

# Speaker Notes — Combined Presentation

## Sensorimotor Habituation in Drosophila Larvae

Gil Raitses · Syracuse University · December 2025

---

### 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 we developed a temporal kernel model. The second covers the follow-up study where we tested whether the same model could phenotype individual larvae.

---

### Slide 2: Executive Summary — Original Study

**Key Points to Emphasize** - The gamma-difference kernel has **two timescales** that govern behavior - Fast excitation at  $\tau_{\text{fast}} \sim 0.3$  seconds captures the initial sensory response - Slow suppression at  $\tau_{\text{slow}} \sim 4$  seconds produces habituation across repeated stimuli - Model 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: The combined kernel showing the full temporal response - Right panel: Decomposition into fast gamma (green) and slow gamma (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**: Direct histogram binning of event times relative to LED onset. No functional form assumed. Uses 100ms bins. - **Parametric PSTH**: Derived from the

Bernoulli event model. At each timestep, event probability  $p(t) = 1 - \exp(-\lambda(t) * dt)$ .

**Key Distinction** Empirical PSTH is what we observe. 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 we generate simulated tracks: 1. Compute hazard rate at each timestep:  $\lambda(t) = \lambda_0 * \exp(K(t) + \eta_t)$  2. Convert to probability:  $p(t) = 1 - \exp(-\lambda(t) * dt)$  3. Draw Bernoulli sample at each timestep 4. Enforce 2-second refractory period

**Key Parameters** - Baseline intercept controls mean event rate - Track-level std ( $\eta_t$ ) controls across-track variance

**Connection to Phenotyping** Simulations with known ground truth let us test whether clustering can recover true phenotypes.

---

### Slide 6: Parameter Sweep for Simulation Calibration

**The Optimization Problem** We need simulations that match empirical data. Two parameters control the event distribution: - Intercept: controls mean event rate - Track std: controls variance across tracks

**Grid Search** Swept intercept from -7.0 to -6.0 and std from 0.1 to 0.8. Compared each combination to empirical distribution using KS test.

**Optimal Values Found** - Intercept = -6.54 - Track std = 0.38

These 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, we need to confirm it generates realistic data. Panel A shows the histograms overlap well. Panel B shows the box plots match.

**Key Numbers** - 260 empirical tracks - 300 simulated tracks - Both show median around 15 events per track - Simulated using optimized intercept = -6.54 and std = 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** - 0-250 Cycling shows the strongest habituation effect with slope +0.031 per pulse - 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: Forward running - Pink: Turning - Blue: Pausing - Orange: 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: LOEO 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 we will see in the follow-up study.

**Transition** This result motivated the follow-up question: Can we phenotype individual larvae using their unique kernel parameters?

---

## Slide 11: Executive Summary — Follow-Up Study

**Key Points to Emphasize** - The answer to individual phenotyping is **negative** with current protocols - Sparse data with only 18-25 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. We could not phenotype individuals. 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: PCA reveals unimodal distribution, not discrete clusters - Panel B: All four validation methods fail with ARI below 0.13 - Panel C: 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 - 6 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: Continuous design produces high bias and RMSE - Panel B: Burst design extracts 10 $\times$  more Fisher Information per event - Panel C: MLE recovery differs dramatically by design - Panel D: 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  $\tau_{11}$ .

**Recommendation Preview** Switch to burst stimulation to sample the early excitatory window repeatedly.

---

## Slide 16: Stimulation Protocol Comparison

**Four Designs Shown** - A: Current continuous 10s ON, 20s OFF - B: Recommended burst  $10 \times 0.5\text{s}$  with 2s spacing - C: Alternative  $4 \times 1\text{s}$  with 5s spacing - D: Alternative  $2 \times 2\text{s}$  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** We chose the gamma-difference kernel for interpretability, but we need to verify it fits as well as flexible alternatives.

**Results** - Raised cosine basis:  $R^2 = 0.974$  with 12 parameters - Gamma-difference:  $R^2 = 0.968$  with 6 parameters

**Interpretation** The gamma-difference captures 96.8% of the variance explained by the flexible model with half the parameters. The timescales  $\tau_{\text{a}1}$  and  $\tau_{\text{a}2}$  are not just curve-fitting artifacts. They represent genuine temporal structure.

---

## Slide 18: Recommendation 1 — 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: Recommendation 2 — Extended Recording

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

**Current State** Current 10-20 minute recordings yield only 18-25 events.

**Power Analysis Result** 100 events are required for 80% power to detect a 0.2-second difference in  $\tau_{\text{a}1}$  at the individual level.

---

## Slide 20: Recommendation 3 — Model Simplification

**Approach** Reduce the parameter space by fixing population-derived parameters.

**Specific Suggestion** - Fix  $\tau_{\text{a}2}$  at the population estimate of 3.8 seconds - Fix the amplitude ratio  $B/A$  at the population value - Estimate only the fast timescale  $\tau_{\text{a}1}$  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: Recommendation 4 — 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: Time from LED onset to first reorientation

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

---

### Slide 22: Recommendation 5 — 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: Conclusions — Original Study

**Summary of Success** - Gamma-difference kernel accurately models population-level dynamics - Two timescales govern behavior:  $\tau_{\text{au1}} \sim 0.3\text{s}$  for excitation,  $\tau_{\text{au2}} \sim 4\text{s}$  for suppression - Model is robust across experimental conditions - Biological interpretability with equivalent goodness of fit

---

### Slide 24: Conclusions — Follow-Up Study

**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 20-30% 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 26-30: FAQ

### Prepared Answers

**Q: What is the sequence of processes in the original study?** Data collection → MAGAT trajectory extraction → Event detection → Population kernel fitting → LOEO validation

**Q: What processes were used in the follow-up study?** Individual MLE fitting → K-means/hierarchical clustering → Round-trip validation → Power analysis → Identifiability analysis

**Q: Why does population modeling succeed but individual fails?** Data-to-parameter ratio. Population pools ~15,000 events for 6 parameters (2500:1). Individual uses ~25 events for 6 parameters (4:1).

**Q: What is hierarchical shrinkage?** Bayesian regularization that pulls individual estimates toward the population mean proportionally to data sparsity.

**Q: How should clustering results be interpreted?** With extreme skepticism. K-means will always produce k clusters. The gap statistic shows k=1 is optimal. Round-trip validation shows ARI < 0.2.

---

### Anticipated Questions and Answers

**Q: Could a different kernel form work better for individual phenotyping?** A: Unlikely. The problem is data quantity and quality, not kernel form. Simpler models like single-timescale exponentials might help by reducing parameters.

**Q: What about using machine learning instead of kernel fitting?** A: ML methods face the same fundamental problem. With 25 events per track, there is insufficient information to distinguish individuals regardless of the algorithm.

**Q: How confident are you in the 100-event threshold?** A: The 100-event threshold comes from simulation-based power analysis targeting 80% power for a 0.2-second effect. Different effect sizes would require different thresholds.

**Q: Are there any individual larvae that do show reliable phenotypes?** A: Yes, 8.6% of tracks show individual variation exceeding measurement noise. These are the “outliers” that retain individual estimates after hierarchical shrinkage. But 8.6% is too few for systematic phenotyping.

**Q: What is the next step for this research?** A: Implement burst stimulation protocol and collect new data with 40+ minute recordings. Rerun the phenotyping analysis with the improved data.

---

### Timing Guide

Slides	Section	Target Time
1-2	Introduction	2 min
3-7	PSTH/Simulation Methods	5 min
8-10	Original Study Results	6 min

Slides	Section	Target Time
11-17	Follow-Up Study	10 min
18-22	Recommendations	5 min
23-25	Conclusions	3 min
26-30	FAQ (if needed)	5 min

**Total: 31-36 minutes**

---

### Technical Terms to Define if Asked

**Gamma-difference kernel:** Difference of two gamma distributions, one fast (excitatory) and one slow (suppressive)

**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