# Speaker Notes
# Sensorimotor Habituation in *Drosophila* Larvae

Gil Raitses

Syracuse University • December 2025

## Contents

# Slide 1: Title

**Opening**

Thank you for having me. I will present work on sensorimotor habituation in *Drosophila* larvae, covering both our population-level modeling success and our subsequent attempt to extend the approach to individual phenotyping.

**Transition**

The presentation has two parts. The first covers the original study where I developed a temporal kernel model. The second covers the follow-up study where I tested whether the same model could phenotype individual larvae.

# Slide 2: Original Study Overview

**Key Points to Emphasize**

The gamma-difference kernel has two timescales that govern behavior. Fast excitation at $\tau_1 \approx 0.3$ seconds captures the initial sensory response. Slow suppression at $\tau_2 \approx 4$ seconds produces habituation across repeated stimuli. The model was validated across 14 experiments and 701 unique tracks.

**Audience Anchor**

If there is one thing to remember from the original study, it is that larval reorientation dynamics can be captured by a simple parametric model with biologically interpretable timescales.

# Slide 3: Kernel Structure

**Figure Walkthrough**

Left panel shows the combined kernel representing the full temporal response. Right panel shows the decomposition into fast gamma in green and slow gamma in red. The fast component peaks at 0.3 seconds and drives immediate response. The slow component peaks around 4 seconds and produces delayed suppression.

**Mathematical Point**

The kernel $K(t)$ modulates reorientation hazard rate. Positive values increase turning probability. Negative values suppress it. The crossover from positive to negative creates the characteristic excitation-then-inhibition pattern.

**Connection to Biology**

These timescales may correspond to distinct neural circuit mechanisms. The fast component could reflect direct sensory activation. The slow component could reflect adaptation or inhibitory feedback.

# Slide 4: PSTH Computation Methods

**Two Approaches**

Empirical PSTH uses direct histogram binning of event times relative to LED onset. No functional form is assumed. Bins are 100ms wide.

Parametric PSTH is derived from the Bernoulli event model. At each timestep, event probability $p(t) = 1 - \exp(-\lambda(t) \cdot \Delta t)$.

**Key Distinction**

Empirical PSTH is what is observed. Parametric PSTH is what the model predicts. Model fitting minimizes the gap between them.

**Why This Matters**

Understanding both approaches is essential for interpreting model validation and simulation accuracy.

## Slide 5: Simulation Track Generation

**Bernoulli Point Process**

This is how simulated tracks are generated.

Step 1: Compute hazard rate at each timestep as $\lambda(t) = \lambda_0 \exp(K(t) + \eta)$.

Step 2: Convert to probability as $p(t) = 1 - \exp(-\lambda(t) \cdot \Delta t)$.

Step 3: Draw Bernoulli sample at each timestep.

Step 4: Enforce 2-second refractory period.

**Key Parameters**

Baseline intercept controls mean event rate. Track-level standard deviation $\eta$ controls across-track variance.

**Connection to Phenotyping**

Simulations with known ground truth allow testing whether clustering can recover true phenotypes.

## Slide 6: Parameter Sweep for Simulation Calibration

**The Optimization Problem**

Simulations must match empirical data. Two parameters control the event distribution: intercept controls mean event rate, and track standard deviation controls variance across tracks.

**Grid Search**

Swept intercept from $-7.0$ to $-6.0$ and standard deviation from $0.1$ to $0.8$. Compared each combination to empirical distribution using KS test.

**Optimal Values Found**

Intercept $= -6.54$ and track standard deviation $= 0.38$ produce 14.9 events per 10-minute track with realistic variance.

## Slide 7: Simulated vs Empirical Event Counts

**Validation Message**

Before using the model for anything, it must be confirmed that it generates realistic data. Panel A shows the histograms overlap well. Panel B shows the box plots match.

**Key Numbers**

260 empirical tracks and 300 simulated tracks both show median around 15 events per track. Simulated tracks used optimized intercept $= -6.54$ and standard deviation $= 0.38$.

**Why This Matters**

The simulation framework is the foundation for power analysis. If simulations do not match empirical data, power calculations are meaningless.

## Slide 8: Habituation Dynamics

**Behavioral Phenomenon**

Turn fraction increases across LED pulses in all four experimental conditions. Larvae spend more time turning and less time running as the session progresses.

**Condition Comparison**

The 0-250 Cycling condition shows the strongest habituation effect with slope $+0.031$ per pulse. The 50-250 conditions show weaker effects. Shaded bands are 95% confidence intervals.

**Interpretation**

Habituation is the behavioral manifestation of the slow suppressive component accumulating across pulses. The kernel model predicts this effect.

## Slide 9: Behavioral State Analysis

**Detailed State Breakdown**

Gray represents forward running. Pink represents turning. Blue represents pausing. Orange represents reverse crawling.

**Key Observation**

Turning fraction increases dramatically. Pausing remains below 5% throughout. Habituation manifests as increased turning, not increased pausing or freezing.

**Quantitative Point**

By pulse 17 in the 50-250 Cycling condition, larvae spend nearly 40% of their time turning compared to about 20% at pulse 0.

## Slide 10: Leave-One-Experiment-Out Validation

**What This Shows**

Leave-one-experiment-out cross-validation tests whether kernel parameters estimated from 13 experiments generalize to the held-out experiment.

**Key Result**

Pass rate of 50% falls within the null distribution with $p = 0.618$. Cross-experiment generalization is no better than chance.

**Interpretation**

The population model fits well overall, but individual experiments show high variability. This foreshadows the individual-level problems addressed in the follow-up study.

**Transition**

This result motivated the follow-up question: Can individual larvae be phenotyped using their unique kernel parameters?

## Slide 11: Follow-Up Study Overview

**Key Points to Emphasize**

The answer to individual phenotyping is negative with current protocols. Sparse data with only 18 to 25 events per track makes 6-parameter estimation unreliable. Apparent clusters are statistical artifacts of fitting high-dimensional models to low-event tracks. Only 8.6% of tracks show genuine individual differences.

**Audience Anchor**

The follow-up study is a negative result. Individual phenotyping could not be achieved. But the negative result is informative because it identifies the root cause and points toward solutions.

## Slide 12: The Clustering Illusion

**Figure Walkthrough**

Panel A shows PCA reveals unimodal distribution, not discrete clusters. Panel B shows all four validation methods fail with ARI below 0.13. Panel C shows gap statistic is minimized at $k = 1$, indicating no clusters.

**Key Message**

K-means will always produce $k$ clusters regardless of whether true clusters exist. The gap statistic tells us $k = 1$ is optimal. There are no discrete phenotypes in this data.

**Why It Matters**

Clusters identified by unsupervised learning are artifacts of sparse data, not genuine biological phenotypes. Publishing these clusters would be misleading.

## Slide 13: Data Sparsity Explains Instability

**The Math Problem**

Mean 25 events per track. Six kernel parameters to estimate. Data-to-parameter ratio is 4:1. Reliable MLE requires at least 10:1.

**Visual Explanation**

Panel C shows the calculation: 4 parameters divided by 25 events equals a ratio of 6:1. This is fundamentally underdetermined.

**Key Number**

100 events per track is the target for stable estimation. Current protocols deliver only 25.

## Slide 14: Hierarchical Shrinkage

**What Shrinkage Does**

Bayesian hierarchical estimation pulls individual estimates toward the population mean. Tracks with sparse data shrink more. Tracks with abundant data retain their individual estimates.

**Key Insight**

Shrinkage is not a bug. It is optimal regularization under the assumption that individuals are exchangeable members of a population.

**Limitation**

Shrinkage cannot create information that is absent. With only 25 events, almost all individual estimates shrink heavily toward the population mean.

## Slide 15: The Identifiability Problem

**Figure Walkthrough**

Panel A shows continuous design produces high bias and RMSE. Panel B shows burst design extracts 10 times more Fisher Information per event. Panel C shows MLE recovery differs dramatically by design. Panel D shows continuous fails because inhibition dominates during LED-ON.

**Key Insight**

The problem is not just data quantity but data quality. Continuous 10-second LED pulses produce events during the suppressive phase of the kernel. These events carry almost no information about $\tau_1$.

**Recommendation Preview**

Switch to burst stimulation to sample the early excitatory window repeatedly.

# Slide 16: Stimulation Protocol Comparison

**Four Designs Shown**

Panel A shows current continuous 10s ON, 20s OFF. Panel B shows recommended burst 10 pulses of 0.5s with 2s spacing. Panel C shows alternative 4 pulses of 1s with 5s spacing. Panel D shows alternative 2 pulses of 2s with 10s spacing.

**Key Numbers**

Burst design provides 8 times more Fisher Information than continuous. This could reduce the number of events required for reliable estimation from 100 to 30.

# Slide 17: Kernel Model Comparison

**Why Compare Models**

The gamma-difference kernel was chosen for interpretability, but verification is needed that it fits as well as flexible alternatives.

**Results**

Raised cosine basis achieves $R^2 = 0.974$ with 12 parameters. Gamma-difference achieves $R^2 = 0.968$ with 6 parameters.

**Interpretation**

The gamma-difference captures 96.8% of the variance explained by the flexible model with half the parameters. The timescales $\tau_1$ and $\tau_2$ are not just curve-fitting artifacts. They represent genuine temporal structure.

# Slide 18: Protocol Modification

**Primary Recommendation**

Replace continuous 10-second ON periods with burst trains. Each burst event carries 10 times more Fisher Information.

**Quantitative Benefit**

This modification alone could reduce the number of events required for reliable estimation from 100 to approximately 30.

**Implementation**

Change the LED control code to deliver 10 pulses of 0.5 seconds each with 2-second spacing instead of a single 10-second pulse.

# Slide 19: Extended Recording

**Secondary Recommendation**

Target 40 minutes or more of recording to achieve at least 50 reorientation events per track.

**Current State**

Current 10 to 20 minute recordings yield only 18 to 25 events.

**Power Analysis Result**

100 events are required for 80% power to detect a 0.2-second difference in $\tau_1$ at the individual level.

# Slide 20: Model Simplification

**Approach**

Reduce the parameter space by fixing population-derived parameters.

**Specific Suggestion**

Fix $\tau_2$ at the population estimate of 3.8 seconds. Fix the amplitude ratio $B/A$ at the population value. Estimate only the fast timescale $\tau_1$ per individual track.

**Rationale**

Hierarchical Bayesian estimation provides natural regularization toward the population mean. With only one free parameter, even 25 events may be sufficient.

# Slide 21: Alternative Phenotypes

**Pragmatic Alternative**

Use robust composite phenotypes that avoid kernel fitting entirely.

**Examples**

ON/OFF event ratio measures whether larvae respond preferentially during LED-ON versus LED-OFF. First-event latency measures time from LED onset to first reorientation.

**Advantage**

These phenotypes require only event counts, not full 6-parameter kernel estimation.

# Slide 22: Within-Condition Analysis

**Methodological Point**

Analyze individual differences within experimental conditions rather than pooling across conditions.

**Why This Matters**

When data from different stimulation intensities and temporal patterns are pooled, condition effects dominate and mask genuine individual variation.

**Evidence**

The ARI near zero across all validation methods indicates no reproducible structure when pooling.

# Slide 23: Original Study Summary

**Summary of Success**

The gamma-difference kernel accurately models population-level dynamics. Two timescales govern behavior: $\tau_1 \approx 0.3$s for excitation and $\tau_2 \approx 4$s for suppression. The model is robust across experimental conditions. Biological interpretability comes with equivalent goodness of fit.

## Slide 24: Follow-Up Study Summary

**Summary of Challenge**

Individual phenotyping fails with current protocols due to sparse data. Apparent clusters are statistical artifacts. Only 8.6% of tracks show individual variation exceeding noise. Current protocols achieve only 20 to 30% statistical power.

**Bottom Line**

Population-level analysis is robust and biologically meaningful. Individual phenotyping requires experimental redesign before kernel-based classification becomes reliable.

## Slide 25: Thank You

**Transition to Questions**

I am happy to take questions. For common questions, I have prepared some FAQ slides.

## Slides 26 to 30: FAQ

**Original Study Methods Sequence**

Data collection, then MAGAT trajectory extraction, then event detection, then population kernel fitting, then LOEO validation.

**Follow-Up Study Methods Sequence**

Individual MLE fitting, then K-means and hierarchical clustering, then round-trip validation, then power analysis, then identifiability analysis.

**Why Population Succeeds but Individual Fails**

Data-to-parameter ratio. Population pools approximately 15,000 events for 6 parameters, a ratio of 2500 to 1. Individual uses approximately 25 events for 6 parameters, a ratio of 4 to 1.

**What is Hierarchical Shrinkage**

Bayesian regularization that pulls individual estimates toward the population mean proportionally to data sparsity.

**How to Interpret Clustering Results**

With extreme skepticism. K-means will always produce $k$ clusters. The gap statistic shows $k = 1$ is optimal. Round-trip validation shows ARI below 0.2.

## Timing Guide

| Slides | Section | Target Time |
|---|---|---|
| 1 to 2 | Introduction | 2 min |
| 3 to 7 | PSTH and Simulation Methods | 5 min |
| 8 to 10 | Original Study Results | 6 min |
| 11 to 17 | Follow-Up Study | 10 min |
| 18 to 22 | Recommendations | 5 min |
| 23 to 25 | Summary | 3 min |
| 26 to 30 | FAQ if needed | 5 min |

**Total: 31 to 36 minutes**

## Technical Terms

**Gamma-difference kernel** Difference of two gamma distributions, one fast for excitation and one slow for suppression.

**PSTH** Peri-stimulus time histogram, the empirical distribution of event times relative to stimulus onset.

**Fisher Information** Measure of how much information an observable contains about an unknown parameter.

**Hierarchical shrinkage** Bayesian regularization toward population mean.

**Gap statistic** Method for determining optimal number of clusters by comparing within-cluster dispersion to null reference.

**ARI** Adjusted Rand Index, measure of agreement between two clusterings corrected for chance.

**MLE** Maximum likelihood estimation.

**LOEO** Leave-one-experiment-out cross-validation.

## Slide 31: From Counting to Simulation

**Connecting to Traditional Methods**

Most behavioral analysis uses counting and visualization. Event counts per condition. Heatmaps of spatial position. Histograms of event timing. These methods describe what happened but cannot predict new outcomes.

**The Extension**

The kernel-based hazard model is generative. Given stimulation timing and kernel parameters, it produces event probabilities at every moment. The model does not just summarize past data. It predicts future behavior under conditions that have not been tested.

**Practical Value for Experiment Design**

Fisher Information analysis revealed that burst stimulation extracts 10 times more information per event than continuous. This insight emerged from simulation, not from running additional experiments. Power analysis determined that 100 events are needed for individual phenotyping. Protocol designers can use this threshold before collecting any new data.

**Round-Trip Validation**

Simulation allows testing of analysis pipelines with known ground truth. Clustering algorithms can be validated by generating data from known phenotypes and checking whether the algorithm recovers them. This is impossible with empirical data where ground truth is unknown.

**Key Message**

Simulation modeling transforms descriptive behavioral science into predictive science. It bridges the gap between what was observed and what will happen, enabling rational experiment design and rigorous validation.