

# **WHERE INTERPRETABILITY WORKS AND WHERE IT BREAKS**

**A Speaker's Screenplay**

Gil Raitses  
Syracuse University

Mirna Lab General Meeting  
December 2025

## Contents

1	ACT ONE: THE POPULATION STORY	3
2	INTERLUDE: THE GAMMA PARAMETERS	5
3	ACT TWO: WHERE IT BREAKS	6
4	INTERRUPTION HANDLING	8
5	ACT THREE: FISHER INFORMATION	9
6	FALLBACK EXPLANATIONS	11
7	ACT FOUR: THE CLOSE	12
8	QUESTION HANDLING	14
9	FINAL REMINDER	16

# 1 ACT ONE: THE POPULATION STORY

## OPENING — BEFORE SLIDE 1

*Take a breath. Stand still. Make eye contact with the room.*

What I want to do today is walk through how I've been thinking about this data, rather than just show results. The lab collects complex behavioral data, and the question I've been asking is how far interpretability can be pushed before structure gets overread. I'm using modeling, simulation, and validation the same way controls are used in experiments, to figure out what kinds of explanations are justified at different scales.

*Small pause. Let the frame settle.*

What you'll see is that this works really well at the population level. The analysis produces clean, interpretable structure that lines up nicely with established thinking about sensory response and adaptation. But what's more interesting to me is what happens when the same ideas get pushed down to individuals. That's where things stop working in a useful way, and that failure turns out to be very informative.

[Advance to Slide 1]

So I'm going to start at the population level, where everything behaves nicely, and then move to the individual level to show where the boundaries are and how they might be pushed experimentally.

*Pause. Point to the slide.*

This is sensorimotor behavior in larvae under optogenetic stimulation, pooled across experiments and conditions.

[Advance to Slide 2]

At this scale, the behavior is very regular. When

the light turns on, reorientation probability rises quickly and then drops over time as the system adapts. You can see that across conditions and across experiments. The point here isn't just that the model fits, it's that the structure is consistent enough to summarize in a small number of parameters that actually mean something biologically.

[Advance to Slide 3]

This kernel is just a compact way of describing that structure over time. There's a fast component right after light onset and a slower component that suppresses behavior later on. I'm not going to dwell on the math here. Conceptually, this is just a way of separating early response from longer term adaptation so they don't get mixed together.

*Small pause.*

At the population level, this representation is very stable, and that's what makes it useful.

[Advance to Slide 4]

This slide shows the transition from counting events to modeling behavior as a time varying probability. The histogram tells you what happened. The model lets you ask what would happen under different timing, which becomes important later when I talk about simulation.

*Pause.*

Up to this point, everything is working well, and this is the version of the story that's comfortable.

*Stop. End of first two minutes.*

## 2 INTERLUDE: THE GAMMA PARAMETERS

### TECHNICAL CLARIFICATION — IF ASKED

*This section is for when someone asks about the gamma difference parameters. Deliver calmly.*

Each gamma component is describing the shape of a response over time, very much like how you'd describe the envelope of a sound. There is an onset or attack phase, a peak or sustain region, and a decay. But instead of naming those explicitly, the gamma distribution uses parameters that jointly control those features.

*Brief pause.*

Conceptually, each gamma component answers three questions. When does this process start to matter after the stimulus. How sharply does it rise to its peak. How long does its influence persist before fading out.

Those three properties together define the temporal gesture of the process. In the fast component, that gesture is brief and early. In the slow component, it is delayed and long lasting. The amplitudes then set how much each gesture contributes relative to the other.

*If someone wants a one-liner:*

Each gamma component defines a temporal envelope, how fast it turns on, when it peaks, and how long it lasts.

*If someone pushes mathematically:*

The shape and scale parameters jointly control rise time and decay, so they are not redundant even though they both affect timing.

*That's enough. Do not go deeper in a general lab talk.*

### 3 ACT TWO: WHERE IT BREAKS

#### THE PIVOT — AFTER SLIDE 4

*You've just said the model lets you ask what would happen under different timing.*

At this point, everything is still very comfortable. The population data are dense, the structure is clear, and the model behaves as expected.

[Advance to Slide 8]

Here's one example of that. Across repeated light pulses, turning behavior gradually increases. That slow change is exactly what the later part of the kernel is capturing. So the model isn't just fitting a curve, it's tracking something already recognized as habituation.

[Advance to Slide 9]

You can see the same thing from a different angle here. As pulses progress, larvae spend less time running and more time turning. Again, this is population level structure, and it's very consistent.

*Pause.*

Up to here, the story is straightforward. The model is doing something sensible.

*Now slow down before advancing.*

The natural question at this point, and the one I was most interested in, is whether these same parameters can tell us something meaningful about individual larvae.

[Advance to Slide 11]

This is where things start to change.

At first glance, it looks promising. When you fit the same model to individual tracks and cluster the parameters, you appear to get distinct groups.

[Advance to Slide 12]

But when you actually test whether those groups are stable or reproducible, they fall apart. Different methods disagree, resampling doesn't preserve structure, and the data look continuous rather than clustered.

*Pause.*

This is not because the clustering algorithm is bad. It's because the analysis is asking more of the data than they can support.

[Advance to Slide 13]

Most individual larvae only produce about twenty re-orientation events. Trying to estimate a full temporal shape from that is mathematically underdetermined.

*Pause here. Let it land.*

## 4 INTERRUPTION HANDLING

### WHEN SOMEONE INTERRUPTS

*If someone asks "why not just fit fewer parameters" or "why not trust the clusters":*

That's exactly what I tried next, and it turns out to be very revealing.

[Then advance to Slide 14]

*General interruption deflection:*

That's a good question, and I'm going to answer it explicitly in about two slides, so let me show you the failure mode first.

*This keeps you in control without shutting anyone down.*

*If someone asks a simulation question too early:*

I'll get to simulation shortly, but the short version is that it helps explain why this breaks, not rescue it.

*The pivot sentence that carries you from failure into design insight:*

Once the individual level fails, the question stops being how to fit this model, and becomes what needs to change experimentally for this to work at all.

*This sentence is the bridge into identifiability, Fisher information, and protocol redesign.*



## 5 ACT THREE: FISHER INFORMATION

### THE INFORMATION GEOMETRY — AFTER SLIDE 14

*You are on Slide 14. Pause briefly.*

So at this point, the model is telling something important. When the analysis tries to estimate individual parameters, almost everything collapses back to the population mean. That's not a failure of the statistics. It's a sign that the individual data just don't contain enough information.

[Advance to Slide 15]

This slide is about why that happens.

What matters here isn't just how many events are collected, but when those events occur relative to the stimulus. Some events are informative about the fast response, and others really aren't.

*Point gently, don't trace curves.*

In the continuous stimulation design, most events happen during periods dominated by suppression. That means they tell us almost nothing about the fast timescale, even though they still count as events.

*Pause.*

So you can collect more data and still not learn the thing you care about.

[Advance to Slide 16]

This is where stimulation geometry starts to matter. By breaking the light into brief pulses, you repeatedly sample the early response window. Each event now carries more information about the fast process.

*Pause.*

This is the key idea. It's not that burst stimulation

magically creates data. It just makes each event more useful for the question being asked.

*If someone looks confused, use this translation:*

You can think of it like probing a system at the wrong time versus the right time. Same number of measurements, very different insight.

[Advance to Slide 18]

Once you frame it this way, you can actually ask practical questions. How many events are needed. How long must recording last. What kind of protocol would make individual differences detectable at all.

*Pause.*

Under the current design, individual phenotyping just isn't feasible. Under a redesigned protocol, it might be.

[Advance to Slide 20]

And when information is limited, the model naturally simplifies. Instead of trying to recover a full temporal shape, the data support much simpler summaries, things like response latency or ON versus OFF bias.

*Pause.*

That's not giving up. That's being honest about what the data can tell us at this scale.

## 6 FALLBACK EXPLANATIONS

### FISHER INFORMATION QUESTIONS

**If asked: “What is Fisher information?”**

It’s a way of quantifying how sensitive the data are to a parameter. If changing a parameter barely changes what you observe, you can’t estimate it reliably.

**If asked: “Why not just collect more animals?”**

More animals help only if the experiment samples the right part of the response. Otherwise you just get more of the same uninformative data.

**If asked: “Isn’t this circular since you assume the model?”**

The simulation isn’t claiming the model is true. It’s asking whether, if the model were true, the experiment could ever detect its parameters.

*This last sentence is extremely strong and shuts down the circularity worry cleanly.*

### BRIDGE SENTENCE

Once you see where information comes from, the question shifts from fitting models to designing experiments that make interpretation possible.

*This is the intellectual heart of your talk.*

## 7 ACT FOUR: THE CLOSE

### FINAL THREE MINUTES – AFTER SLIDE 21

*Slow your pace slightly here.*

So stepping back, what this analysis really shows is not just what works, but why it works at one scale and not another. At the population level, behavior is dense and regular enough that interpretable structure emerges cleanly. At the individual level, the same approach stops being informative, and that boundary turns out to be very sharp.

*Pause.*

What I find useful about this is that it tells where effort is best spent. Instead of forcing increasingly complex models onto sparse data, the options are to simplify what gets estimated or redesign the experiment to extract information more efficiently.

**[Advance to the final summary slide]**

Simulation plays a very specific role here. It doesn't replace experiments, and it doesn't tell us what biology is doing. It helps us see whether a question is even answerable with a given protocol, and what would need to change for it to become answerable.

*Pause.*

In that sense, the simulations aren't predictions. They're diagnostics.

*Short pause.*

So the takeaway for me is that interpretability isn't something you add after the fact. It's something that emerges when the scale of analysis, the structure of the data, and the experimental design line up.

*Now stop advancing slides.*

## SENSORIMOTOR HABITUATION

That's where I'll stop, and I'm very happy to talk through questions or dig into any part of the pipeline.

## 8 QUESTION HANDLING

### FIRST QUESTION PROTOCOL

*The first question almost always sets the tone.*

*If the first question is broad or vague:*

That's a good place to zoom in. Are you asking more about the modeling side, the simulation side, or the experimental implications?

*This lets the questioner self-specify and gives you control.*

*If the first question is aggressive or skeptical:*

That concern is exactly what motivated this analysis, so let me answer it in the context of what worked and what didn't.

*Then answer briefly and stop.*

*If you don't want to answer in detail yet:*

I don't want to give a sloppy answer to that, but I can tell you what the analysis says at a high level and details can be followed up afterward.

*This reads as integrity, not evasion.*

### RECOVERY SENTENCE

*If discussion starts to wander, use this to refocus:*

The main point I wanted to make today is that the limits being observed are not failures of modeling, they're measurements of where the experiments stop being informative.

## SENSORIMOTOR HABITUATION

*This sentence brings everything back into focus.*

## 9 FINAL REMINDER

### BEFORE YOU WALK IN

*Read this to yourself before entering the room.*

*You are not defending a model.*

*You are presenting a careful investigation  
into what your data can and cannot support.*

*That stance is unusually strong,  
especially in a lab setting,  
and it will be recognized as such.*

**END OF SCREENPLAY**