Modeling Flux | Part 4 | Model Comparison & Recoverability

Developmental Computational Psychiatry lab

HARTLEY LAB

Model selection, comparison, and validation

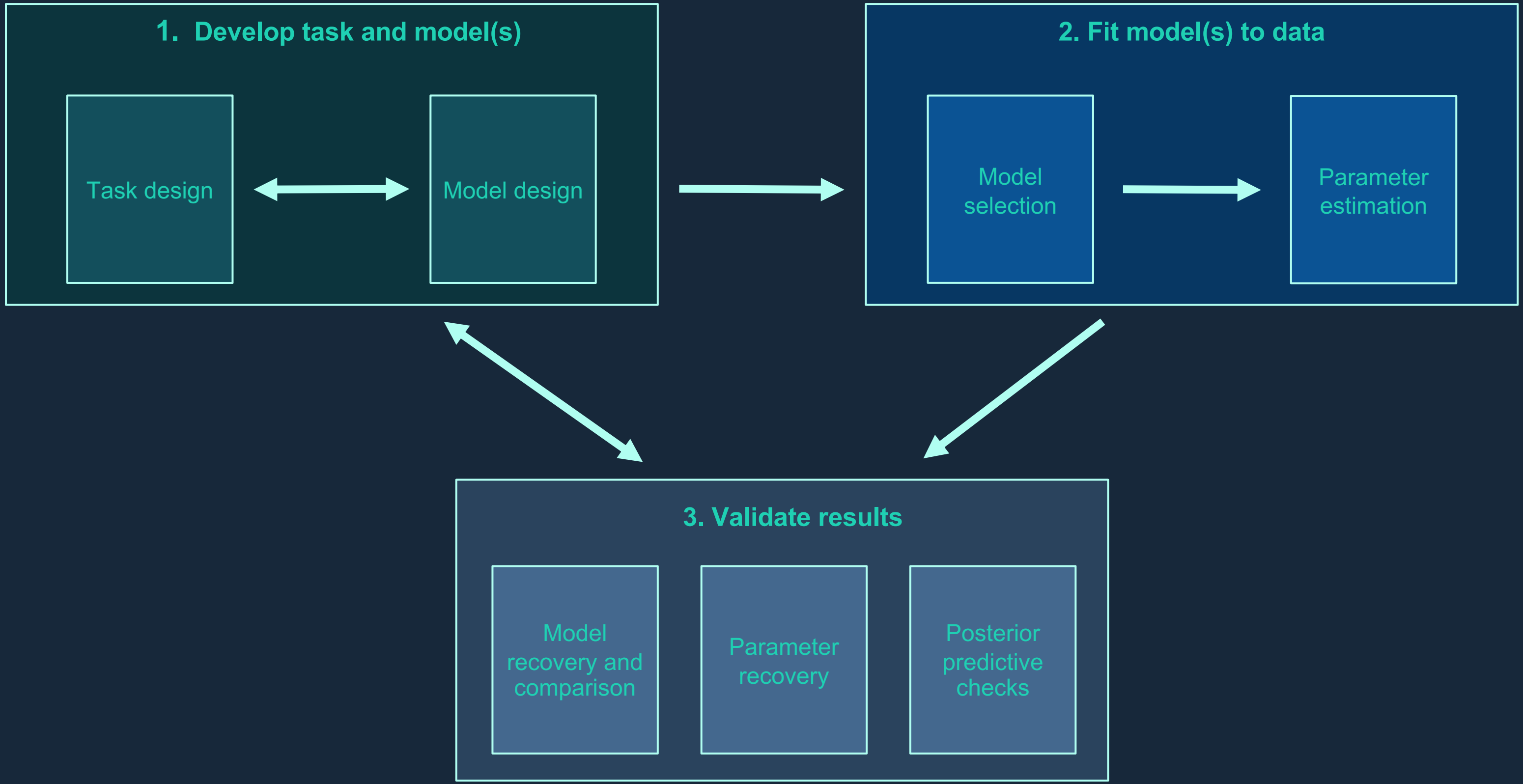
Kate Nussenbaum | Vasilisa Skvortsova | Johanna Habicht

flux

Developmental
Computational
Psychiatry lab

HARTLEY LAB

Modeling Flux | Part 4 | Model Comparison & Recoverability



flux

Developmental Computational Psychiatry lab

HARTLEY LAB

3. Validate results

Model recovery and comparison	Parameter recovery	Posterior predictive checks

flux

Developmental Computational Psychiatry lab

HARTLEY LAB

What do all these procedures have in common?

MODEL SIMULATION

3. Validate results

- Model recovery and comparison
- Parameter recovery
- Posterior predictive checks

Model simulation

Participants' data (choices) → Model fitting (RL model) → Best fitted parameters τ_{fitted} and α_{fitted}

➔ **You start with the data and unknown model & parameters**

Model simulation

Participants' data (choices) → Model fitting (RL model) → Best fitted parameters τ_{fitted} and α_{fitted}

➔ **You start with the data and unknown model & parameters**

Known model with some parameters (pilot, previous experiments) → Simulate participants' data based on known $\tau_{simulated}$ and $\alpha_{simulated}$ → Use simulated data to run

- parameter recovery
- model recovery
- predictive checks

➔ **You start with the known model & parameters to generate "fake" data**

Developmental
Computational
Psychiatry lab

HARTLEY LAB

Parameter recovery

How reliable are model parameters?
How do parameters change relative to one another?

→ **We need to perform parameter recovery checks.**
→ **For example, with current task and 1 LR model:**

- Recover softmax decision temperature
- Recover learning rates

flux

Developmental Computational Psychiatry lab

HARTLEY LAB

Parameter recovery

Steps:

1. Fit the model to behavior and define parameters' range (average/median/min-max)
2. Simulate the model varying one of the parameter values while keeping other parameters fixed
3. Fit simulated data with the same model used for the simulation
4. Compare true and recovered parameters.

flux

Developmental Computational Psychiatry lab

HARTLEY LAB

Parameter recovery

Example:

2-arm bandit task with binary outcomes (reward, no reward)
Model with 1 learning rate and softmax decision rule

N = 30 subjects

T = 100 trials, $p_{reward} = 0.8$

Best fitted model group parameters:

chosen learning rate $\alpha_{chosen} = 0.5$,

softmax temperature $\tau = 0.3$

flux

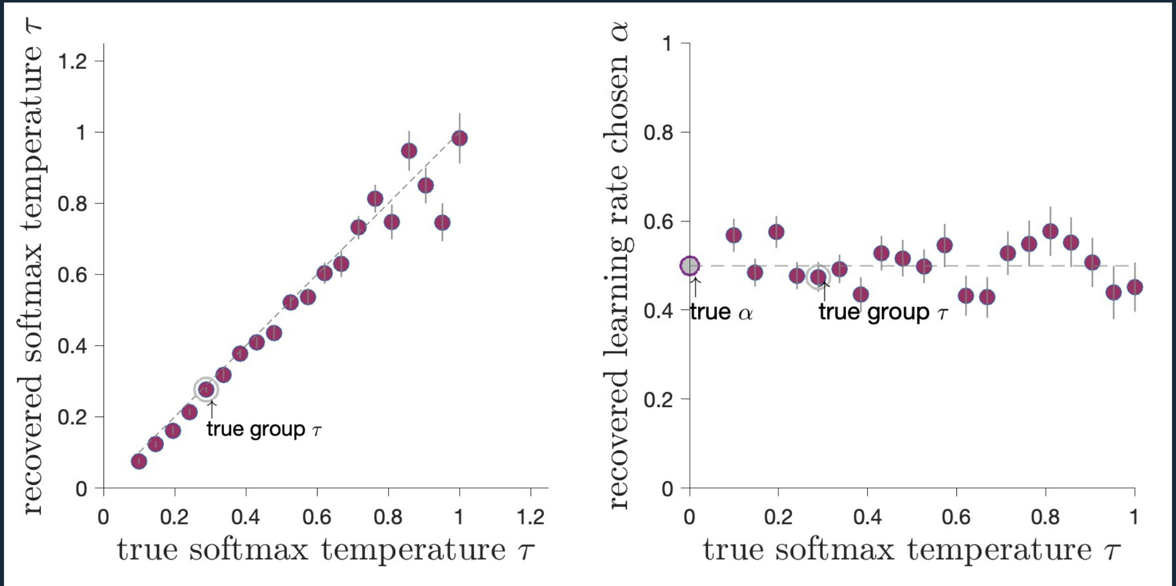
Developmental Computational Psychiatry lab

HARTLEY LAB

Parameter recovery

For softmax temperature τ

Simulate $\tau \in [0.1, 1]$ and $\alpha_{chosen} = 0.5$



flux

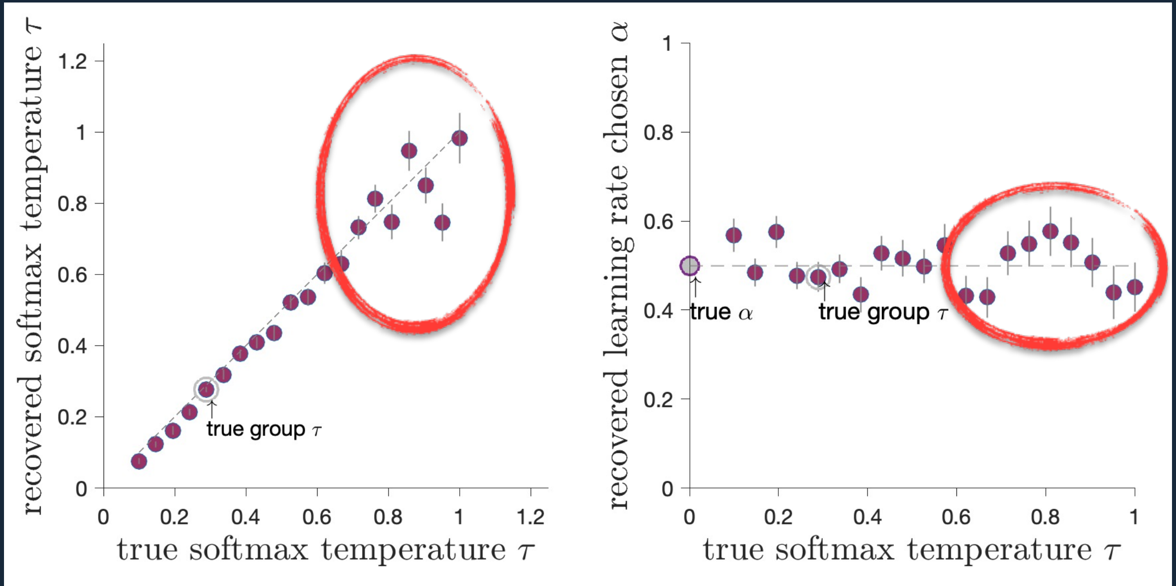
Developmental Computational Psychiatry lab

HARTLEY LAB

Parameter recovery

For softmax temperature τ

Simulate $\tau \in [0.1, 1]$ and $\alpha_{chosen} = 0.5$



flux

Developmental
Computational
Psychiatry lab

HARTLEY LAB

Parameter recovery

For softmax temperature τ

Simulate $\tau \in [0.1, 1]$ and $\alpha_{chosen} = 0.5$

- → **Recovery shows problem with high values of the softmax inverse temperature: $\uparrow \tau \rightarrow$ more exploration $\rightarrow$ less stable estimates**
- → **True values for both parameters might be overestimated**
- → **Try: increase the number of trials: $100 \rightarrow 300$**

flux

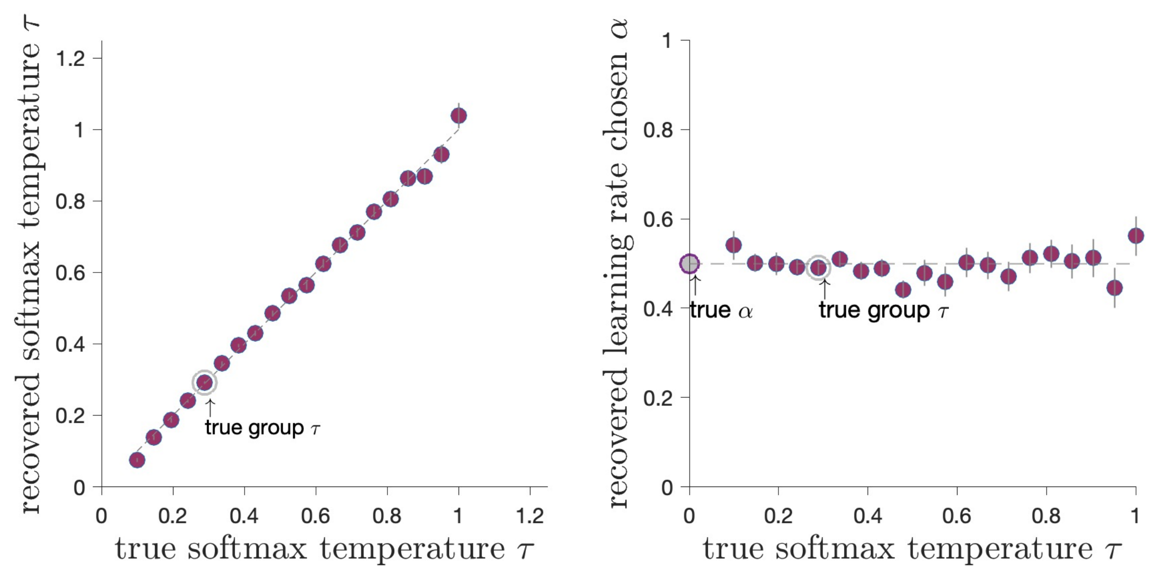
Developmental Computational Psychiatry lab

HARTLEY LAB

Parameter recovery

For softmax temperature τ

Simulate $\tau \in [0.1, 1]$ and $\alpha_{chosen} = 0.5$



flux

What is “good” recoverability?

- No hard rule or accepted values
- Experimental design and nature of the data: example laboratory well-controlled study vs. online experiment with much noisier data
- Simulations help setting a benchmark & optimise the design
    - Vary number of trials
    - Vary parameter ranges

Model recovery

→ Questions about cognition (and developmental change in cognitive processes) can be addressed through examining **parameter estimates** from a single model

Developmental
Computational
Psychiatry lab

HARTLEY LAB

Model recovery

→ Questions about cognition (and developmental change in cognitive processes) can be addressed through examining **parameter estimates** from a single model and/or by **comparing *different* models.**

flux

Model recovery

- → Questions about cognition (and developmental change in cognitive processes) can be addressed through examining **parameter estimates** from a single model and/or by **comparing *different* models.**

- → In the domain of reinforcement learning, different models typically formalize different **value-updating processes** or **choice functions**.

What if there are multiple plausible models of behavior?

→ Typically, more than one hypothesis about behavior can be formalized algorithmically

Developmental
Computational
Psychiatry lab

HARTLEY LAB

What if there are multiple plausible models of behavior?

→ Typically, more than one hypothesis about behavior can be formalized algorithmically

→ For example, with task described earlier:

1. One learning-rate model
2. Decay model
3. Null model

flux

Defining different models

1. One learning-rate model – 2 parameters (τ, α)

   Single learning rate scales prediction errors

Defining different models

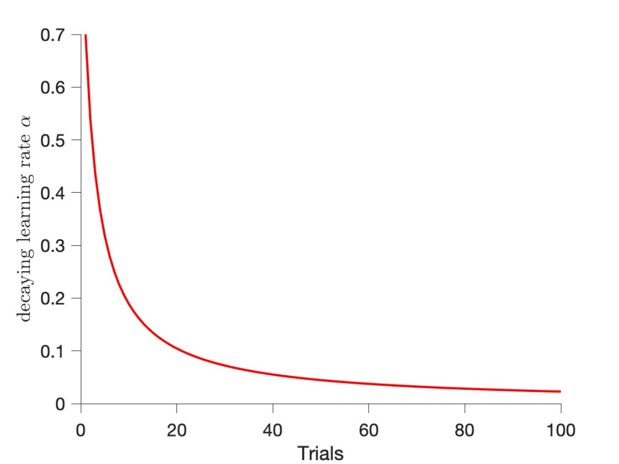
1. One learning-rate model – 2 parameters (τ, α)

   Single learning rate scales prediction errors

2. Decay model – 3 parameters $(\tau, alpha_{initial}, \eta)$

   Learning rate decays over time

$$\alpha_{decay} \ at \ t = \frac{\alpha_{initial}}{1 + \eta * trial_{t-1}}$$



Defining different models

1. One learning-rate model – 2 parameters (τ, α)

   Single learning rate scales prediction errors

2. Decay model – 3 parameters $(\tau, alpha_{initial}, \eta)$

   Learning rate decays over time

3. Null model – 0 parameters

   No learning — random choice on every trial

Model comparison: determining which model best captures data

Steps:

1. Fit all possible models to behavior
2. Compare indices of “fit”

Model comparison: determining which model best captures data

Steps:

1. **Fit all possible models to behavior**
2. Compare indices of "fit"

→ Take into account likelihood — *the probability of the observed choices given the algorithm* — AND penalize more complex models

Developmental
Computational
Psychiatry lab

HARTLEY LAB

Model comparison: determining which model best captures data

Steps:

1. **Fit all possible models to behavior**
2. **Compare indices of "fit"**

→ Take into account likelihood — *the probability of the observed choices given the algorithm* — AND penalize more complex models

→ Two common metrics: AIC and BIC

flux

Developmental
Computational
Psychiatry lab

HARTLEY LAB

Model comparison: determining which model best captures data

Steps:

1. **Fit all possible models to behavior**
2. **Compare indices of "fit"**

→ Take into account likelihood — *the probability of the observed choices given the algorithm* — AND penalize more complex models

→ Two common metrics: AIC and BIC

AIC: $2k - 2\ln(L)$

BIC: $k\ln(n) - 2\ln(L)$

flux

Developmental Computational Psychiatry lab

HARTLEY LAB

AIC and BIC

AIC: $2k - 2\ln(L)$

BIC: $k\ln(n) - 2\ln(L)$

k: number of parameters

L: max likelihood

n: number of observations

→ Smaller values are better

flux

Finding the best-fitting model

Developmental Computational Psychiatry lab

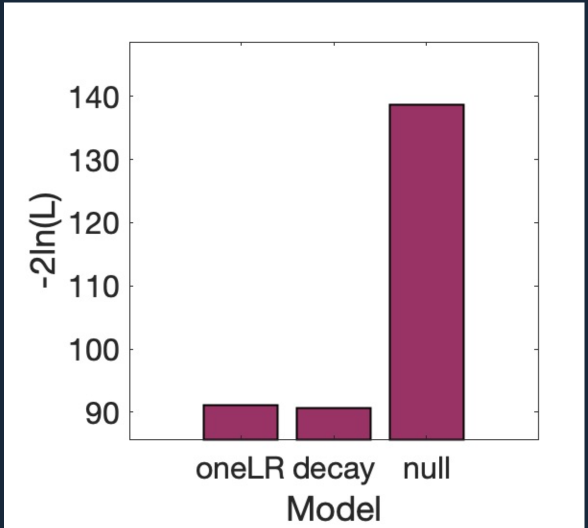
HARTLEY LAB

Finding the best-fitting model

Likelihood term

worse

better



flux

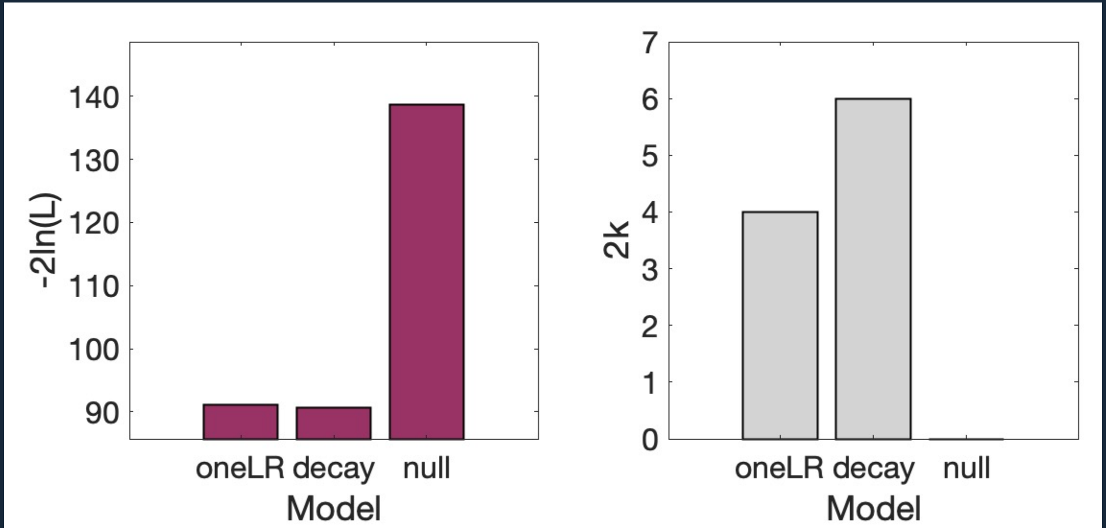
Finding the best-fitting model

Likelihood term

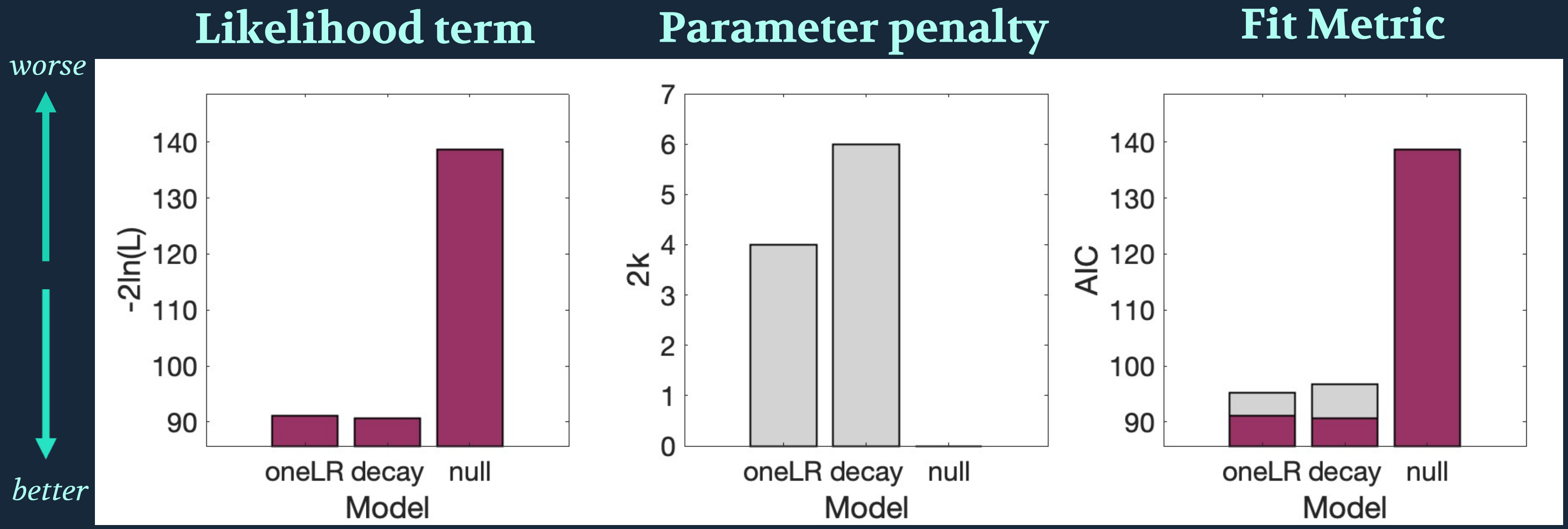
Parameter penalty

worse

better



Finding the best-fitting model



Are our models ‘recoverable’?

→ Critical to ensure that different models are actually distinguishable from one another, given the task design.

Are our models 'recoverable'?

→ Critical to ensure that different models are actually distinguishable from one another, given the task design.

→ Extreme example:

Imagine a task that involves 3 trials.

Can quantitatively assess model fit, but it's unlikely you will really be able to learn anything about the cognitive processes behind a participant's choices.

Developmental Computational Psychiatry lab

HARTLEY LAB

How do we know whether our model-fitting results reflect reality?

→ *Problem:* No way to know the 'true' algorithm a participant used to make choices.

→ *Solution:* Simulate fake participants so that we *know* the algorithm that generated the choice data.

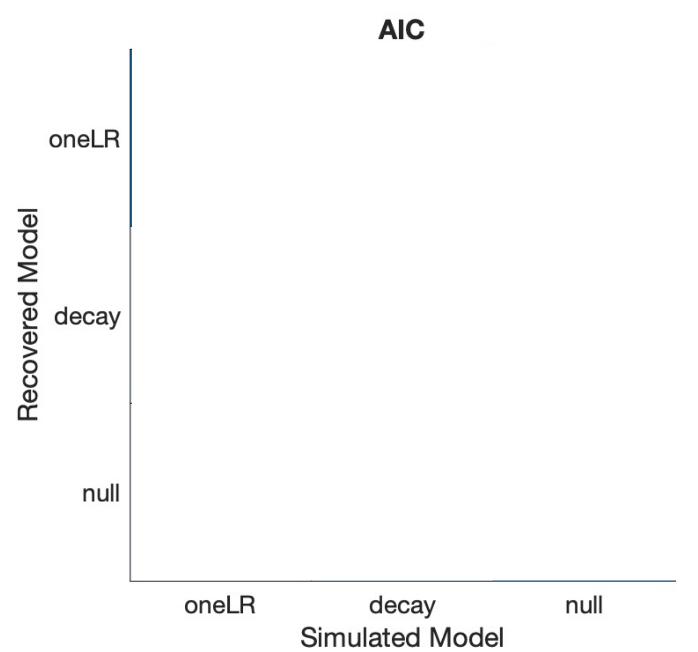
flux

Model recoverability analyses

Steps:

1. Simulate data from all models.
2. Fit models to all simulated datasets.
3. Determine which model best fits each data set.
4. Determine the proportion of datasets for which the 'recovered' model matches the 'ground truth' model.

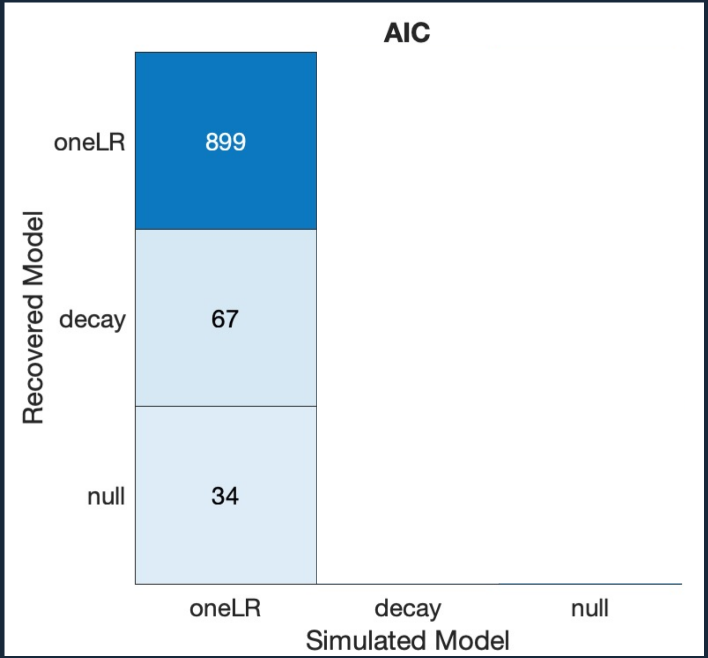
Model recoverability analyses: Confusion matrices



Developmental Computational Psychiatry lab

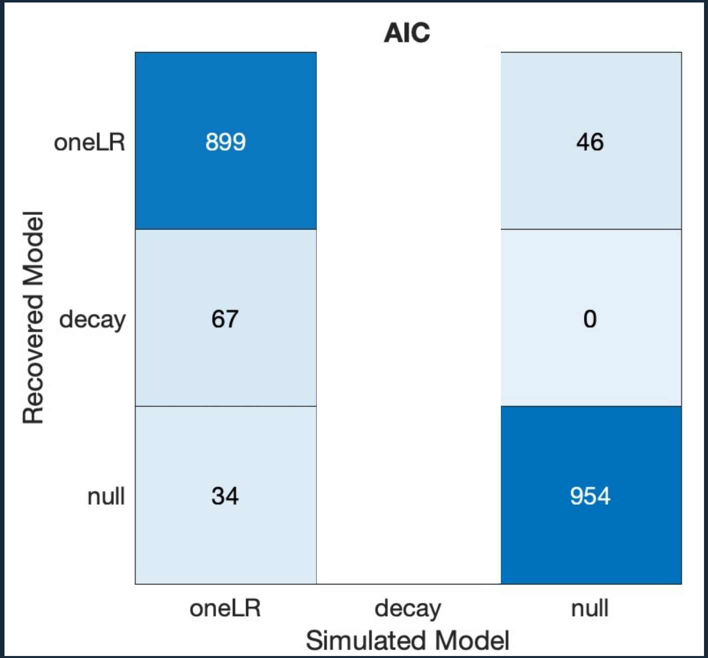
HARTLEY LAB

Model recoverability analyses: Confusion matrices



flux

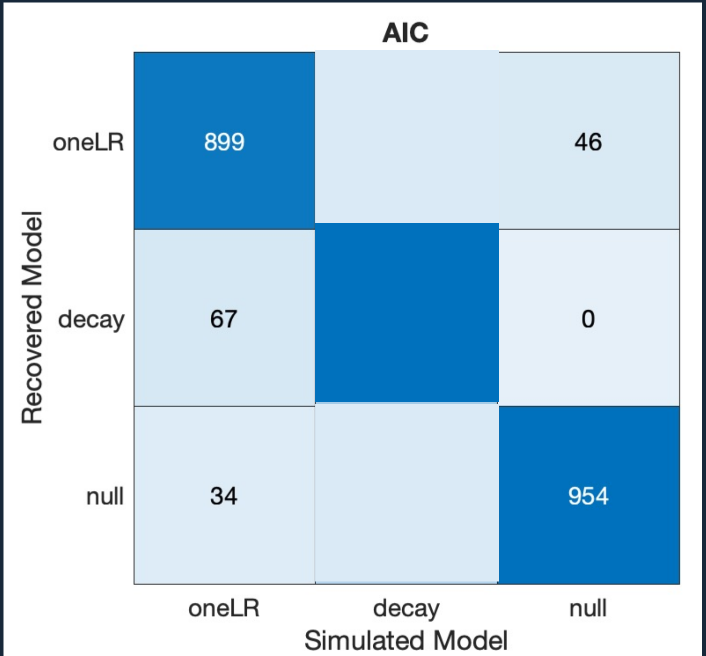
Model recoverability analyses: Confusion matrices



Developmental Computational Psychiatry lab

HARTLEY LAB

Model recoverability analyses: Confusion matrices

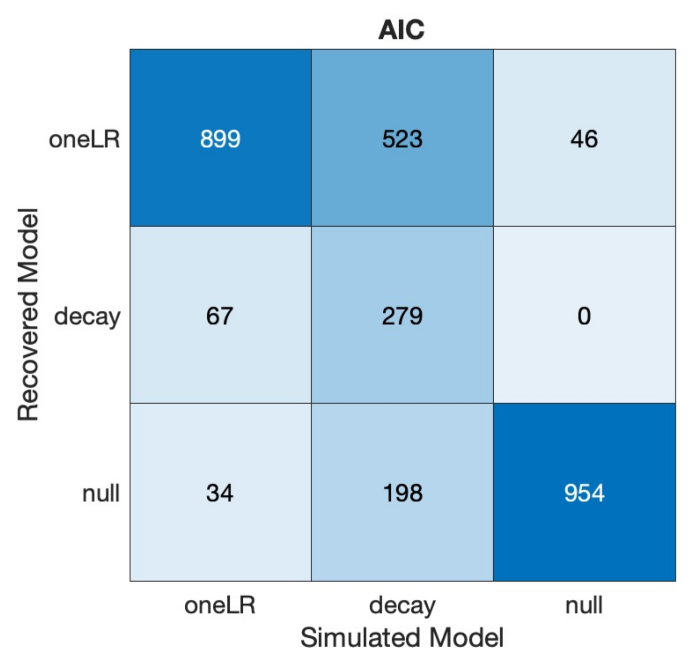


flux

Developmental Computational Psychiatry lab

HARTLEY LAB

Model recoverability analyses: Confusion matrices



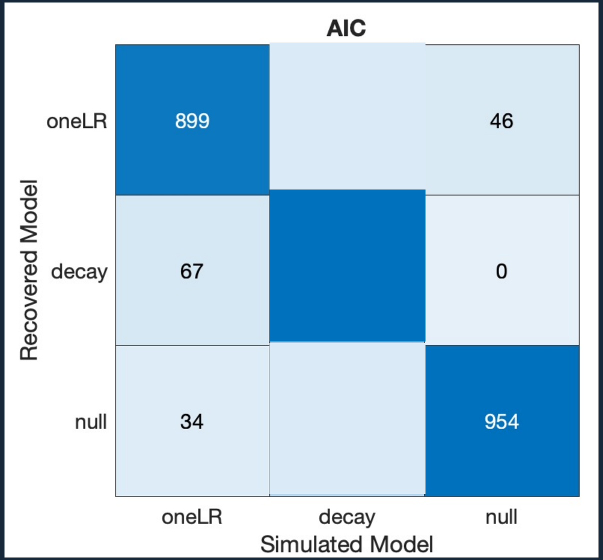
flux

Developmental Computational Psychiatry lab

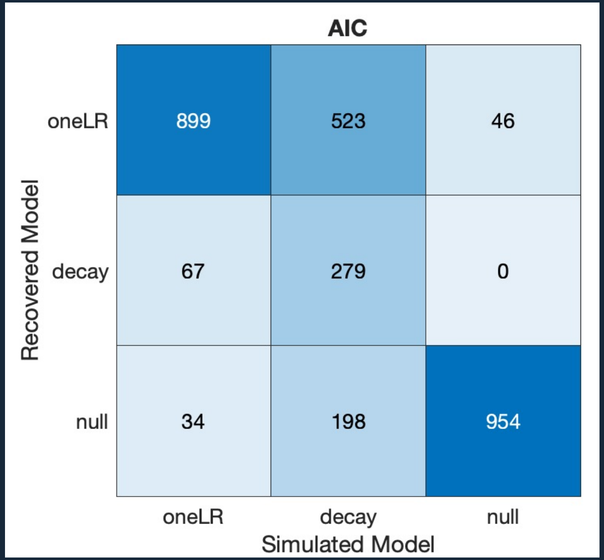
HARTLEY LAB

Model recoverability analyses: Dream vs. reality

Dream



Reality



flux

Developmental
Computational
Psychiatry lab

HARTLEY LAB

Task optimization

What if model recoverability is poor?

→ **Change aspects of task design to improve it.**

flux

Developmental Computational Psychiatry lab

HARTLEY LAB

1. Develop task and model(s)

Task design

3. Validate results

Model recovery and comparison

Parameter recovery

Posterior predictive checks

Developmental
Computational
Psychiatry lab

HARTLEY LAB

Task optimization

What if model recoverability is poor?

→ **Change aspects of task design to improve it.**

→ Examples: Number of trials, number of stimuli, changes in reward probabilities, etc.

flux

Developmental
Computational
Psychiatry lab

HARTLEY LAB

Task optimization

What if model recoverability is poor?

→ **Change aspects of task design to improve it.**

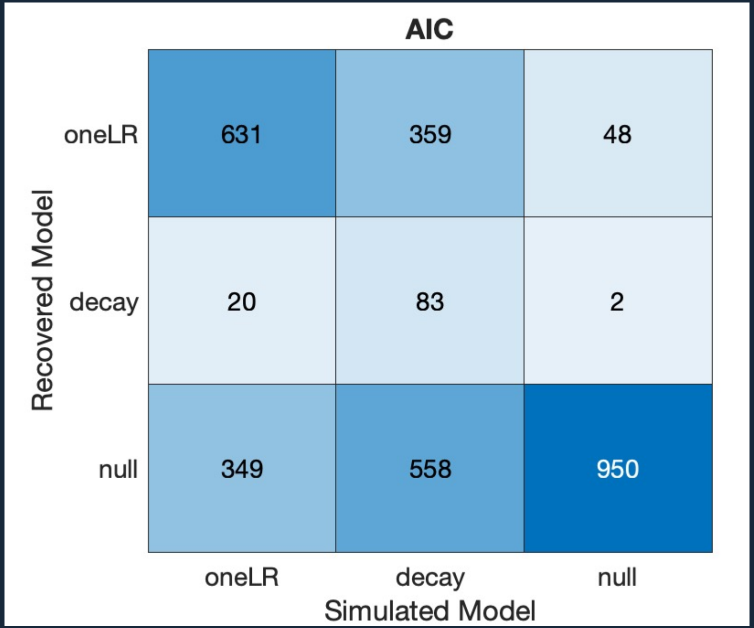
→ Examples: Number of trials, number of stimuli, changes in reward probabilities, etc.

→ **Repeat.**

flux

Comparing task versions

20 trials



200 trials

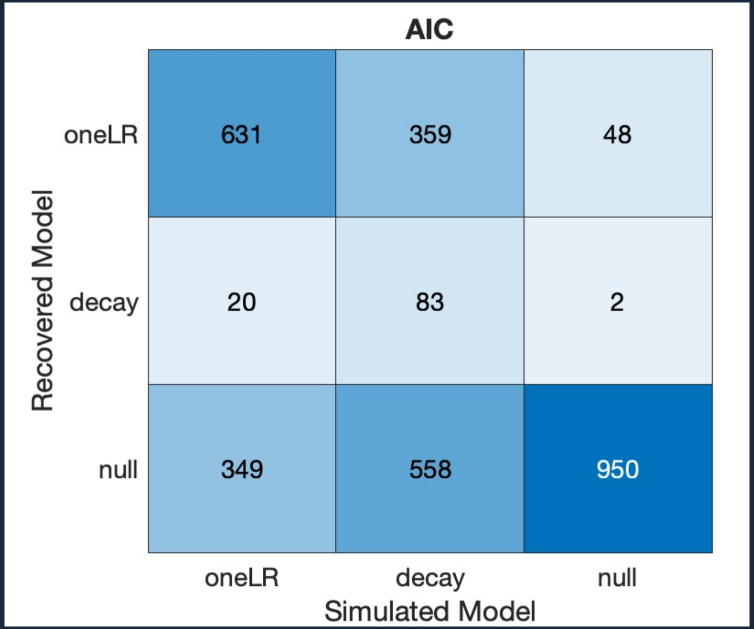
Developmental
Computational
Psychiatry lab

HARTLEY LAB

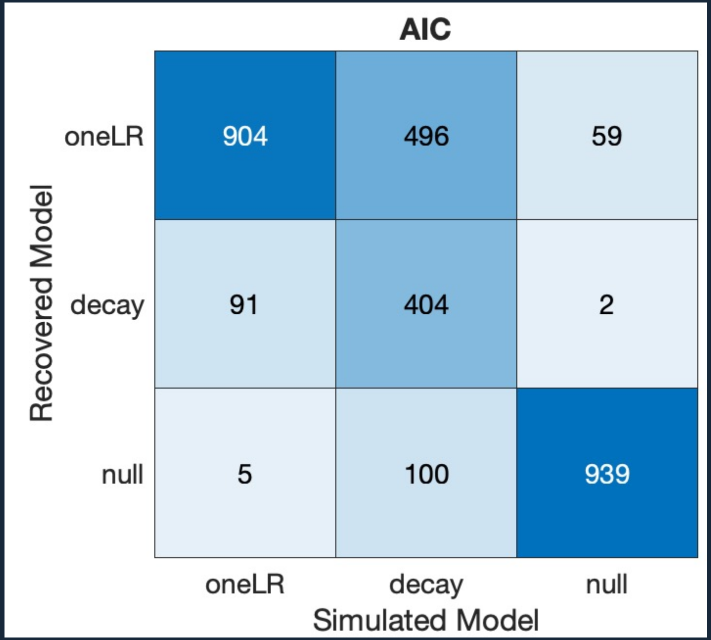
flux

Comparing task versions

20 trials



200 trials



Predictive performance and model checks

1. **All models could be wrong — *the model comparisons are relative***

→ Better "fit" does not guarantee the model reproduces behaviour

→ Better "fit" does not guarantee the model could be recovered

Developmental
Computational
Psychiatry lab

HARTLEY LAB

Predictive performance and model checks

1. **All models could be wrong — *the model comparisons are relative***
   - ➔ Better “fit” does not guarantee the model reproduces behaviour
   - ➔ Better “fit” does not guarantee the model could be recovered
2. **Generative performance of the model: how well the model can reproduce behaviour**

flux

Developmental Computational Psychiatry lab

HARTLEY LAB

Predictive performance and model checks

1. **All models could be wrong — *the model comparisons are relative***

   → Better "fit" does not guarantee the model reproduces behaviour

   → Better "fit" does not guarantee the model could be recovered

2. **Generative performance of the model: how well the model can reproduce behaviour**

   → Check model performance against behavioural data **qualitatively**

   → Try to find a behavioural pattern that **dissociates** between the models

The importance of model falsification (Palminteri et al., Trends Cog Sci 2017)

flux

Developmental Computational Psychiatry lab

HARTLEY LAB

Generative performance of the model

Steps:

1. Simulate the models with the best fitted parameters
2. Define a behavioural marker: where models' behaviour won't generate the same predictions
3. Compare model performance to subjects' actual behaviour

flux

Developmental Computational Psychiatry lab

HARTLEY LAB

Generative performance of the model

Example:

Model 1: *fixed* learning rate and softmax decision rule
Model 2: *decaying* learning rate and softmax decision rule:
the agent progressively decreases the update of the option values and "ignores" the irrelevant non-rewarding events

The agent is *less likely* to switch choice after a negative prediction error

Benchmark behaviour: $P(switch)$ after a *negative* prediction error

flux

Developmental Computational Psychiatry lab

HARTLEY LAB

Generative performance of the model

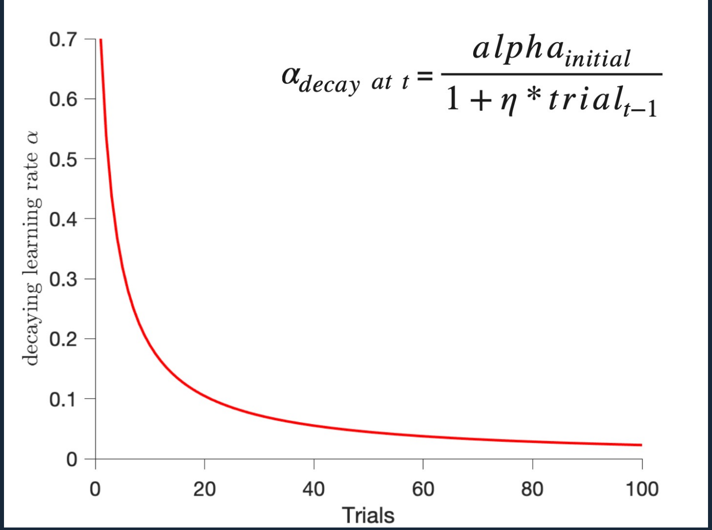
Example:

N = 30 subjects who played 2-arm bandit task, T = 100 trials, $p_{reward} = 0.8$

Best fitted model group parameters:

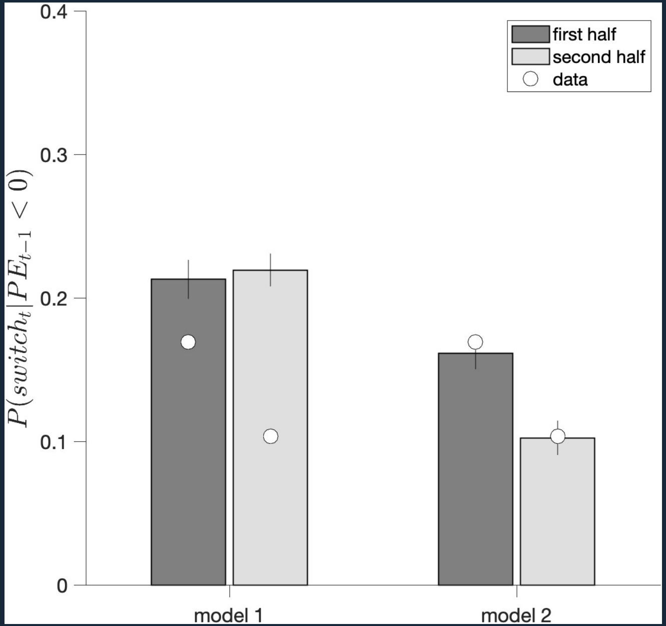
Model 1: chosen learning rate $\alpha_{chosen} = 0.7, \tau = 0.2$

Model 2: initial chosen learning rate $\alpha_{chosen} = 0.7$, decay parameter $\eta = 0.02$



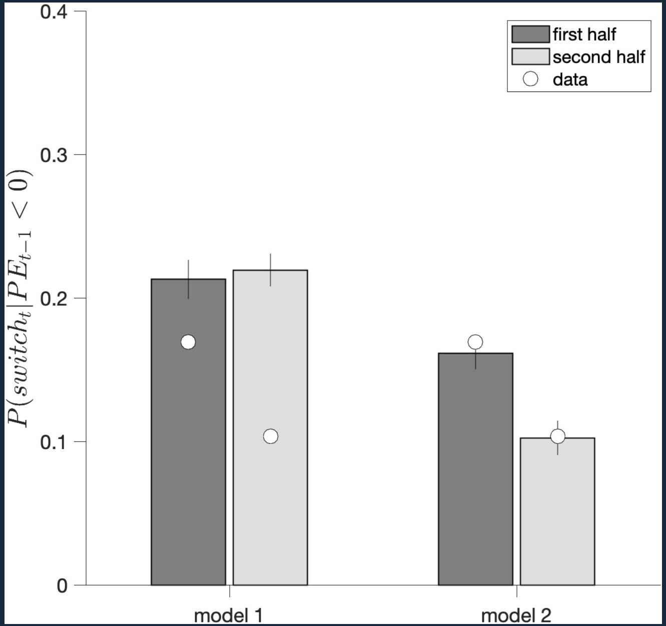
Generative performance of the model

Example:



Generative performance of the model

Example:



Only the model with *decaying* learning rate was capable of generating this behavioural pattern → the two models can be dissociated based on this behavioural marker.

Developmental Computational Psychiatry lab

HARTLEY LAB

Model validation summary

- It's *extremely important* to ensure parameters and models of interest are recoverable *prior* to data collection.
- Simulation is a valuable tool to test modeling approaches.
- Tasks and models should be developed and refined *concurrently.*

flux

Q & A