Presenter's Companion

Sensorimotor Habituation in *Drosophila* Larvae

Population-Level Modeling and Individual Phenotyping Validation

Gil Raitses
Syracuse University

December 2025

Total Slides	27
Target Duration	30 minutes + Q&A
Presentation Arcs	6
Version	v3

Contents

Arc 1: Introduction

Slides 1–2 · Target: 2 minutes

Arc Purpose
Establish context and preview the two-part structure. The audience should understand that the presentation covers both a successful population-level model and a failed attempt to extend it to individual phenotyping.

Core Message
Larval reorientation dynamics can be modeled with a parametric kernel that has biologically interpretable timescales. The same model fails at the individual level due to data sparsity.

Audience Anchor
Two timescales govern behavior: $\tau_1 \approx 0.3$s for excitation and $\tau_2 \approx 4$s for suppression.

Slide 1.1 Title Slide

Opening Statement

"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."

Timing

30 seconds. Do not linger. Move directly to slide 2.

Slide 1.2 Executive Summary — Original Study

Points to Emphasize

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

Key Point

If the audience remembers only one thing from the original study, it should be: Larval reorientation dynamics follow a simple parametric model with biologically interpretable timescales.

Transition

"Let me show you the kernel structure that captures these dynamics."

Arc 2: Original Study — Population-Level Modeling

Slides 3–7 · Target: 8 minutes

Arc Purpose
Present the gamma-difference kernel model, validate it against empirical data, demonstrate habituation dynamics, and reveal the first hint of individual-level problems through LOEO validation.

Core Message
The population-level model works. It captures the essential temporal structure of sensorimotor transformation. But LOEO validation shows that individual experiments are highly variable, foreshadowing the individual-level problems.

Key Figures

Slide 3	Kernel Structure	Central theoretical construct
Slide 4	Simulation Validation	Generative model works
Slide 5	Habituation Dynamics	Behavioral phenomenon
Slide 6	Behavioral States	Detailed state breakdown
Slide 7	LOEO Validation	First hint of problems

Slide 2.1 Kernel Structure

Figure Walkthrough

- Left panel: Combined kernel showing full temporal response
- Right panel: Decomposition into fast gamma (green) and slow gamma (red)
- Fast component peaks at 0.3 seconds and drives immediate response
- 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."

Timing

2 minutes. Spend time here—this is the core theoretical contribution.

Slide 2.2 Simulated vs Empirical Event Counts

Validation Message

"Before using the model for anything, we 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

Key Point

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

Timing

1 minute. Quick validation slide—do not over-explain.

Slide 2.3 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: Strongest habituation, slope +0.031 per pulse
- 50-250 conditions: Weaker effects
- Shaded bands: 95% confidence intervals

Interpretation

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

Timing

1.5 minutes. Link habituation to the kernel.

Slide 2.4 Behavioral State Analysis

State Breakdown

- Gray: Forward running
- Pink: Turning
- Blue: Pausing (remains below 5%)
- Orange: Reverse crawling

Key Point

Habituation manifests as increased turning, not increased pausing or freezing. By pulse 17, larvae spend nearly 40% of time turning versus 20% at pulse 0.

Timing

1.5 minutes. Emphasize that pausing stays low.

Slide 2.5 Leave-One-Experiment-Out Validation

What This Shows

"LOEO 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 next."

Transition

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

Timing

2 minutes. This is the pivot point to the follow-up study.

Arc 3: Follow-Up Study — Individual Phenotyping Failure

Slides 8–14 · Target: 10 minutes

Arc Purpose
Present the negative result: Individual phenotyping fails. Explain why: data sparsity and identifiability problems. Demonstrate that apparent clusters are artifacts.

Core Message
The model is not wrong—the data is insufficient. With only 25 events per track and 6 parameters to estimate, the problem is fundamentally underdetermined. The solution requires more data or simpler models.

Key Figures

Slide 8	Executive Summary	Frame the negative result
Slide 9	Clustering Illusion	Clusters are artifacts
Slide 10	Data Sparsity	Root cause explanation
Slide 11	Hierarchical Shrinkage	Partial mitigation
Slide 12	Identifiability	Protocol design matters
Slide 13	Stimulation Protocols	Visual comparison
Slide 14	Model Comparison	Gamma-diff vs flexible

Slide 3.1 Executive Summary — Follow-Up Study

Points to Emphasize

- The answer to individual phenotyping is **negative** with current protocols
- Sparse data: Only 18–25 events per track
- Apparent clusters are statistical artifacts
- Only 8.6% of tracks show genuine individual differences

Key Point

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.

Timing

1.5 minutes. Set expectations clearly.

Slide 3.2 The Clustering Illusion

Figure Walkthrough

- Panel A: PCA reveals unimodal distribution, not discrete clusters

- Panel B: All four validation methods fail with ARI < 0.13
- Panel C: Gap statistic minimized at $k = 1$—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."

Key Point

Publishing these clusters would be misleading. They are artifacts of sparse data, not genuine biological phenotypes.

Timing

1.5 minutes. Emphasize the validation failures.

Slide 3.3 Data Sparsity Explains Instability

The Math Problem

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

Key Number

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

Timing

1.5 minutes. The math is simple—make it memorable.

Slide 3.4 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 individual estimates."

Limitation

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

Timing

1 minute. Quick explanation of mitigation attempt.

Slide 3.5 The Identifiability Problem

Figure Walkthrough

- Panel A: Continuous design produces high bias and RMSE
- Panel B: Burst design extracts **10× more Fisher Information**
- Panel C: MLE recovery differs dramatically by design
- Panel D: Continuous fails because inhibition dominates during LED-ON

Key Point

The problem is not just data quantity but data quality. Continuous 10-second pulses produce events during the suppressive phase. These events carry almost no information about τ_1.

Timing

2 minutes. This is the key mechanistic insight.

Slide 3.6 Stimulation Protocol Comparison

Four Designs

- A: Current continuous 10s ON, 20s OFF
- B: Recommended burst 10×0.5s
- C: Alternative 4×1s
- D: Alternative 2×2s

Key Number

Burst design provides **8× more Fisher Information** than continuous.

Timing

1 minute. Visual comparison—quick.

Slide 3.7 Kernel Model Comparison

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 variance with half the parameters. The timescales τ_1 and τ_2 represent genuine temporal structure, not curve-fitting artifacts."

Transition

"Now let me present five recommendations for enabling individual phenotyping in future work."

Timing

1 minute. Justification for model choice.

Arc 4: Recommendations for Future Work

Slides 15–19 · Target: 5 minutes

Arc Purpose
Transform the negative result into actionable guidance. Present five concrete recommendations that would enable individual phenotyping in future experiments.

Core Message
The failure is not fundamental—it is fixable. Protocol modification (burst stimulation), extended recording (40+ minutes), model simplification, alternative phenotypes, and within-condition analysis can together solve the problem.

Recommendation Hierarchy

Priority	Recommendation	Slide
Primary	Protocol Modification (Burst)	15
Secondary	Extended Recording	16
Tertiary	Model Simplification	17
Alternative	Simpler Phenotypes	18
Methodological	Within-Condition Analysis	19

Slide 4.1 Recommendation 1 — Protocol Modification

Primary Recommendation

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

Quantitative Benefit

This modification alone could reduce required events from 100 to approximately 30.

Implementation

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

Timing

1 minute. Emphasize this is the highest-leverage intervention.

Slide 4.2 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.

Timing

45 seconds. Quick—complements Rec 1.

Slide 4.3 Recommendation 3 — Model Simplification

Approach

"Reduce parameter space by fixing population-derived parameters."

Specific Suggestion

- Fix τ_2 at population estimate of 3.8 seconds
- Fix amplitude ratio B/A at population value
- Estimate only τ_1 per individual track

Timing

45 seconds. One free parameter may be enough.

Slide 4.4 Recommendation 4 — Alternative Phenotypes

Pragmatic Alternative

"Use robust composite phenotypes that avoid kernel fitting entirely."

Examples

- ON/OFF event ratio
- First-event latency

Timing

45 seconds. Simpler may be better.

Slide 4.5 Recommendation 5 — Within-Condition Analysis

Methodological Point

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

Why

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

Timing

45 seconds. Important caveat for future analyses.

Arc 5: Conclusions

Slides 20–22 · Target: 3 minutes

Arc Purpose
Summarize both studies. Leave the audience with a clear understanding of what succeeded, what failed, and why.

Core Message
Population-level modeling succeeds because pooled data provides sufficient statistical power. Individual phenotyping fails because the same model is overparameterized for sparse individual data. The solution requires either more data per individual or simpler models.

Take-Home Numbers

$\tau_1 \approx 0.3$s	Fast excitatory timescale
$\tau_2 \approx 4$s	Slow suppressive timescale
$R^2 = 0.968$	Kernel fit quality
25 events/track	Current data
100 events/track	Required for phenotyping
8.6%	Tracks with genuine individual variation
10×	Information gain from burst protocol

Slide 5.1 Conclusions — Original Study

Summary of Success

- Gamma-difference kernel accurately models population-level dynamics
- Two timescales: $\tau_1 \approx 0.3$s for excitation, $\tau_2 \approx 4$s for suppression
- Model is robust across experimental conditions
- Biological interpretability with equivalent goodness of fit

Timing

1 minute.

Slide 5.2 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

Key Point

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

Timing

1.5 minutes.

Slide 5.3 Thank You

Transition to Questions

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

Timing

30 seconds.

Arc 6: Frequently Asked Questions

Slides 23–27 · Target: 5 minutes (if needed)

Arc Purpose
Provide prepared answers to anticipated questions. Use only if time permits or if specific questions arise.

Prepared Questions

1. What is the sequence of processes in the original study?
2. What processes were used in the follow-up study?
3. Why does population modeling succeed but individual fails?
4. What is hierarchical shrinkage?
5. How should clustering results be interpreted?

Slide 6.1 FAQ — Original Study Methods

Prepared Answer
Data collection → MAGAT trajectory extraction → Event detection → Population kernel fitting → LOEO validation

Slide 6.2 FAQ — Follow-Up Study Methods

Prepared Answer
Individual MLE fitting → K-means/hierarchical clustering → Round-trip validation → Power analysis → Identifiability analysis

Slide 6.3 FAQ — Why Population Succeeds but Individual Fails

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

Slide 6.4 FAQ — What is Hierarchical Shrinkage

Prepared Answer
"Bayesian regularization that pulls individual estimates toward the population mean proportionally to data sparsity. It is optimal regularization under the assumption that individuals are exchangeable members of a population."

Slide 6.5 FAQ — How to Interpret Clustering Results

Prepared Answer

"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. Clusters are artifacts, not phenotypes."

Arc A: Timing Summary

Arc	Slides	Target Time
1. Introduction	1–2	2 min
2. Original Study	3–7	8 min
3. Follow-Up Study	8–14	10 min
4. Recommendations	15–19	5 min
5. Conclusions	20–22	3 min
6. FAQ (if needed)	23–27	5 min
Total	27 slides	28–33 min

Arc B: Technical Glossary

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

PSTH
Peri-stimulus time histogram; 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

Arc C: Anticipated Questions

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 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 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 larvae that do show reliable phenotypes?

A: Yes, 8.6% of tracks show variation exceeding measurement noise. These are the outliers that retain individual estimates after 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 phenotyping analysis with improved data.