Cognitive Psychometrics:

Building hierarchical Bayesian measurement models

Adriana Felisa Chávez De la Peña

Overview

General framework:

1. Bayesian methods in Cognitive Science
2. Cognitive Psychometrics: Implementing cognitive models as measurement tools
3. Cognitive process models of choice and response time

Overview

General framework:

1. Bayesian methods in Cognitive Science
2. Cognitive Psychometrics: Implementing cognitive models as measurement tools
3. Cognitive process models of choice and response time

One main project:
Developing a hyper-efficient hierarchical Bayesian EZ-DDM

1. Introduce the model
2. Showcase applications to perception data

Other projects: Circular Drift Diffusion Model

Bayesian methods in Cognitive Science

Bayes' Theorem

$$P(A \mid B) = \frac{P(B \mid A) \times P(A)}{P(B)}$$

A principled framework for **updating our beliefs about an event being true given some evidence**.

Bayes' Theorem

$$P(A \mid B) = \frac{P(B \mid A) \times P(A)}{P(B)}$$

- $P(A)$: The **prior probability** of event A.
- $P(B \mid A)$: The **likelihood** of observing evidence B given event A.
- $P(B)$: The **marginal likelihood** of the evidence B.
- $P(A \mid B)$: The **posterior probability** of event A given evidence B.

Three Main Applications to Cognitive Science

1) **“Bayes in the Mind”**

   Applications:

   - Optimal cue integration
   - Probabilistic inference
   - Causal reasoning

Three Main Applications to Cognitive Science

2) Bayesian Data Analysis

$$P(H_0 \mid \text{Data}) = \frac{P(\text{Data} \mid H_0) \times P(H_0)}{P(\text{Data})}$$

- $P(H_0)$: The prior probability of the null hypothesis.
- $P(\text{Data} \mid H_0)$: The likelihood of the data under the null hypothesis.
- $P(H_0 \mid \text{Data})$: The posterior probability of the null hypothesis given the data.

Three Main Applications to Cognitive Science

2) Bayesian Data Analysis

$$P(H_0 \mid \text{Data}) = \frac{P(\text{Data} \mid H_0) \times P(H_0)}{P(\text{Data})}$$

- $P(H_0)$: The prior probability of the null hypothesis.
- $P(\text{Data} \mid H_0)$: The likelihood of the data under the null hypothesis.
- $P(H_0 \mid \text{Data})$: The posterior probability of the null hypothesis given the data.

Since hypotheses are expressed in terms of parameter values (e.g., $\theta = 0$), we can rewrite the equation as:

$$P(\theta \mid \text{Data}) \propto P(\text{Data} \mid \theta) \times P(\theta)$$

Three Main Applications to Cognitive Science

3) Bayesian Cognitive Modeling

$$P(\theta \mid \text{Data}) \propto P(\text{Data} \mid \theta) \times P(\theta)$$

- Specify prior distributions for all model parameters.
- Specify the likelihood function of the data given the model parameters.

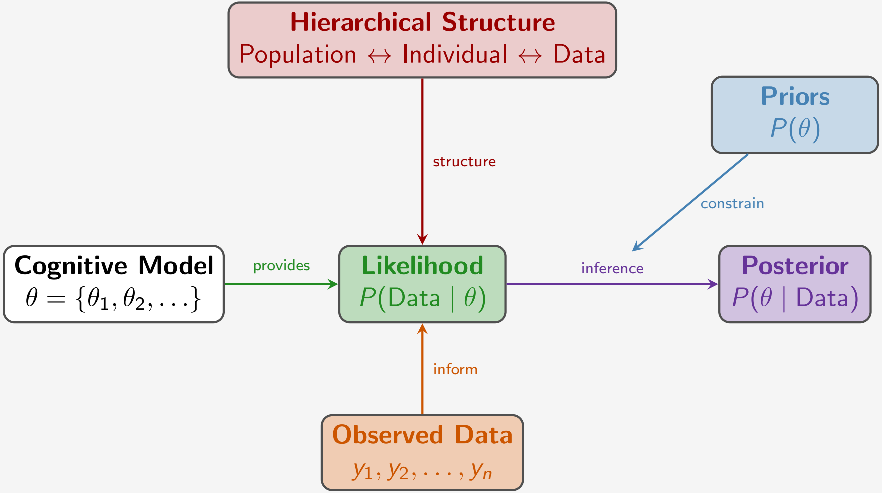
Three Main Applications to Cognitive Science

3) Bayesian Cognitive Modeling

$$P(\theta \mid \text{Data}) \propto P(\text{Data} \mid \theta) \times P(\theta)$$

- Specify prior distributions for all model parameters.
- Specify the likelihood function of the data given the model parameters.
- Add hierarchical structures that allow us to capture variability across individuals, conditions, and interventions.

Implementing Cognitive Models in a Bayesian Framework



Why use Bayesian inference?

1. **Quantify uncertainty through probability distributions**
   - Full posterior distributions provide **credible intervals**
   - Direct **probability statements** about parameters (e.g., “95% probability that θ is between 0.3 and 0.5”)

Why use Bayesian inference?

1. **Quantify uncertainty through probability distributions**
   - Full posterior distributions provide **credible intervals**
   - Direct **probability statements** about parameters (e.g., "95% probability that θ is between 0.3 and 0.5")
2. **Model Comparison and Selection**
   - Bayes factors quantify evidence for competing models

Why use Bayesian inference?

1. **Quantify uncertainty through probability distributions**
   - Full posterior distributions provide **credible intervals**
   - Direct **probability statements** about parameters (e.g., "95% probability that θ is between 0.3 and 0.5")

2. **Model Comparison and Selection**
   - Bayes factors quantify evidence for competing models

3. **Naturally accommodate hierarchical structures and individual differences**
   - Partial pooling: borrow strength across individuals while respecting individual variation
   - Simultaneously estimate population-level and individual-level parameters

Cognitive Psychometrics

Cognitive models as measurement tools

Cognitive modeling: Traditional approach

Cognitive models are fitted to obtain **individual-level parameter estimates**.

Cognitive modeling: Traditional approach

Cognitive models are fitted to obtain **individual-level parameter estimates**.

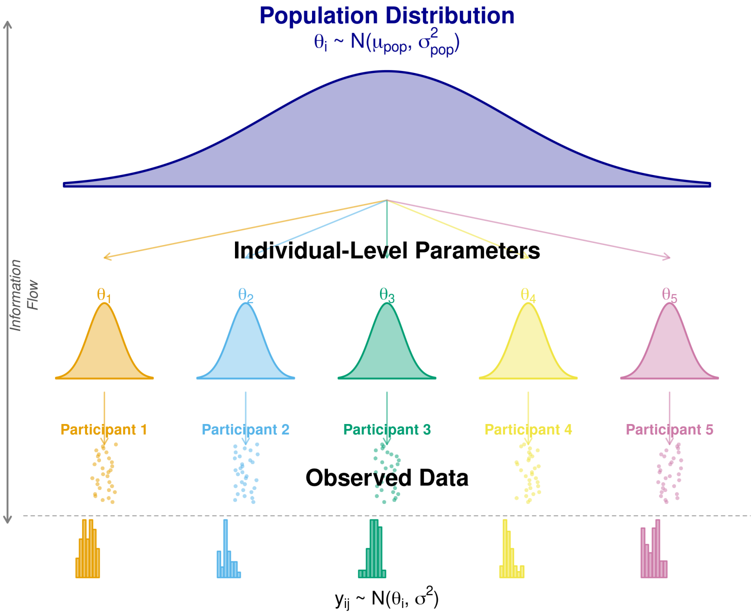
- **Group-level performance** is described and compared using **aggregate measures**.
- **Individual differences** are described in terms of the variability in the estimates observed.

This implies a multi-step approach that neglects the uncertainty associated with parameter estimation.

Hierarchical Bayesian cognitive models

Parameter values across levels of variation are assumed to be sampled from parent distributions.

- Model variability between individuals, tasks, treatments, conditions...
- Consider measurement error

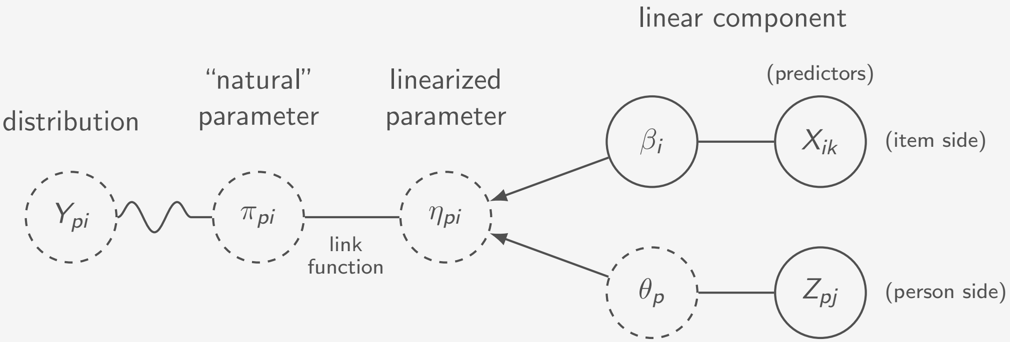

Population Distribution
$\theta_i \sim N(\mu_{pop}, \sigma^2_{pop})$
Individual-Level Parameters
Information Flow
θ_1
θ_2
θ_3
θ_4
θ_5
Participant 1
Participant 2
Participant 3
Participant 4
Participant 5
Observed Data
$y_{ij} \sim N(\theta_i, \sigma^2)$

Cognitive Psychometrics

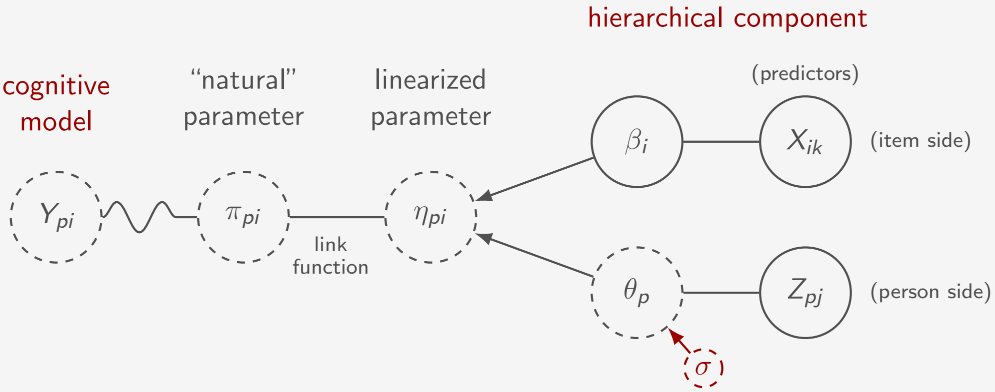
Main focus: The application of cognitive models as psychological measurement tools.

- Benefits from implementing cognitive models within a Bayesian framework.
- Hierarchical Bayesian structures allow us to capture **variability across levels of interest** (e.g., individual differences, interventions, subpopulations, etc.).
- Hierarchical structures can be extended to incorporate **regression structures** to explain the observed variability as a function of covariates of interest (e.g., age, levels of cognitive impairment, test scores, etc.).
- This approach fits nicely with the Explanatory Item Response Model framework.

Explanatory cognitive process models



Explanatory cognitive process models



Case example:

Models of Choice and Response Time

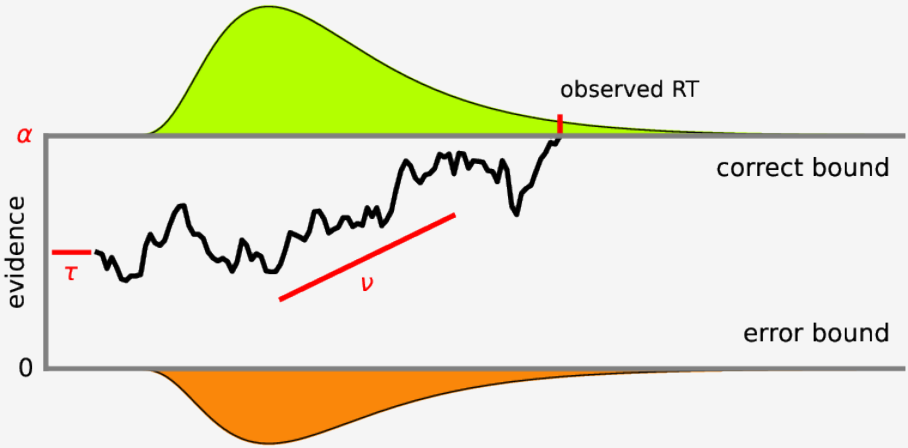
Evidence Accumulation models

Cognitive **process** models that describe decision-making as the result of a stochastic sequential sampling process.

There's always two decisions:

- 1) What to respond (**Choice**)
- 2) When to respond (**Response Time**)

The Drift Diffusion Model (DDM)



EZ-DDM (Wagenmakers et al., 2007)

Forward EZ equations

Let $y = \exp(-\alpha\nu)$.

$$\tilde{R} = \frac{1}{y+1}$$

$$\tilde{M} = \tau + \left(\frac{\alpha}{2\nu}\right)\left(\frac{y-1}{y+1}\right)$$

$$\tilde{V} = \left(\frac{\alpha}{2\nu^3}\right)\left\{\frac{1 - 2\alpha\nu y - y^2}{(y+1)^2}\right\}$$

EZ-DDM (Wagenmakers et al., 2007)

Forward EZ equations

Let $y = \exp(-\alpha\nu)$.

$$\tilde{R} = \frac{1}{y+1}$$

$$\tilde{M} = \tau + \left(\frac{\alpha}{2\nu}\right)\left(\frac{y-1}{y+1}\right)$$

$$\tilde{V} = \left(\frac{\alpha}{2\nu^3}\right)\left\{\frac{1 - 2\alpha\nu y - y^2}{(y+1)^2}\right\}$$

Inverse EZ equations:

Let $L = \log\left(\frac{\dot{R}}{1-\dot{R}}\right)$

$$\hat{\nu} = \text{sign}\left(\dot{R} - \frac{1}{2}\right) \times \cdots$$

$$\cdots \sqrt[4]{\frac{L\left(\dot{R}^2 L - \dot{R} L + \dot{R} - \frac{1}{2}\right)}{\dot{V}}}$$

$$\hat{\alpha} = \frac{L}{\hat{\nu}}$$

$$\hat{\tau} = \dot{M} - \left(\frac{\hat{\alpha}}{2\hat{\nu}}\right)\left[\frac{1 - \exp(-\hat{\nu}\hat{\alpha})}{1 + \exp(-\hat{\nu}\hat{\alpha})}\right]$$

Hyper-efficient hierarchical Bayesian measurement models of choice and response time

Bayesian implementation of the EZ-DDM

- The EZ-DDM provides deterministic estimators $\hat{\nu}$, $\hat{\alpha}$, $\hat{\tau}$.
- In order to build a Bayesian implementation of the EZ-DDM, we need to find *probabilistic estimators* and *a distribution over data* that is conditional on the model parameters.

Bayesian implementation of the EZ-DDM

- The EZ-DDM provides deterministic estimators $\hat{\nu}$, $\hat{\alpha}$, $\hat{\tau}$.
- In order to build a Bayesian implementation of the EZ-DDM, we need to find *probabilistic estimators* and *a distribution over data* that is conditional on the model parameters.

Solution:

- Use the *sampling distributions of the summary statistics* in the EZDDM to build a proxy model
- The proxy model allows for hierarchical Bayesian extensions.

Sampling distributions for the EZ summary statistics

Accuracy:

$$\dot{T} \sim \text{Binomial}\left(\tilde{R}, N\right)$$

Mean RT ($\dot{M}$):

$$\dot{M} \sim \text{Normal}\left(\tilde{M}, \frac{\tilde{V}}{N}\right)$$

Variance of all RTs ($\dot{V}$)

$$(N-1)\frac{\dot{V}}{\tilde{V}} \sim \text{Chi-squared}\,(N-1)$$

$$(N-1)\frac{\dot{V}}{\tilde{V}} \sim \text{Gamma}\left(\frac{N-1}{2}, 2\right)$$

$$\dot{V} \sim \text{Gamma}\left(\frac{N-1}{2}, \frac{2\tilde{V}}{N-1}\right)$$

As N becomes sufficiently large:

$$\dot{V} \sim \text{Normal}\left(\tilde{V}, \frac{2\tilde{V}^2}{N-1}\right)$$

The proxy model

$$\dot{T} \sim \text{Binomial}\left(\tilde{R}, N\right) \qquad \tilde{R} = \frac{1}{y+1}$$

$$\dot{M} \sim \text{Normal}\left(\tilde{M}, \frac{\tilde{V}}{N}\right) \qquad \tilde{M} = \tau + \left(\frac{\alpha}{2\nu}\right)\left(\frac{y-1}{y+1}\right)$$

$$\dot{V} \sim \text{Normal}\left(\tilde{V}, \frac{2\tilde{V}^2}{N-1}\right) \qquad \tilde{V} = \left(\frac{\alpha}{2\nu^3}\right)\left\{\frac{1-2\alpha\nu y - y^2}{(y+1)^2}\right\}$$

Applied example (Data by Ratcliff & Rouder, 1998)

Task: Is the overall brightness “high” or “low”?

Applied example (Data by Ratcliff & Rouder, 1998)

Task: Is the overall brightness "high" or "low"?

Design: There are 66 cells in the design following a **two-factor design**:

- **Accuracy vs Speed instructions**: Conditions 1–33 had an accuracy instruction, 34–66 a speed instruction.
- **More black vs More white pixels**: Conditions 1–16 and 34–49 had more black pixels; conditions 18–33 and 51–66 had more white, and conditions 17 and 50 were ambiguous and will not be used here because they can't provide accuracy measures.

The model

The model incorporates an effect β of instruction x_i on the α.

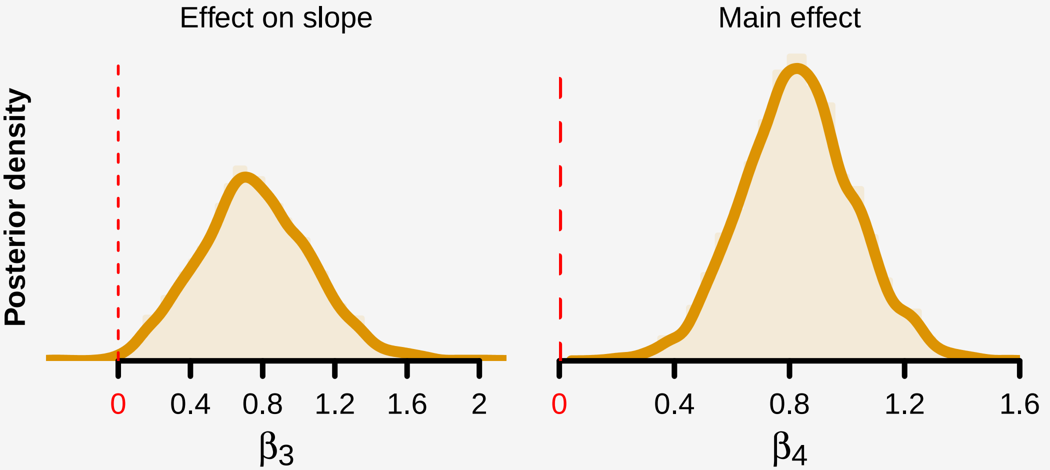
$$\alpha \sim \text{Normal}(\mu_\alpha + \beta X_i, \sigma_\alpha)$$

We also include a nonlinear regression on the drift rate δ using instruction x_i and stimulus configuration x_s as predictors.

$$\begin{aligned} Y &= \Phi\left(\beta_1 + \beta_2|X_s| + \beta_3 X_i|X_s|\right) \\ \delta_{\text{pred}} &= \mu_\delta + \beta_0 Y + \beta_4 X_i \\ \delta &\sim \text{Normal}\left(\delta_{\text{pred}}, \sigma_\delta\right) \end{aligned}$$

We focus on the effect of instruction (β_3 and β_4).

Effect of instruction on the drift rate



The model

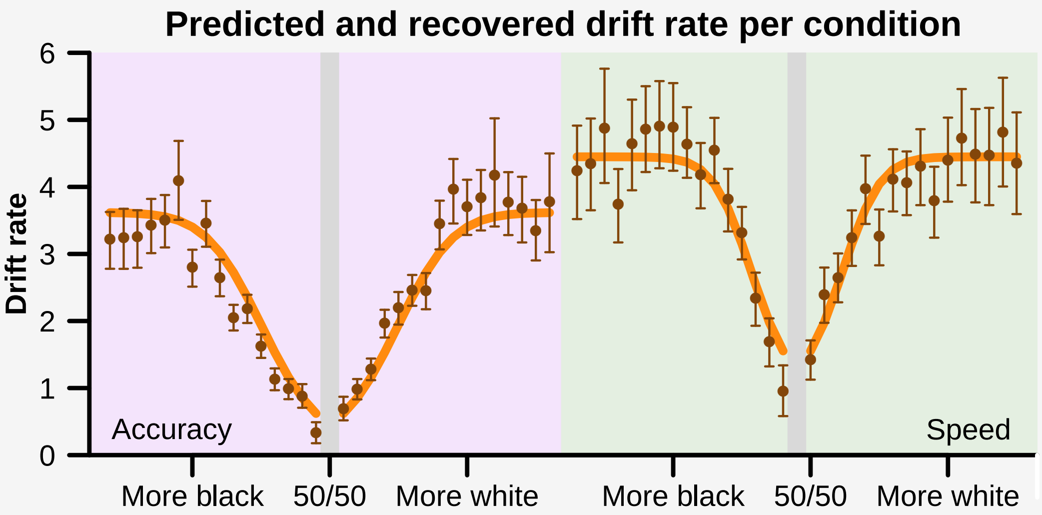
The model incorporates an effect β of instruction x_i on the α.

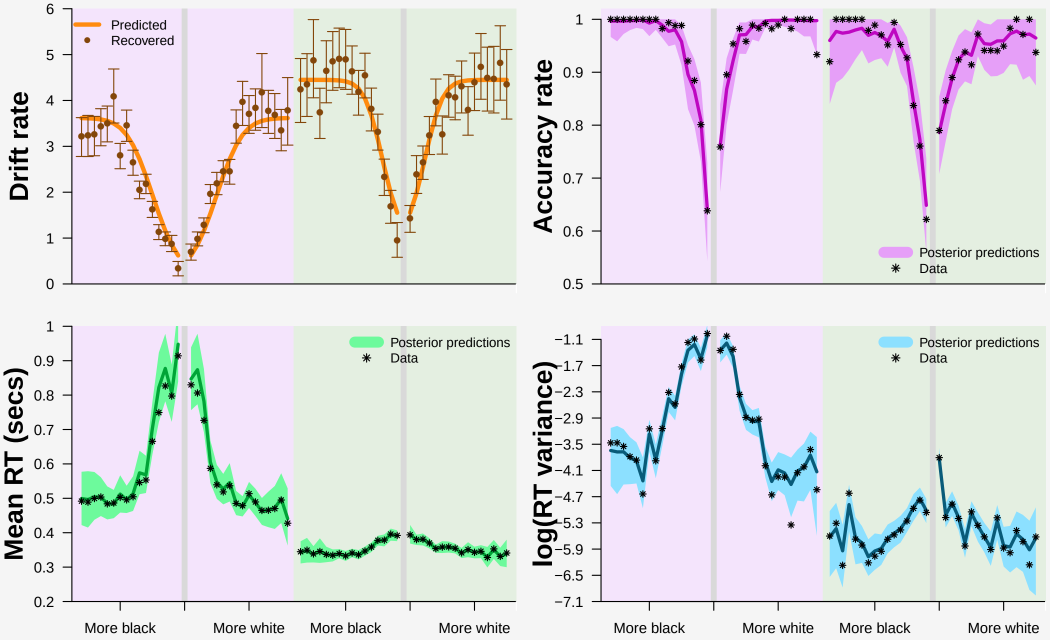
$$\alpha \sim \text{Normal}(\mu_\alpha + \beta X_i, \sigma_\alpha)$$

We also include a nonlinear regression on the drift rate δ using instruction x_i and stimulus configuration x_s as predictors.

$$\begin{aligned} Y &= \Phi\left(\beta_1 + \beta_2 |X_s| + \beta_3 X_i |X_s|\right) \\ \delta_{\text{pred}} &= \mu_\delta + \beta_0 Y + \beta_4 X_i \\ \delta &\sim \text{Normal}\left(\delta_{\text{pred}}, \sigma_\delta\right) \end{aligned}$$

Let us now inspect the predictions of the model for the drift rate.

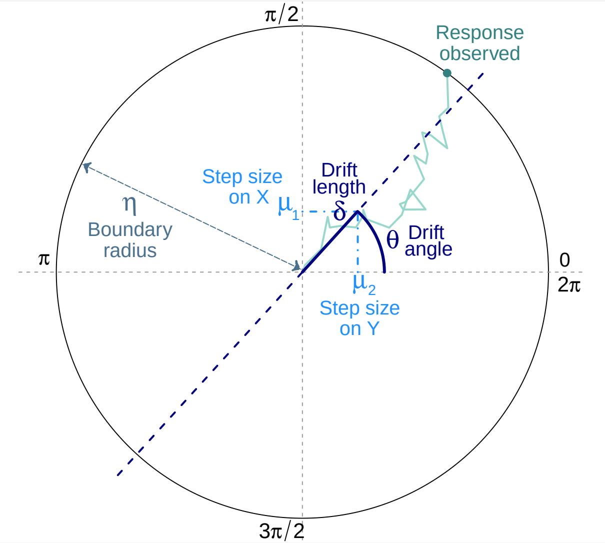

Predicted and recovered drift rate per condition
Drift rate
0
1
2
3
4
5
6
Accuracy
Speed
More black
50/50
More white
More black
50/50
More white



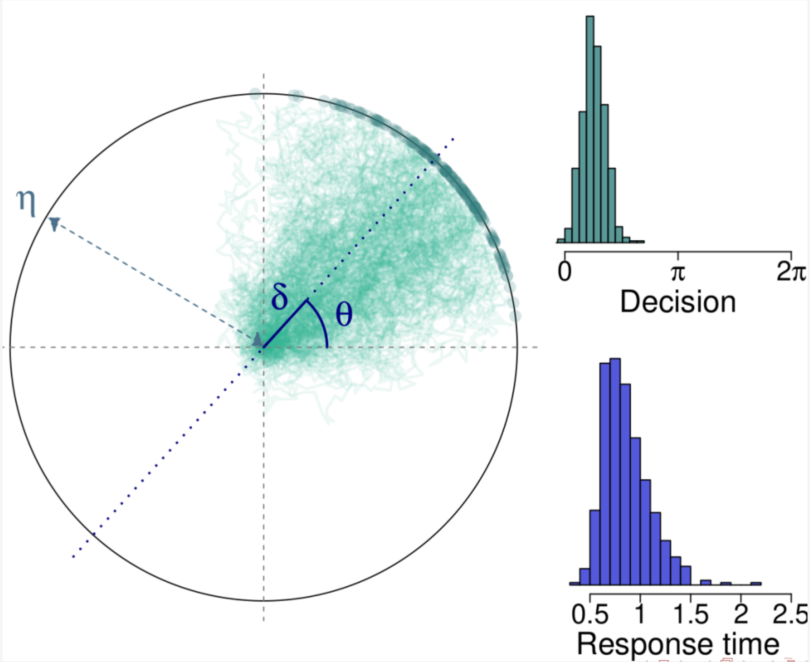
Drift rate
Predicted
Recovered
Accuracy rate
Posterior predictions
Data
Mean RT (secs)
Posterior predictions
Data
log(RT variance)
Posterior predictions
Data
More black
More white
More black
More white
More black
More white
More black
More white

This approach is widely generalizable across contexts and data types!

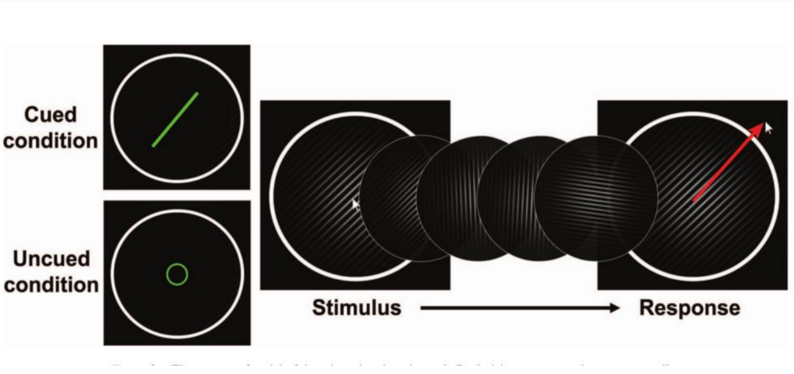
The Circular Drift Diffusion Model



- Nondecision time (τ): Visual encoding and motor control.
- Boundary radius (η): Criterion to be reached to make a decision.
- Drift vector ($\mu = \{\mu_x, \mu_y\}$): Mean step size on X and Y dimensions.


η
δ
θ
0
π
2π
Decision
0.5
1
1.5
2
2.5
Response time

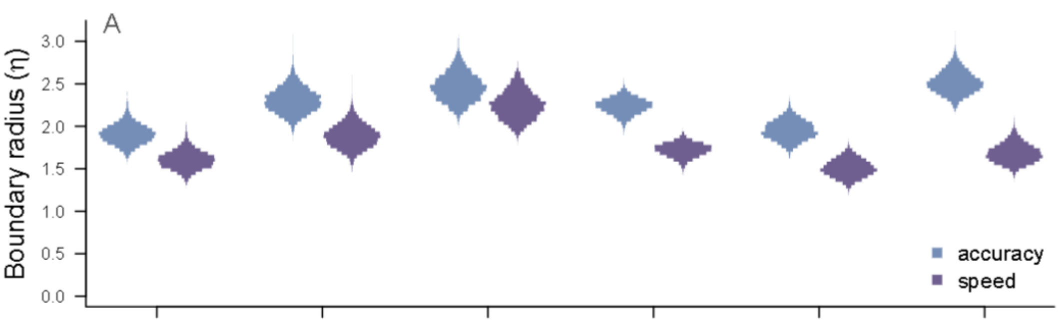
An orientation judgment task (Data by Kvam et al., 2019)



- Two instruction conditions: **Accuracy** and **Speed**.
- Two cue conditions: **Cue** and **No cue**.
- On cue conditions, cue could vary from true orientation by **15**, **30**, or **45** degrees.

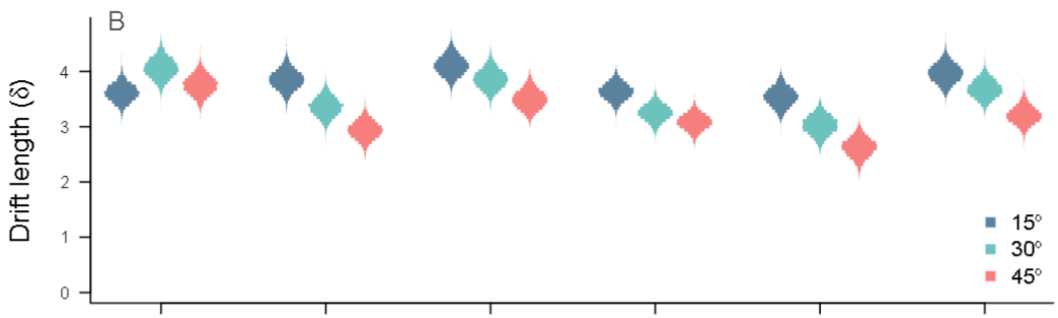
Application on real data (Data by Kvam et al., 2019)

Q1: Are participants more cautious when instructed to prioritize accuracy?



Application on real data (Data by Kvam et al., 2019)

Q2: Does the speed of information processing change with the task difficulty?



Conclusion remarks

- Cognitive model parameters carry **theoretical meaning** allowing for **interpretability**.
- Bayesian hierarchical structures naturally capture **variability across different levels of analysis**.
- **Measurement precision:** Incorporates covariates and regression structures to explain variability in cognitive processes
- **Generalizability:** Framework applies across diverse data types and research domains (perception, memory, decision-making, clinical assessment)

Thank you!

Adriana Felisa Chávez De la Peña

achavezd@uci.edu