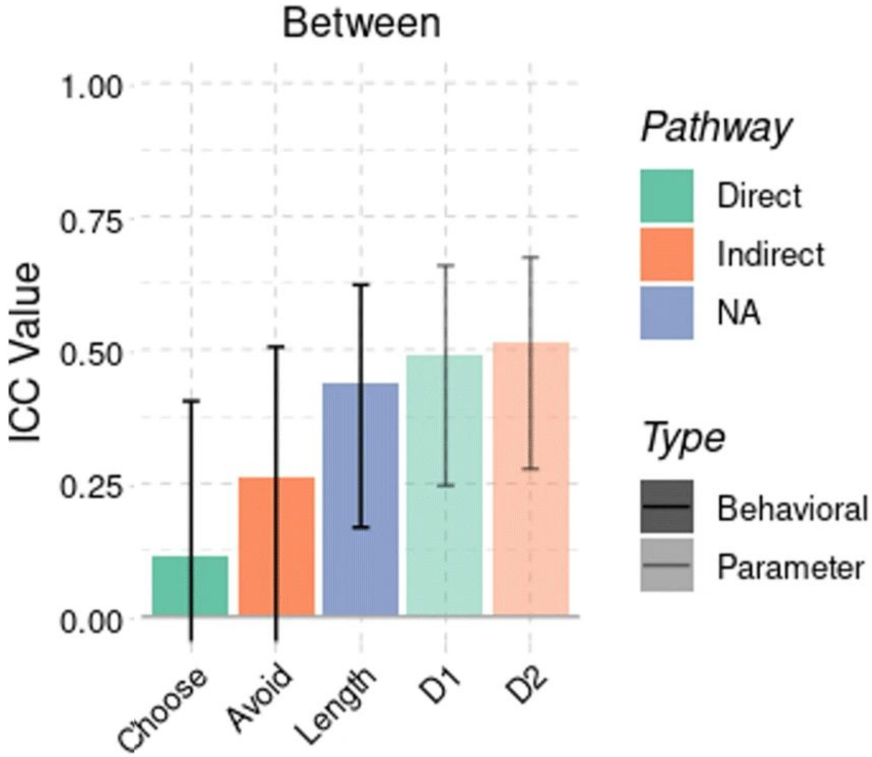


Xu & Stocco, 2021, *Comp. Brain Behav.*

Including D1/D2 parameters in ACT-R model



Parameters have greater reliability!



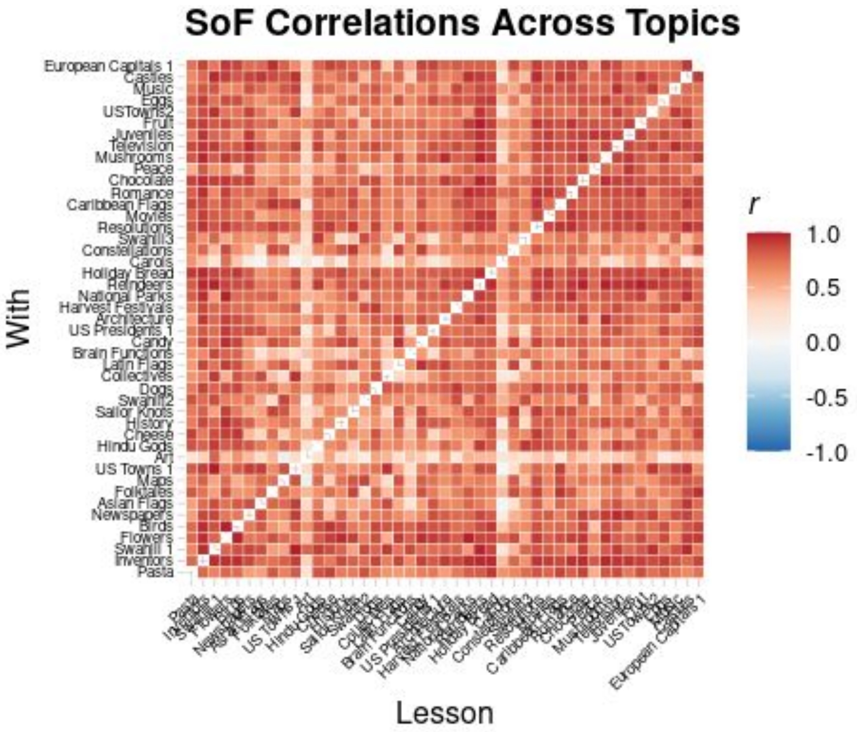
Tracking memory decline

Long-running study to track memory decay in 47 elderly individuals

Really, **α** param in Pavlik & Anderson ("Speed of forgetting", **SOF**)

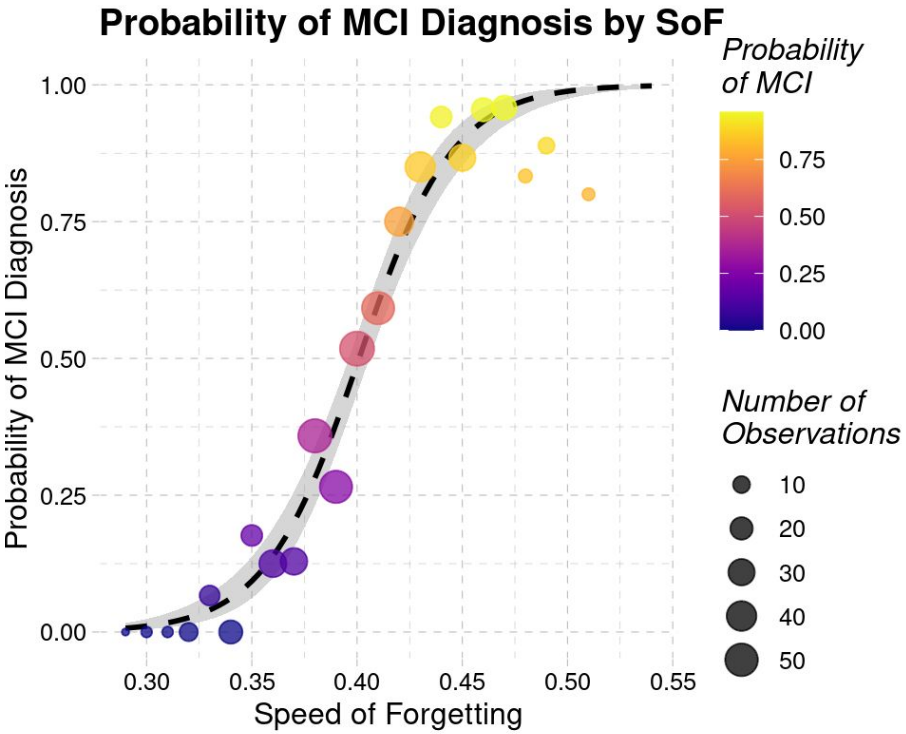
Weekly tests over **one year**

Mean correlation $\boldsymbol{r = 0.72}$



Hake et al, 2023, *CogSci* ♕ Applied Modeling Award

Differences in memory predict cognitive impairment



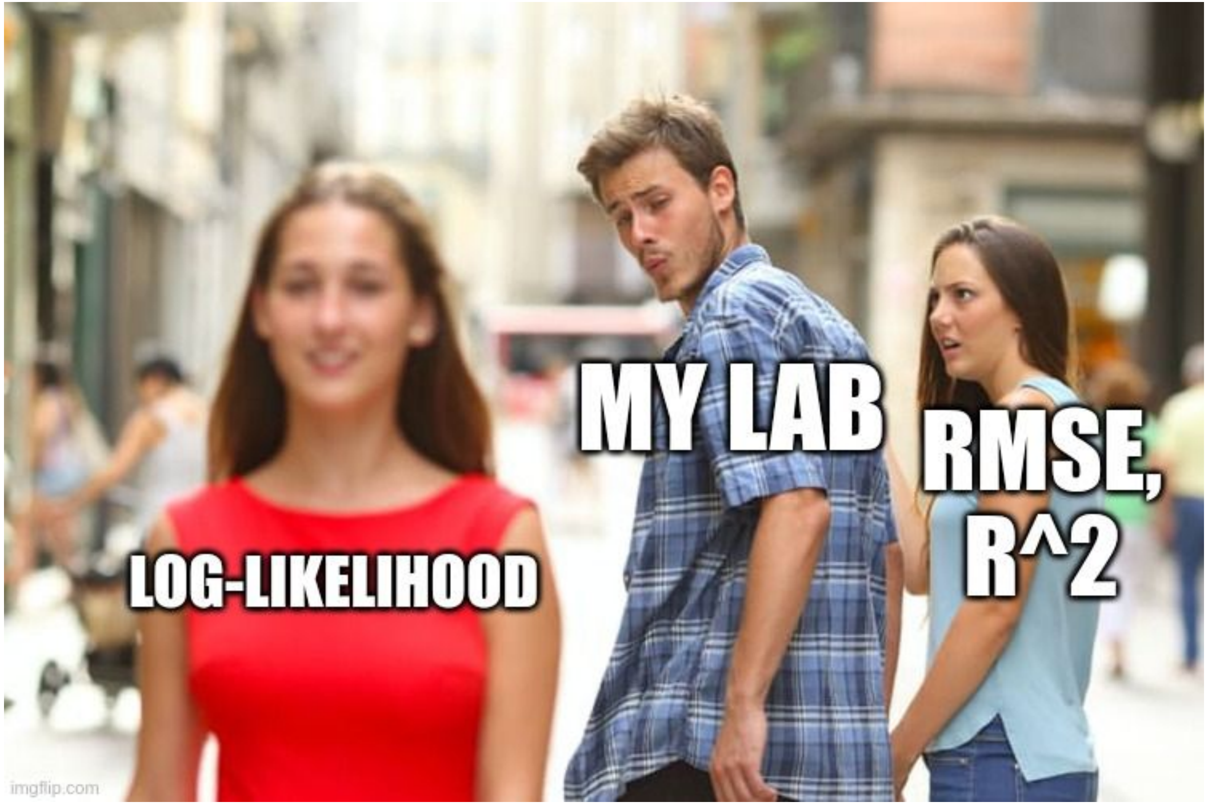
Adjusted $r^2 = 0.38$

Summary

Reasons to use Maximum Likelihood (especially trial-by-trial)

- Clear **interpretation**
- Comparisons between models of **different complexity**
- Can mix **multiple measures**
- Reliable **individual differences**

Even shorter summary



Super special thanks to...