Sequence Prediction

Until now, we've focused on complex sequences

- speech sounds (HMM)
- words (HMM)
- human judgments (CRF)
- location of person/car (particle filter)
- coal mining accidents (OCPD)
- stock market returns (OCPD)
- saccade perturbations (Kalman filter)

All but last have been with an AI focus

Today, more modeling of human behavior, with very simple, *binary* sequences

X X X X X X X _ _ _

X Y X Y X Y X _ _ _

Simple Choice Task

X → 1 Y → 2

Measure response latency

mean RT = 310 ms, with standard deviation = 25 ms

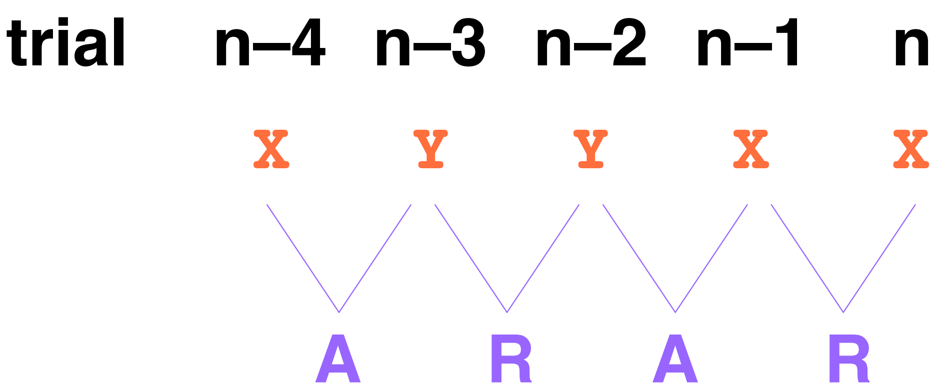
Simple Choice Task

X → 1 Y → 2

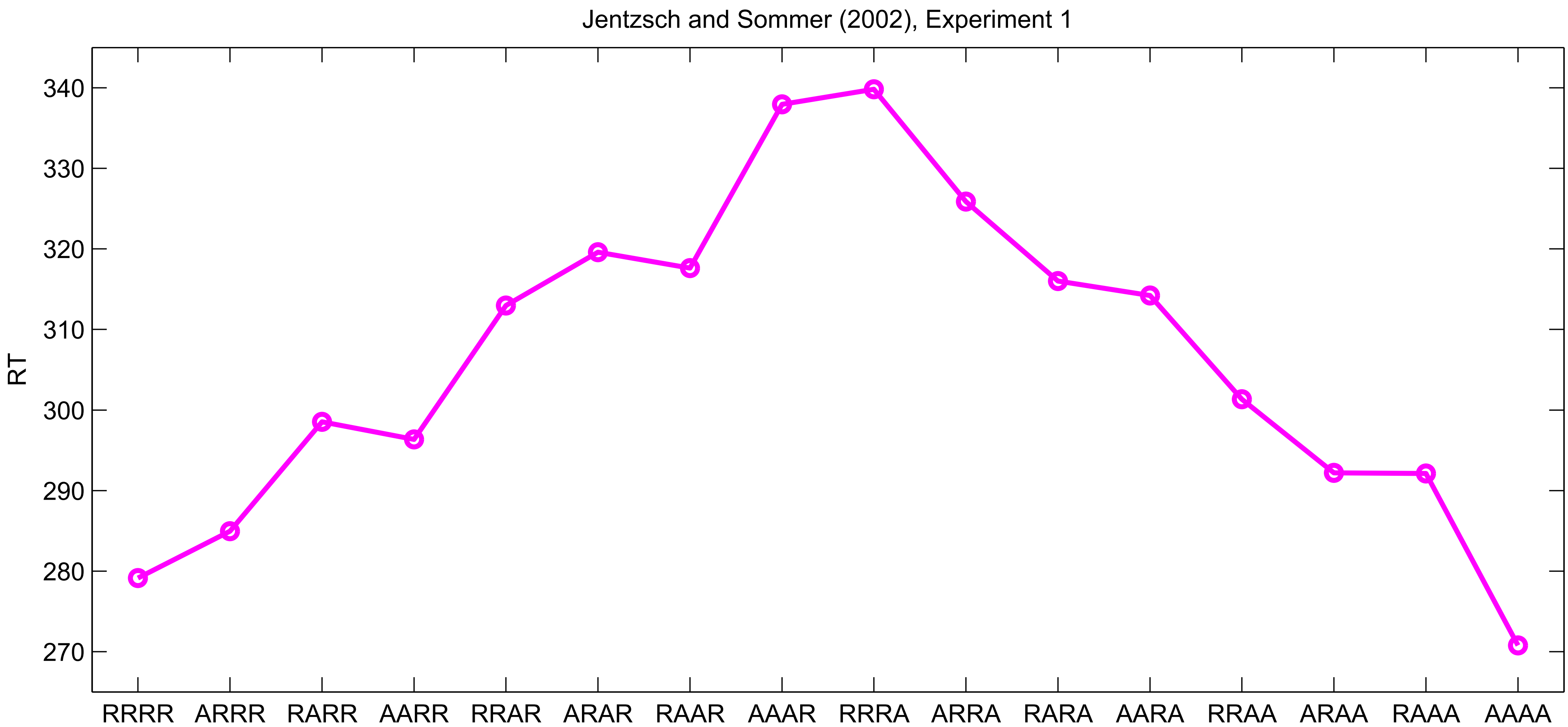
Measure response latency

mean RT = 310 ms, with standard deviation = 25 ms

Suppose we condition performance on recent history



Response Latencies Conditioned on History



Sequential effects

- explain significant variability in behavior
- give us insight into primitive learning mechanisms
- show how adaptive the brain is to a changing environment

What Sequence(s) Causes The Dependencies?

X ⟶ 1 Y ⟶ 2

Stimulus identity sequence	X X X Y Y X Y X Y Y
Response identity sequence	1 1 1 2 2 1 2 1 2 2

What Sequence(s) Causes The Dependencies?

X → 1 Y → 2

Stimulus repetition sequence R R A R A A A A R

Stimulus identity sequence X X X Y Y X Y X Y Y

Response identity sequence 1 1 1 2 2 1 2 1 2 2

Response repetition sequence R R A R A A A A R

What Sequence(s) Causes The Dependencies?

X → 1   Y → 2

Stimulus repetition sequence R R A R A A A A R

Stimulus identity sequence X X X Y Y X Y X Y Y

Response identity sequence 1 1 1 2 2 1 2 1 2 2

Response repetition sequence R R A R A A A A R

FIRST ORDER

What Sequence(s) Causes The Dependencies?

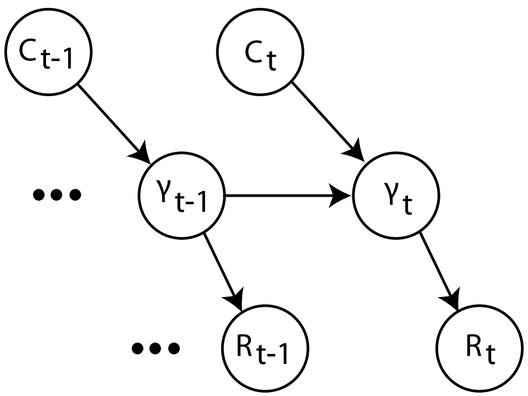
X → 1   Y → 2

Stimulus repetition sequence	R R A R A A A A R
Stimulus identity sequence	X X X Y Y X Y X Y Y
Response identity sequence	1 1 1 2 2 1 2 1 2 2
Response repetition sequence	R R A R A A A A R

SECOND ORDER

Dynamic Belief Network (Yu & Cohen, 2009)

Represents second-order (stimulus or response) sequence



$C \in \{0, 1\}$ changepoint

γ: repetition probability

$R \in \{r, a\}$

Model predicts next element in second-order sequence

Three parameters

changepoint prior α

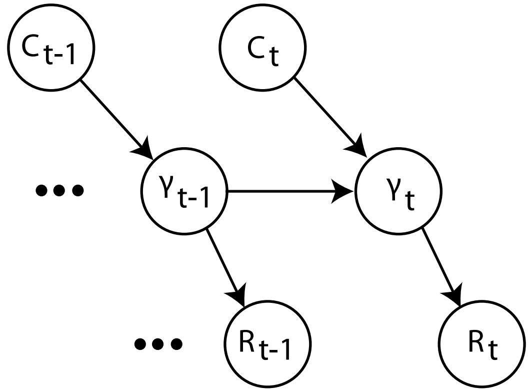
imaginary counts of Beta reset distribution for γ

Assumption

response time inversely related to probability of element that occurs

e.g., $P(R = a) = .7$ predicts fast response if next element is alternation

Inference In DBN



$C \in \{0, 1\}$ changepoint

γ : repetition probability

$R \in \{r, a\}$

Exact inference

$P(\gamma_t | R_1, ..., R_{t-1}) = P(C_t=1)\ Beta(\alpha,\beta) + P(C_t=0)\ P(\gamma_{t-1} | R_1, ..., R_{t-1})$

$P(\gamma_t | R_1, ..., R_t) \sim P(R_t \mid \gamma_t)\ P(\gamma_{t-1} | R_1, ..., R_{t-1})$

γ_t : mixture of beta distributions with t components

Note: related to linear space/time complexity of online changepoint detection

Linear space/time complexity ok for AI, not for cognitive models

Approximate inference

Model γ_t distribution as discrete in, e.g., {0.00, 0.01, 0.02, 0.03,...,1.00}.

Exact Inference

$$p(\gamma \mid \vec{X}_t) \sim p(X_t \mid \gamma)\, p(\gamma \mid \vec{X}_{t-1})$$

$p(X_t \mid \gamma)$: γ if $X_t = 1$ or $(1-\gamma)$ if $X_t = 0$; $\quad p(\gamma \mid \vec{X}_{t-1})$: $\sum w_i \,\text{Beta}(\alpha_i, \beta_i)$

consider $X_t = 1$

$$\sim \sum w_i \cdot \gamma \cdot \frac{\Gamma(\alpha_i + \beta_i)}{\Gamma(\alpha_i)\Gamma(\beta_i)} \gamma^{\alpha_i - 1} (1-\gamma)^{\beta_i - 1}$$

$$\sim \sum w_i \frac{\Gamma(\alpha_i + \beta_i)}{\Gamma(\alpha_i)} \cdot \frac{\Gamma(\alpha_i + 1)}{\Gamma(\alpha_i + \beta_i + 1)} \cdot \frac{\Gamma(\alpha_i + \beta_i + 1)}{\Gamma(\alpha_i + 1)\Gamma(\beta_i)} \gamma^{\alpha_i} (1-\gamma)^{\beta_i - 1}$$

$$\sim \sum w_i \frac{\Gamma(\alpha_i + 1)}{\Gamma(\alpha_i)} \frac{\Gamma(\alpha_i + \beta_i)}{\Gamma(\alpha_i + \beta_i + 1)} \text{Beta}(\alpha_i + 1, \beta_i)$$

$$= \sum v_i \,\text{Beta}(\alpha_i + 1, \beta_i)$$

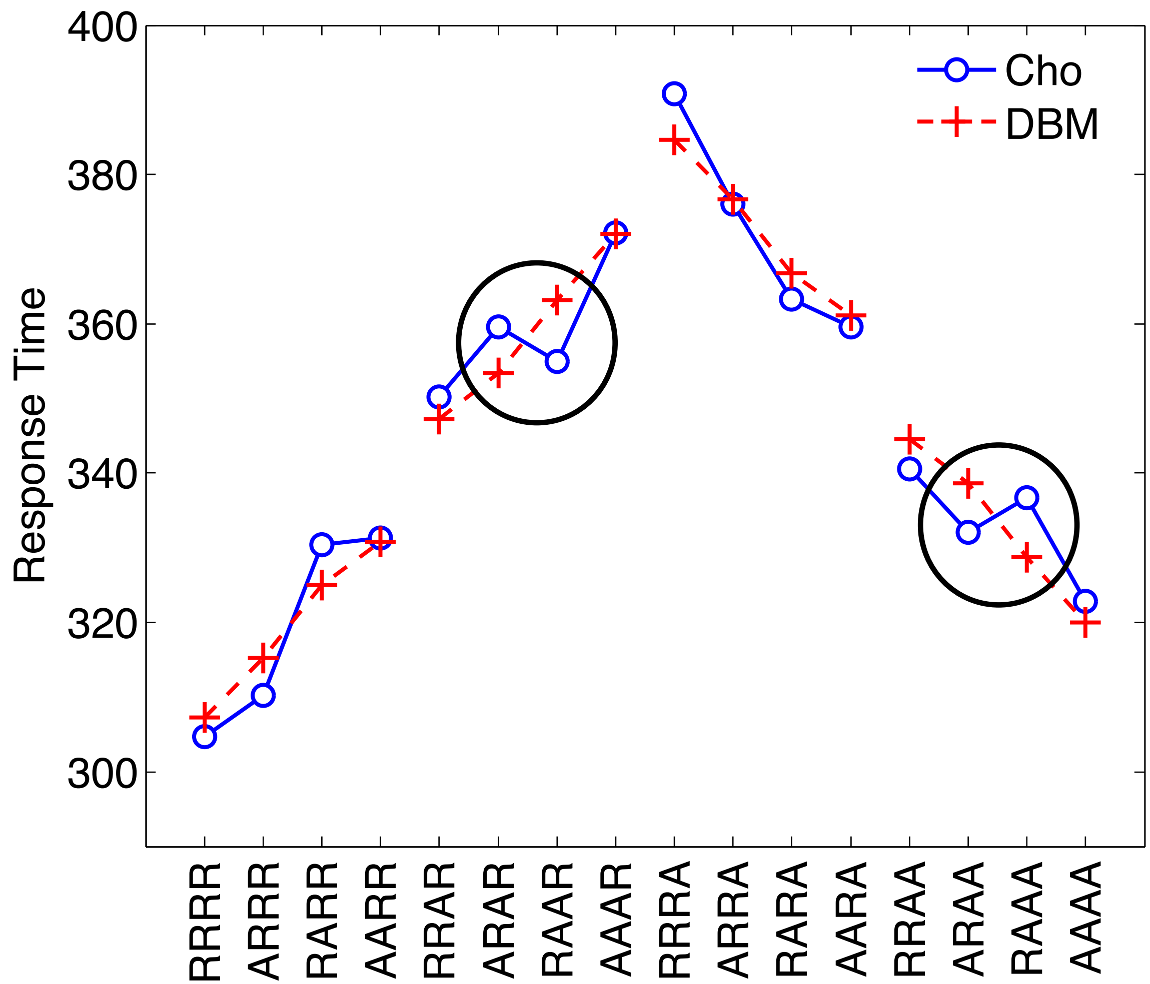
$$w'_i = w_i \frac{\Gamma(\alpha_i + 1)}{\Gamma(\alpha_i)} \cdot \frac{\Gamma(\alpha_i + \beta_i)}{\Gamma(\alpha_i + \beta_i + 1)}$$

$$= w_i \frac{\alpha_i!}{(\alpha_i - 1)!} \frac{(\alpha_i + \beta_i - 1)!}{(\alpha_i + \beta_i)!}$$

$$= w_i \cdot \alpha_i / (\alpha_i + \beta_i)$$

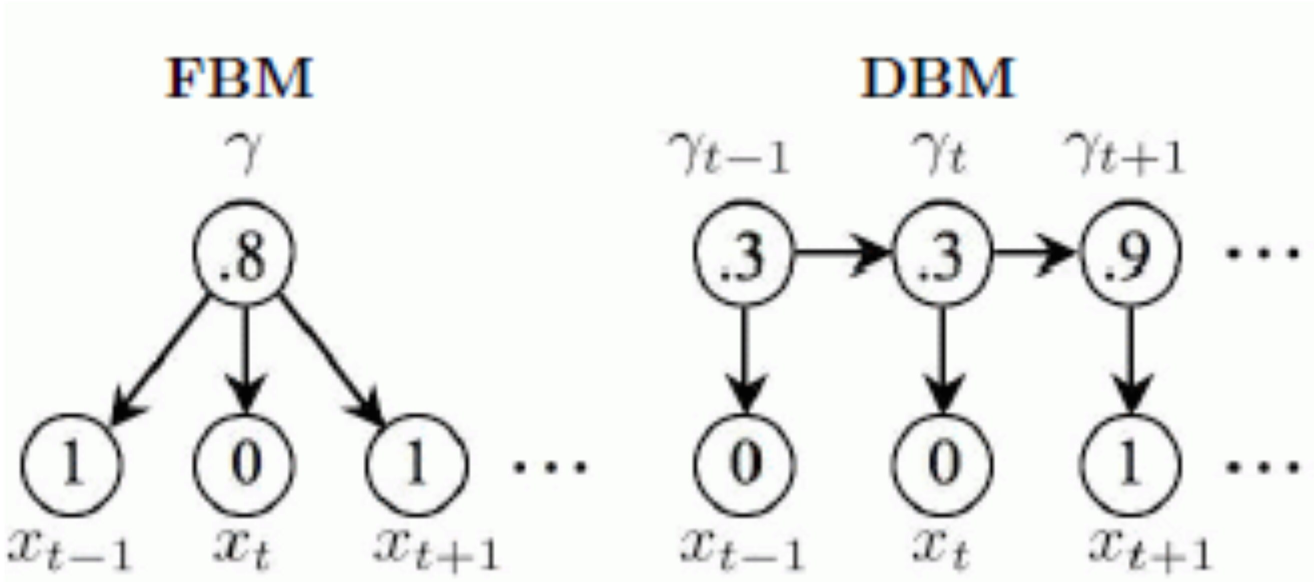
$$v_i = \frac{w'_i}{\sum w'_j}$$

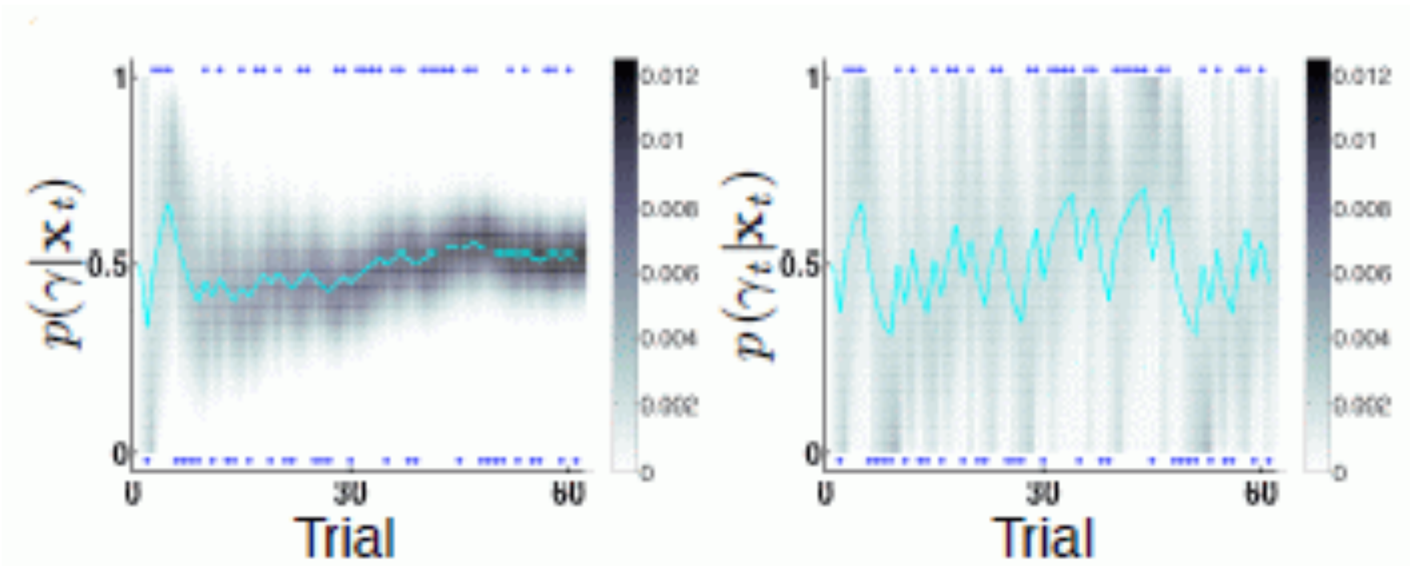
DBM Fit to Data of Cho et al. (2002)



Where does the asymmetry between R and A trials come from?

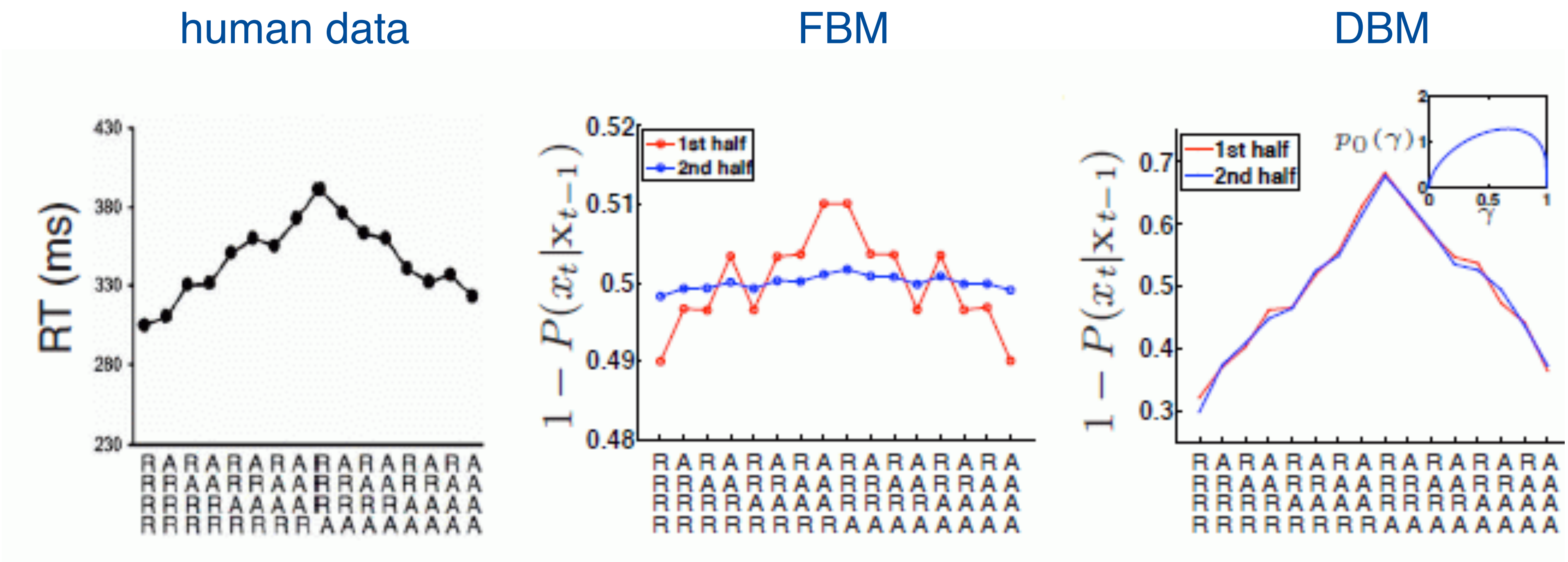
Dynamic Belief Model Versus Fixed Belief Model





FBM predicts less change in γ with experience
-> sequential effects diminish

Fixed Belief Model Fails To Fit Data



Conclusion:

Sequential effects are a *rational* behavior under the assumption of nonstationarity in the environment

Key Result (Yu & Cohen, 2009)

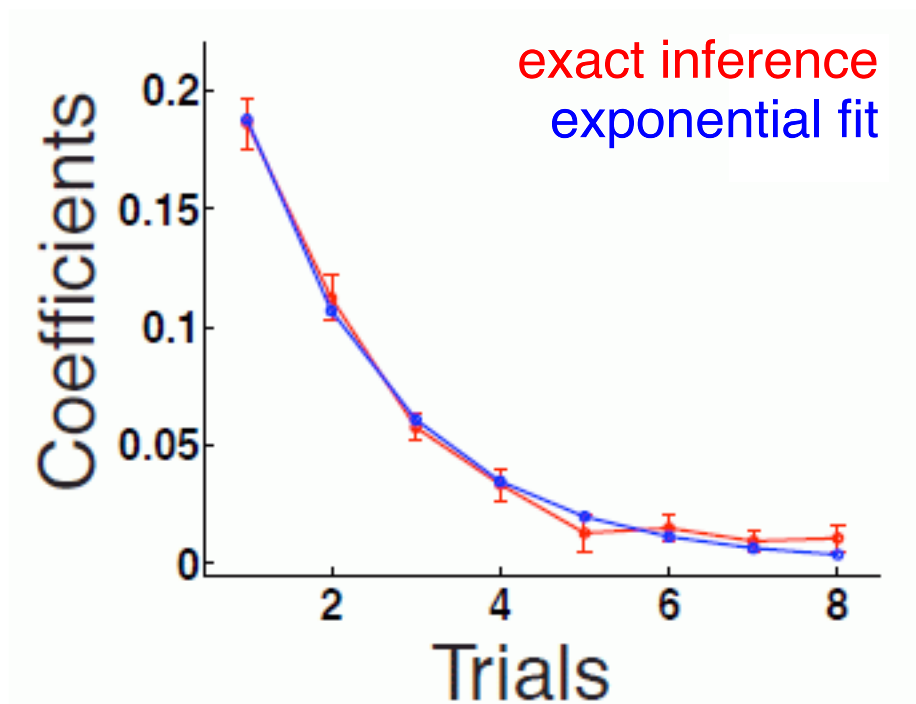
For most γ, DBM is well approximated by a model that maintains an exponentially decaying trace of recent repetitions/ alternations.

That is, if

- **$R_t = +1$ for repetition**
- **$R_t = -1$ for alternation,**

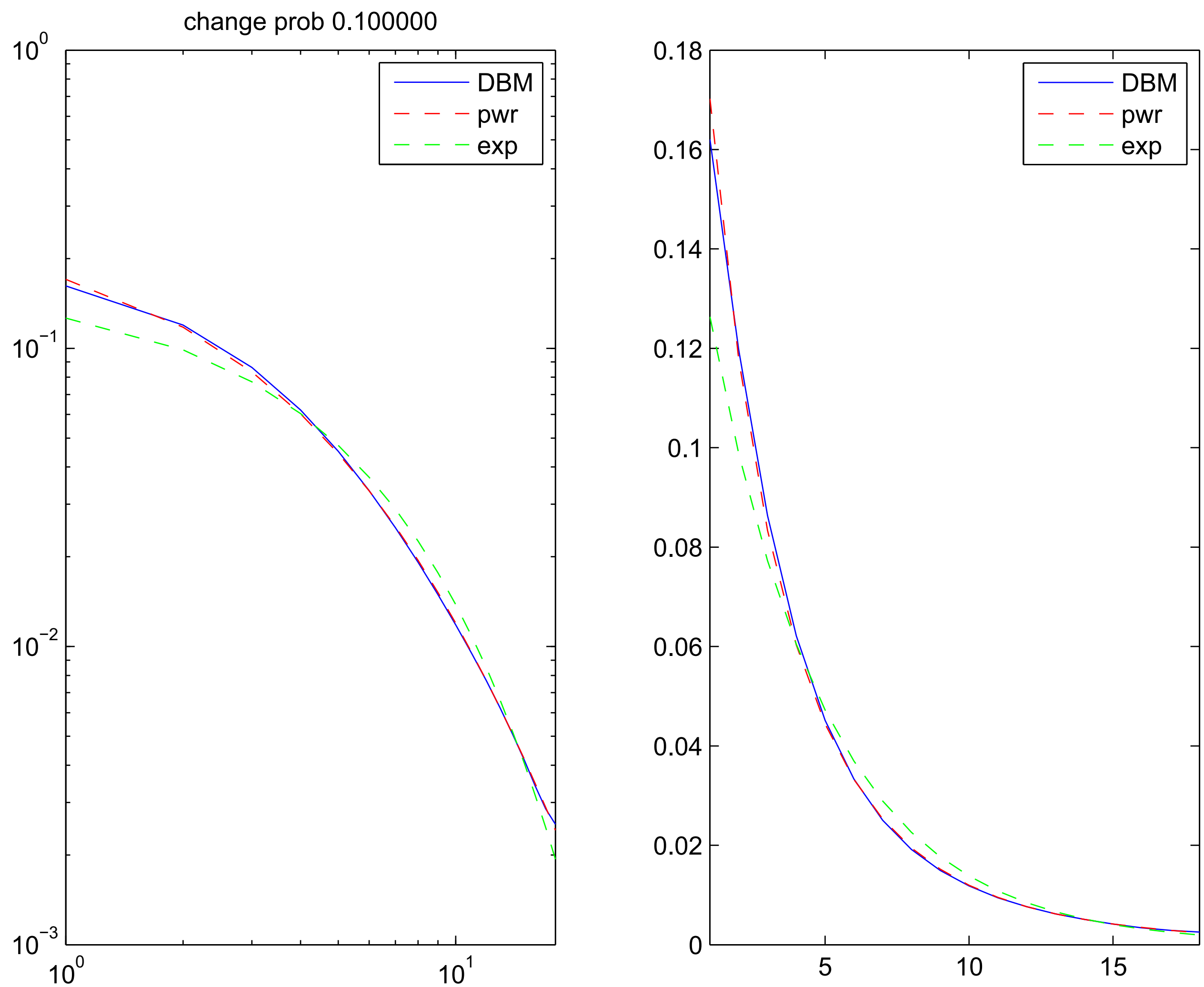
prediction of next trial under DBM is approximately

$$\overline{R}_{t+1} = \sum_{i=0}^{t} \gamma^i R_{t-i}$$

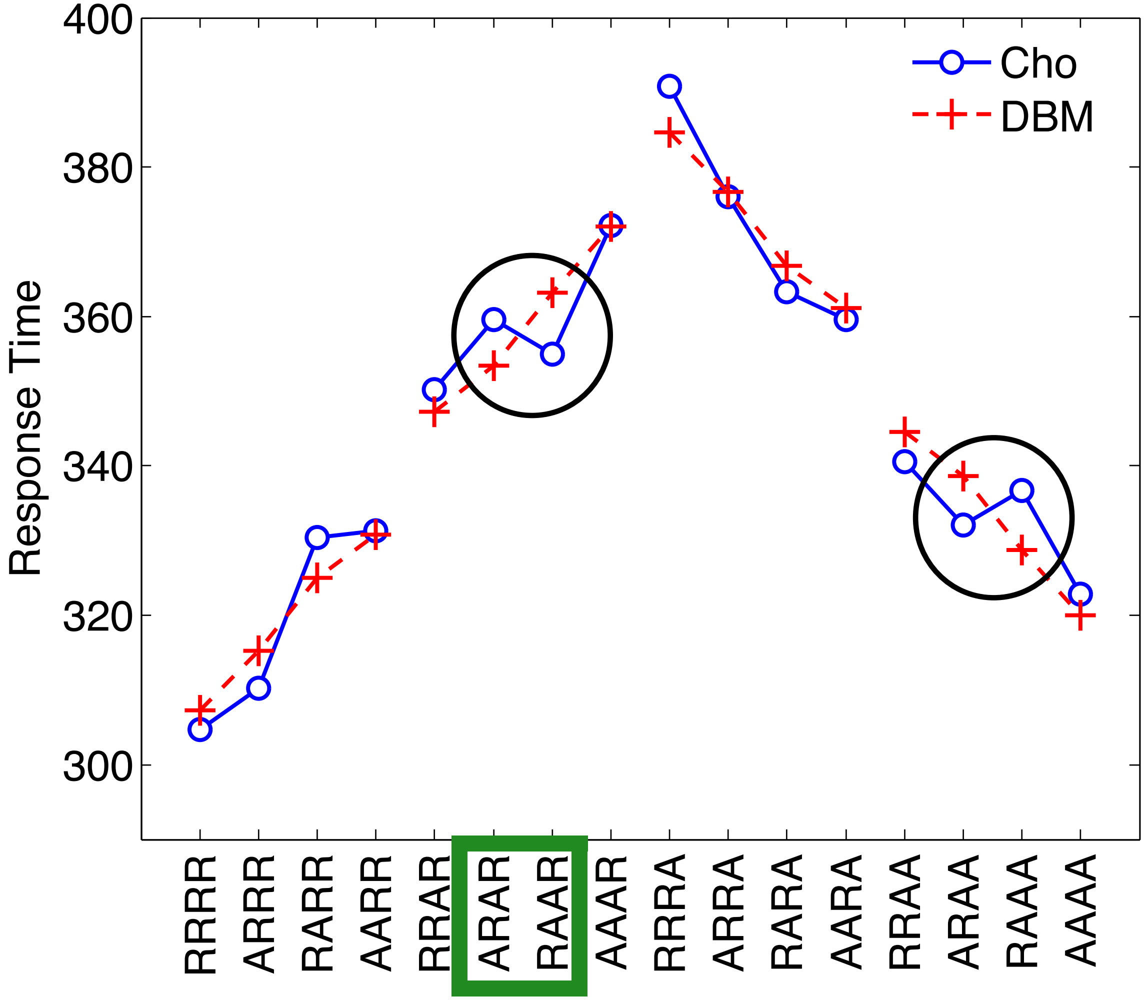


Exact Inference Revisited

Yu & Cohen sampled over histories, but with t-length histories, we can exhaustively sum over the 2^t possibilities



DBM Fit to Data of Cho et al. (2002)



Circled points: mismatch between model and data

First Versus Second Order Predictions

1st order sequence: trial *n-k* is same/different as trial *n*

e.g., XYYXX = YXXYY = SDDS

e.g., XXYXX = YYXYY = SSDS

2nd order sequence: trial *n-k* is a repetition/alternation of *n-k+1*

e.g., XYYXX = YXXYY = ARAR

e.g., XXYXX = YYXYY = RAAR

First Versus Second Order Predictions

1st order sequence: trial *n-k* is same/different as trial *n*

e.g., XYYXX = YXXYY = SDDS

e.g., XXYXX = YYXYY = SSDS

2nd order sequence: trial *n-k* is a repetition/alternation of *n-k+1*

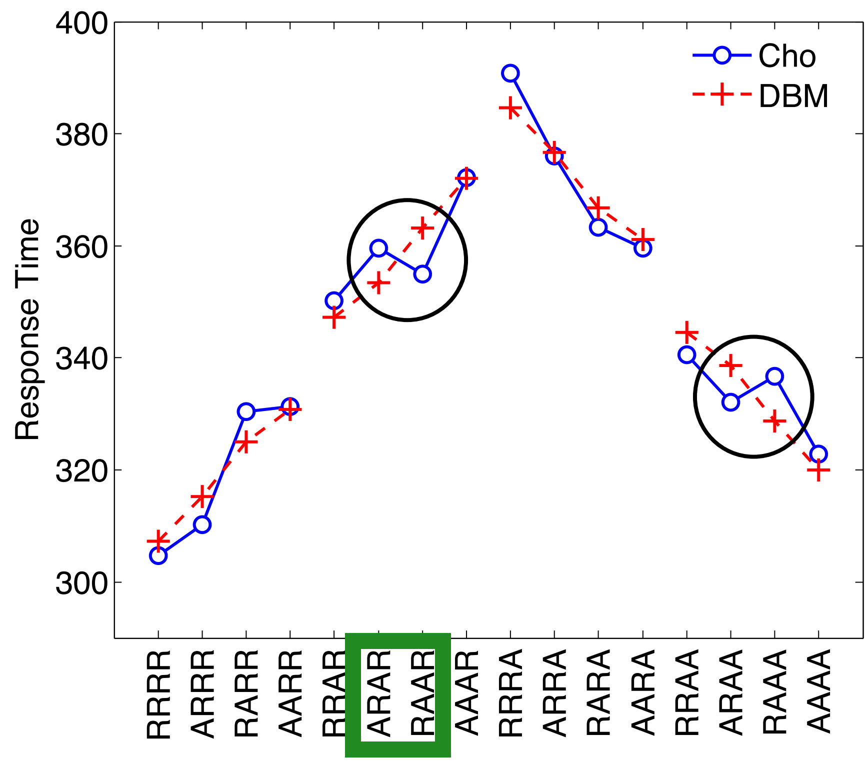
e.g., XYYXX = YXXYY = ARAR

e.g., XXYXX = YYXYY = RAAR

First and second order histories are one-to-one, but predictions can diverge.

P(next element is same as prev. | SSDS) > P(next element is same | SDDS)

P(next element is repetition | RAAR) < P(next element is alternation | ARAR)



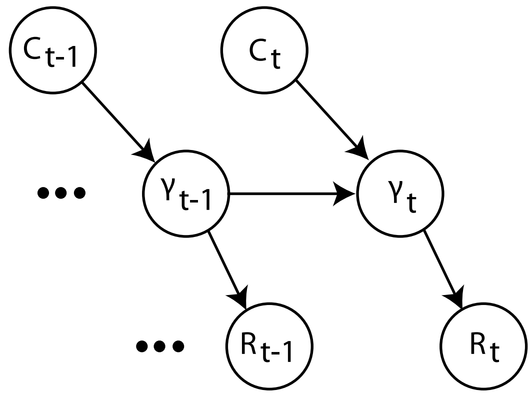
Cho et al. theorized that sequential dependencies in their data are due to *both* first and second order effects

- neural net leaky integrator model
- biased by recency in both first and second order sequences

Can the same type of account work within a more principled (i.e., DBM) framework?

DBM can represent first-order sequence just as well as second-order sequence

2nd order

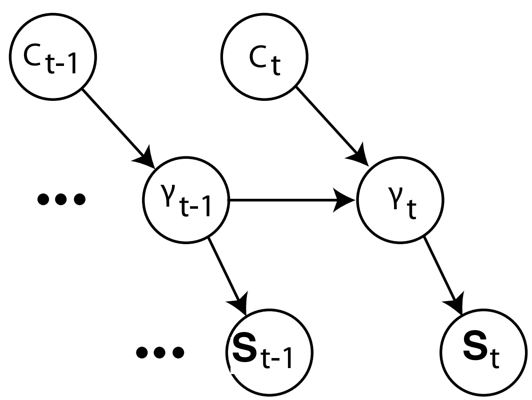


$C \in \{0, 1\}$ changepoint

γ: repetition probability

$R \in \{r, a\}$

1st order



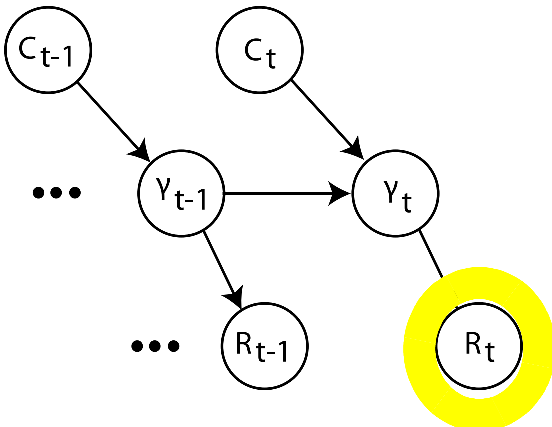
$C \in \{0, 1\}$ changepoint

γ: stimulus probability

$S \in \{x, y\}$

DBM can represent first-order sequence just as well as second-order sequence

2nd order

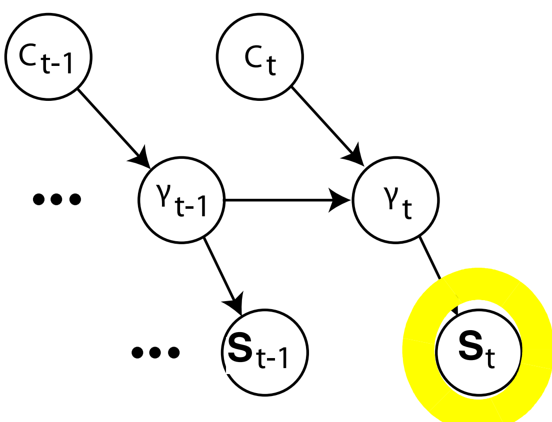


$C \in \{0, 1\}$ changepoint

γ : repetition probability

$R \in \{r, a\}$

1st order



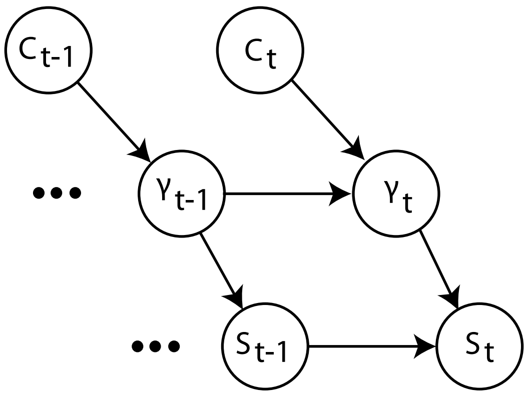
$C \in \{0, 1\}$ changepoint

γ : stimulus probability

$S \in \{x, y\}$

DBM can represent first-order sequence just as well as second-order sequence

2nd order

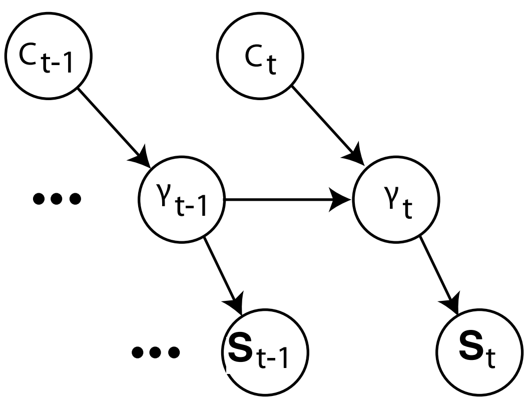


$C \in \{0, 1\}$ changepoint

γ : repetition probability

$S \in \{x, y\}$

1st order



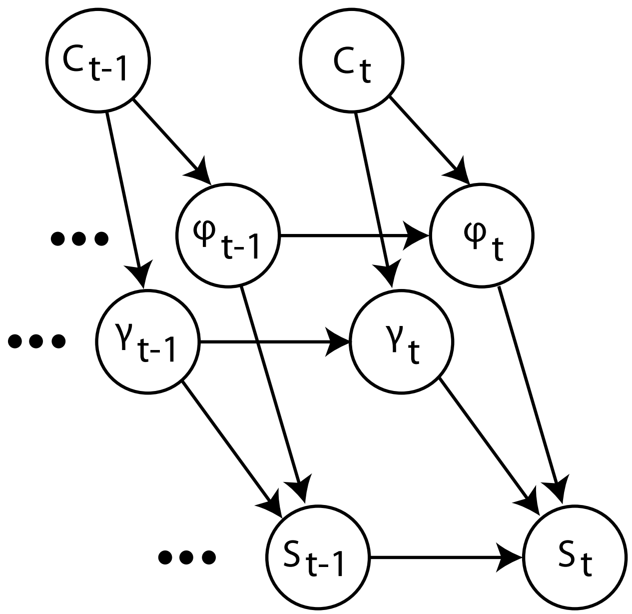
$C \in \{0, 1\}$ changepoint

γ : stimulus probability

$S \in \{x, y\}$

DBM2: Dynamic Belief Mixture Model (Wilder, Jones, & Mozer, 2009)

Current stimulus/response influenced by both 1st and 2nd order sequence properties (base and repetition rates)



$C \in \{0, 1\}$ changepoint

φ : stimulus probability
γ : repetition probability

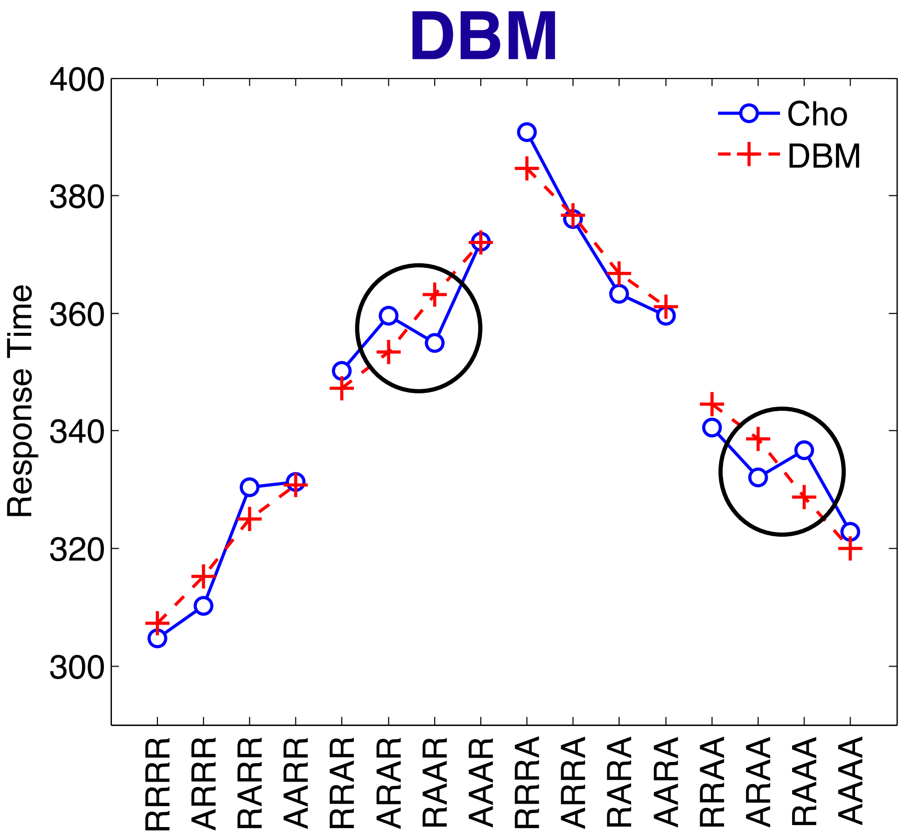
$S \in \{x, y\}$

$$P(S_t = X|\phi_t, \gamma_t, S_{t-1} = X) = w\phi_t + (1-w)\gamma_t$$
$$P(S_t = X|\phi_t, \gamma_t, S_{t-1} = Y) = w\phi_t + (1-w)(1-\gamma_t)$$

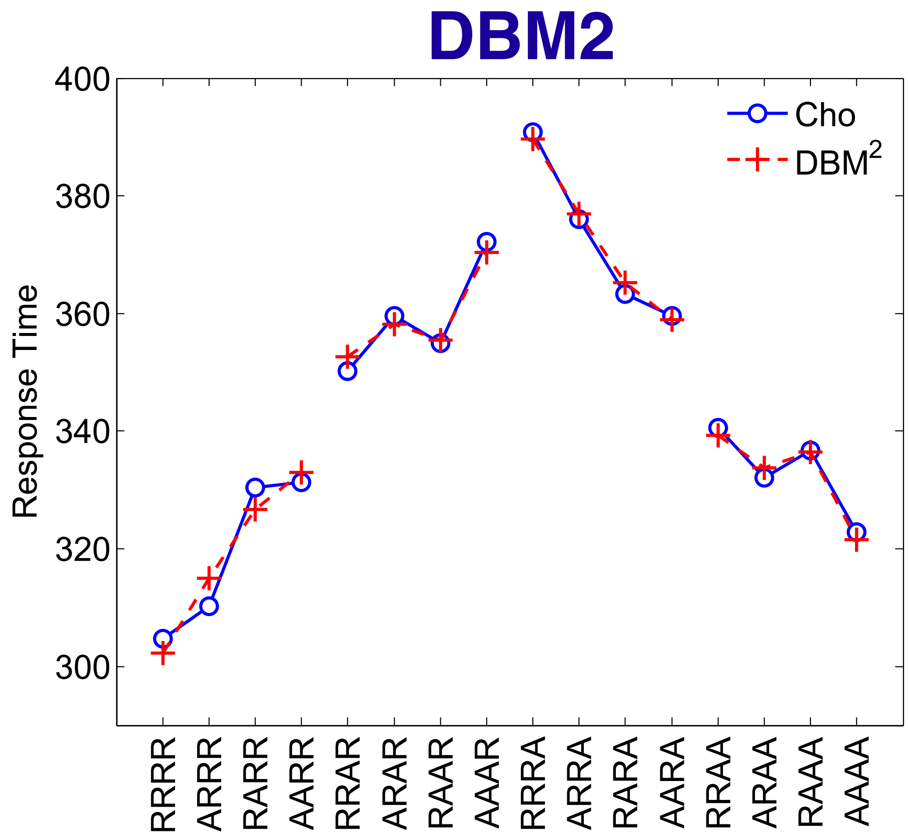
Two free parameters: changepoint prior, *w*

Reset distribution is unbiased Beta(1,1)

Fit to Cho et al. (2002)

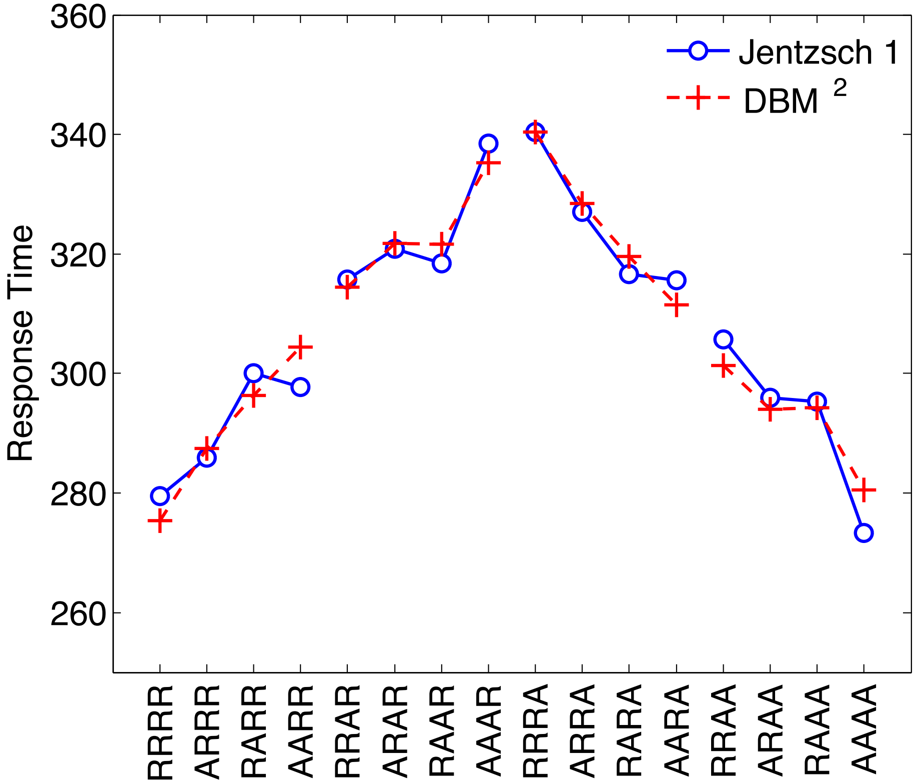


95.8% variance explained
3 free parameters
simple architecture

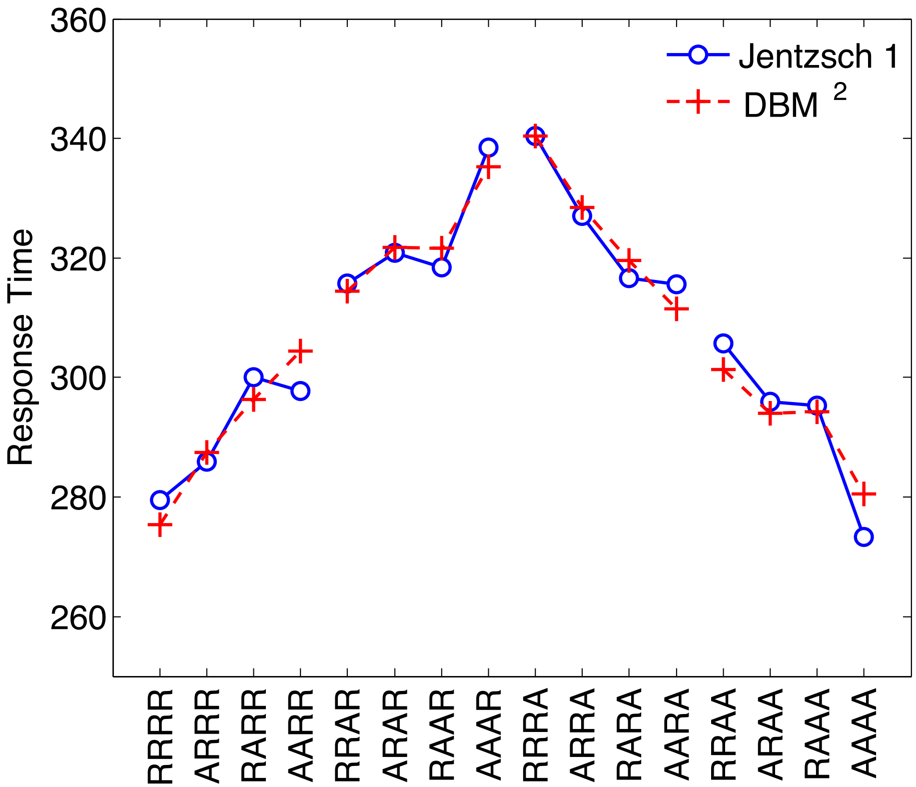


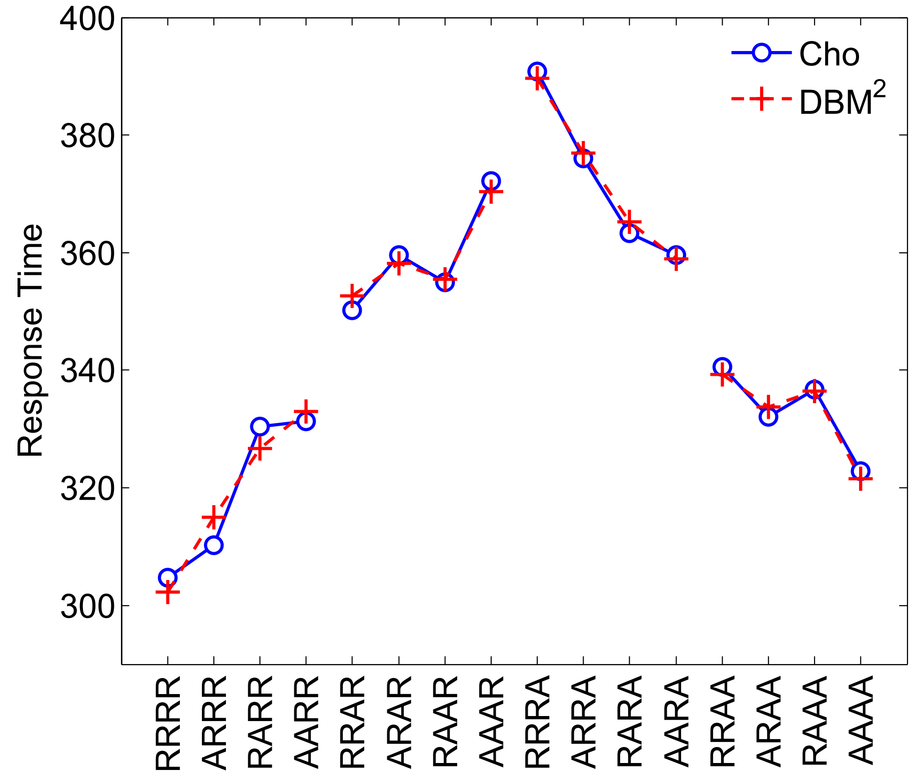
99.2% variance explained
2 free parameters
relatively complex architecture

Jentzsch and Sommer (2002)



Jentzsch and Sommer (2002)





Maloney, Dal Martello, Sahm, and Spillman (2005)

Sequential dependencies in perception of apparent motion

Maloney, Dal Martello, Sahm, and Spillman (2005)

Sequential dependencies in perception of apparent motion

Maloney, Dal Martello, Sahm, and Spillman (2005)

Sequential dependencies in perception of apparent motion

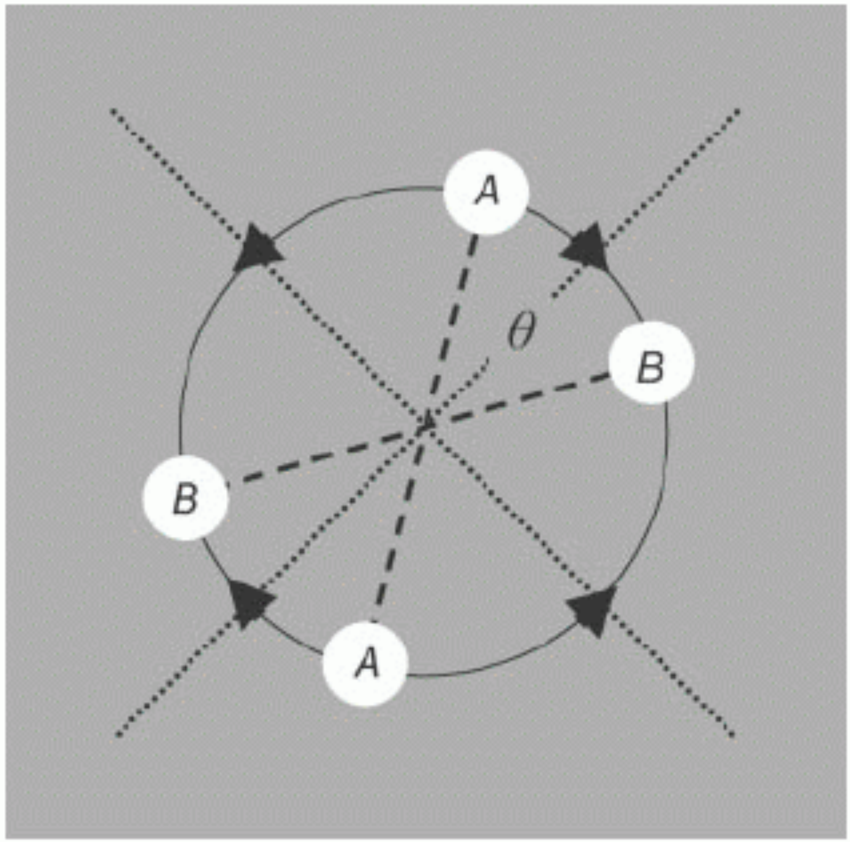
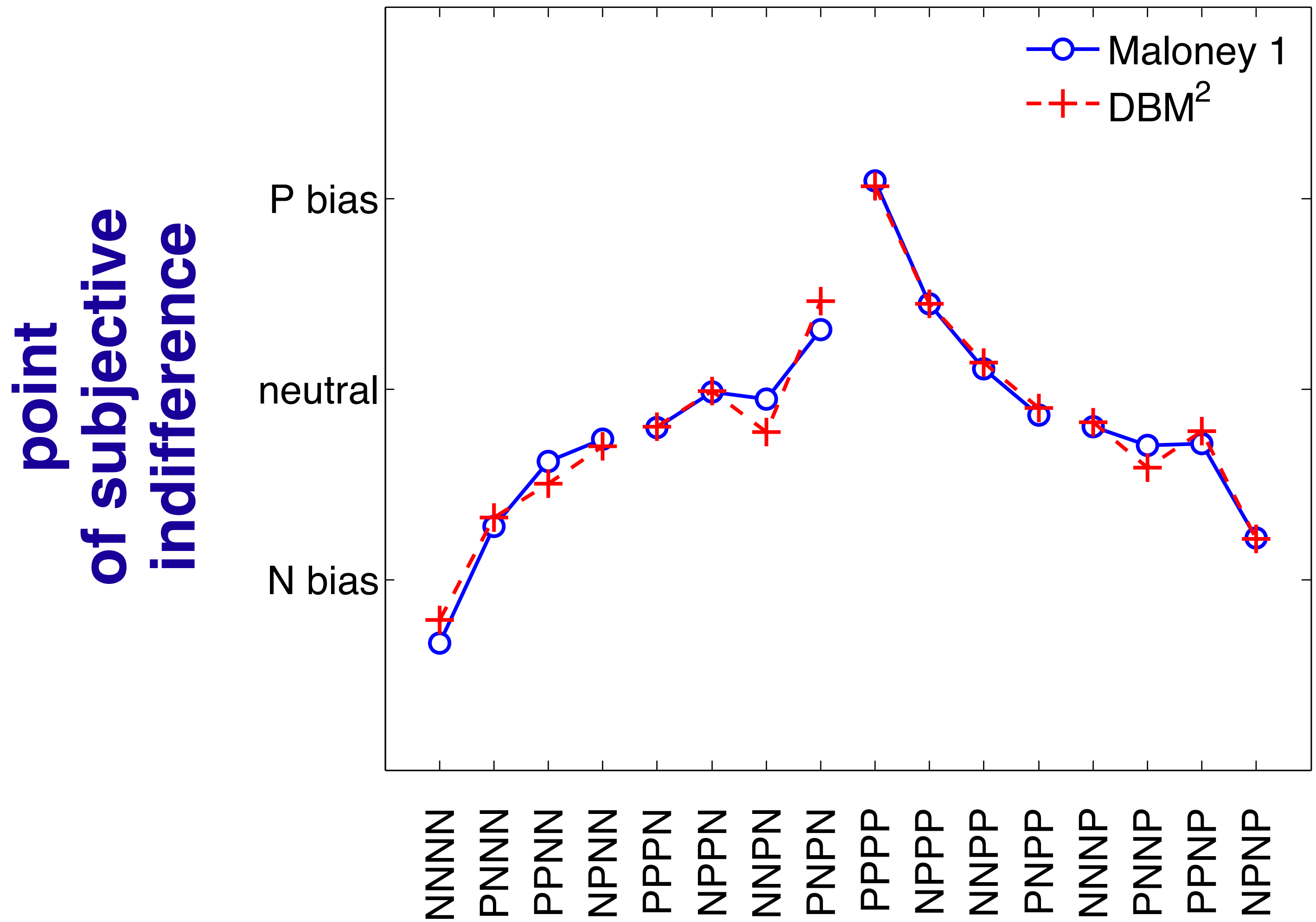


Fig. 1. A motion quartet. The pair of disks marked A appears for 250 ms and then disappears. After a short delay (250 ms), the pair marked B appears for 250 ms. The observer sees apparent rotational motion that carries the first pair of dots into the second. The angle θ between the two diameters affects the probability that the direction of apparent motion is clockwise or counter-clockwise. For many observers, the movement is roughly equally likely to be clockwise as counterclockwise when $\theta = 90°$.

Maloney et al. (2005), Experiment 1



Where Are We At?

DBM2 more complex than DBM

Both models have 3 free parameters

DBM2 fits data a bit better

Table 1: A comparison between the % of data variance explained by DBM and DBM2.

	Cho	Jentzsch 1	Maloney 1
DBM	95.8	95.5	96.1
DBM2	99.2	96.5	97.7

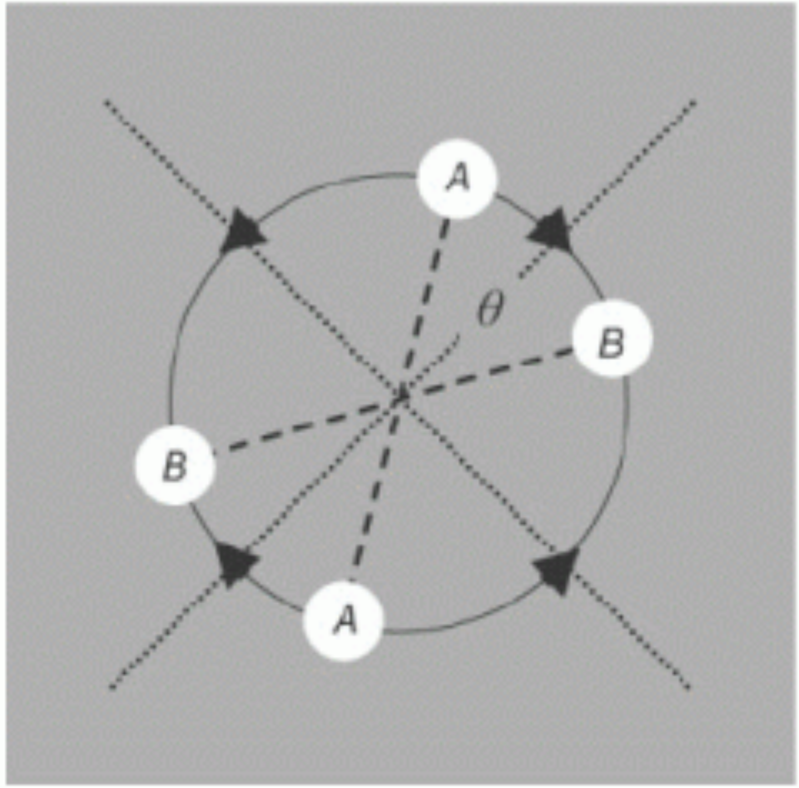
Further Claim of DBM2

First and second order predictions are prediction are distinct, and might correspond to distinct brain mechanisms.

Hypothesis

- Base rates (first order) are computed in response system and based on response properties.
- Repetition rates (second order) are computed in perceptual system and based on stimulus properties.

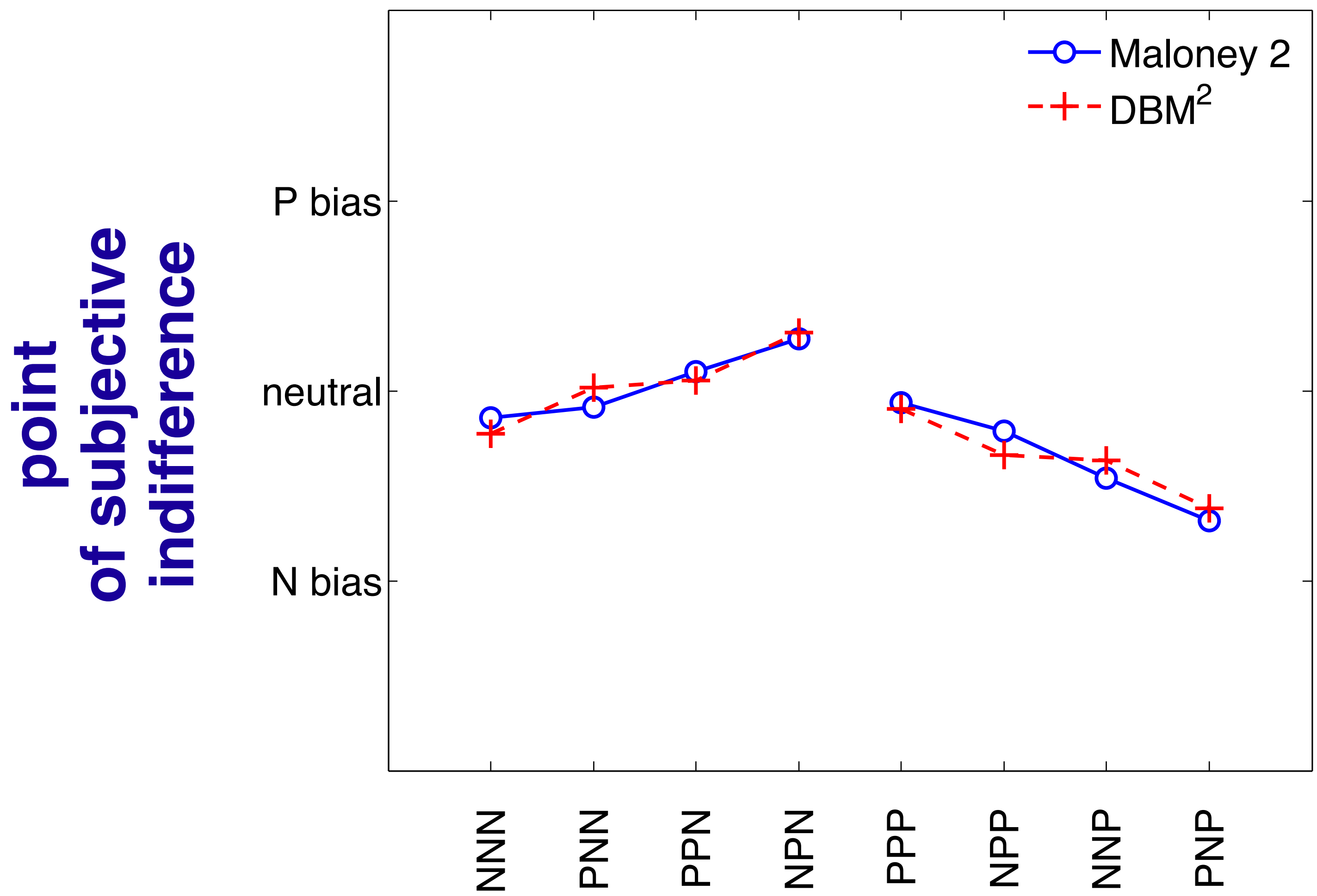
Maloney et al. (2005), Experiment 2



Participants make responses only every 4 trials.

If response mechanisms aren't operating, then according to our hypothesis, base rates will not influence sequential dependencies.

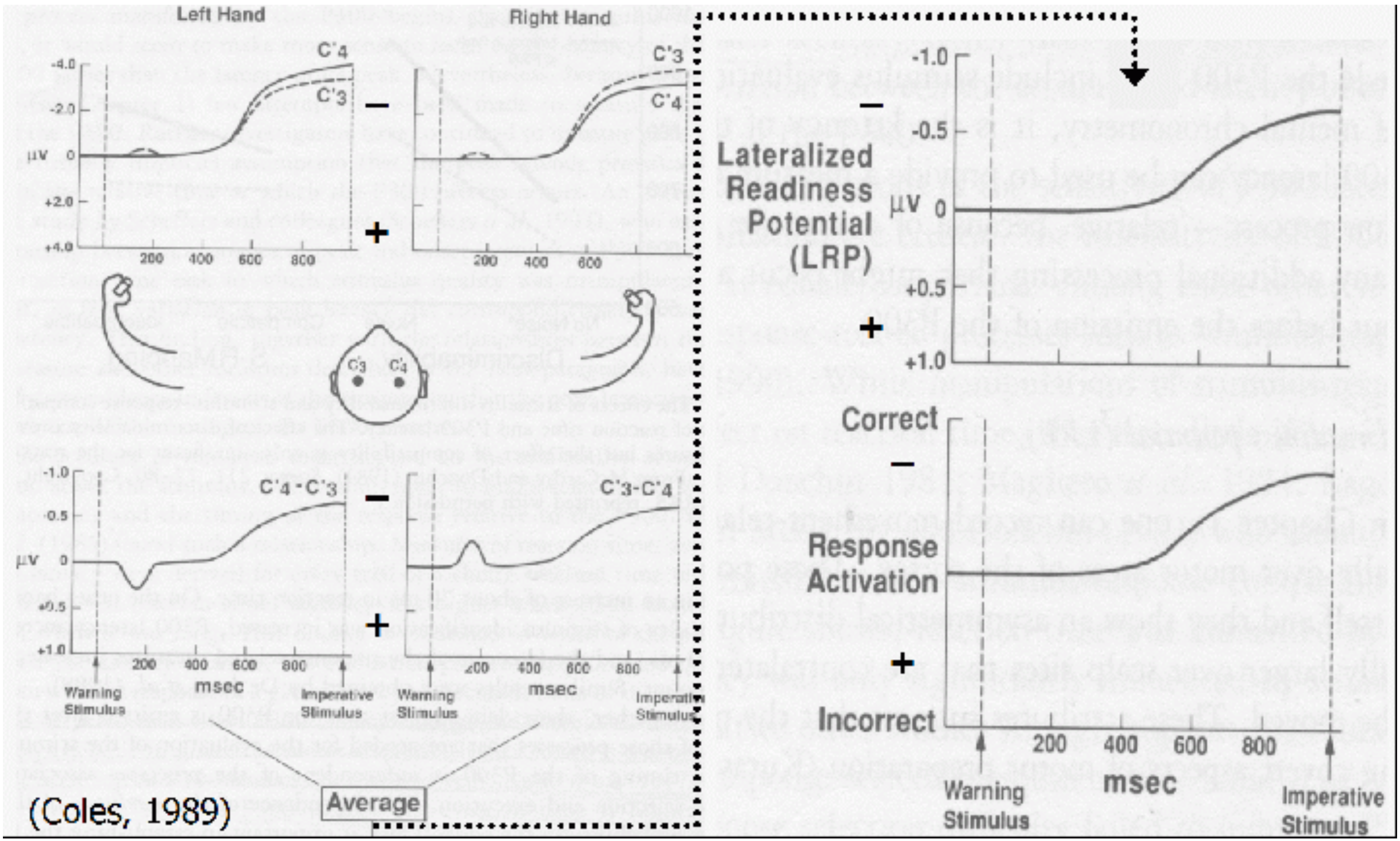
Maloney et al. (2005), Experiment 2



Jentzsch and Sommer (2002)

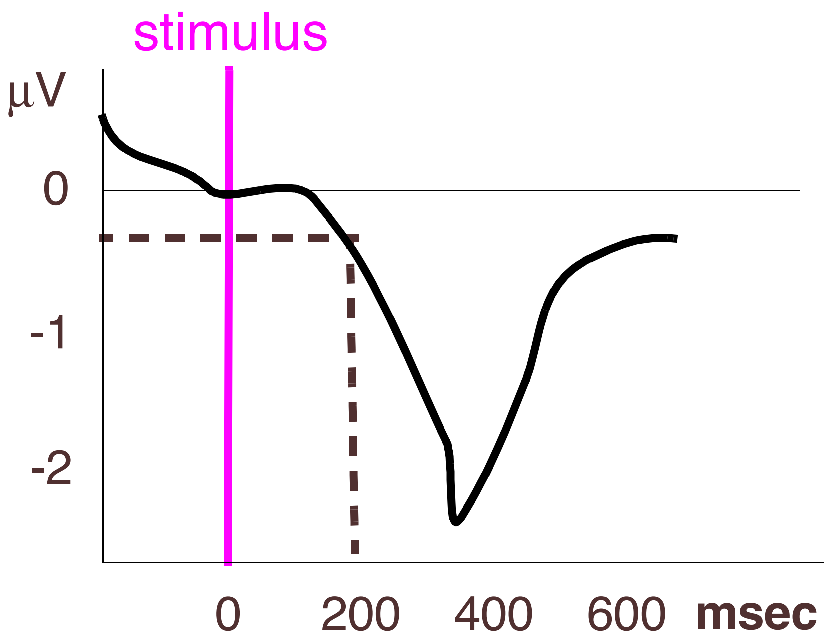
Measured lateralized readiness potential (LRP)

ERP measure of ipsilateral - contralateral motor activity

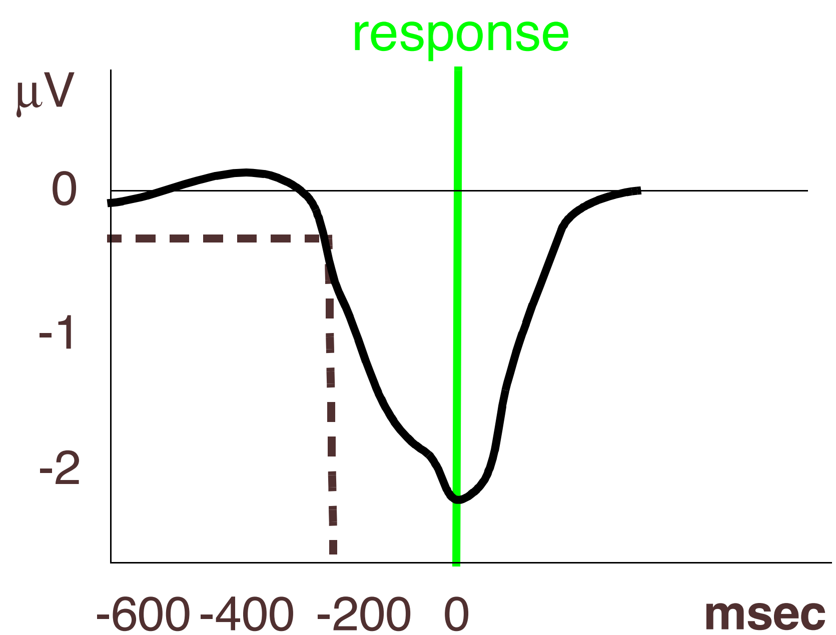


Two LRP measures

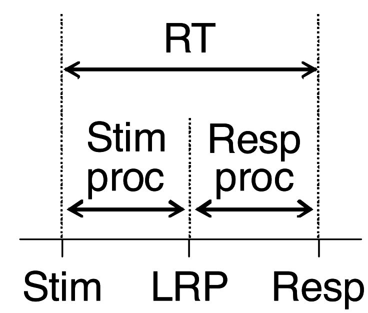
S-LRP: time from stimulus presentation to onset of LRP



LRP-R: time from onset of LRP to initiation of response



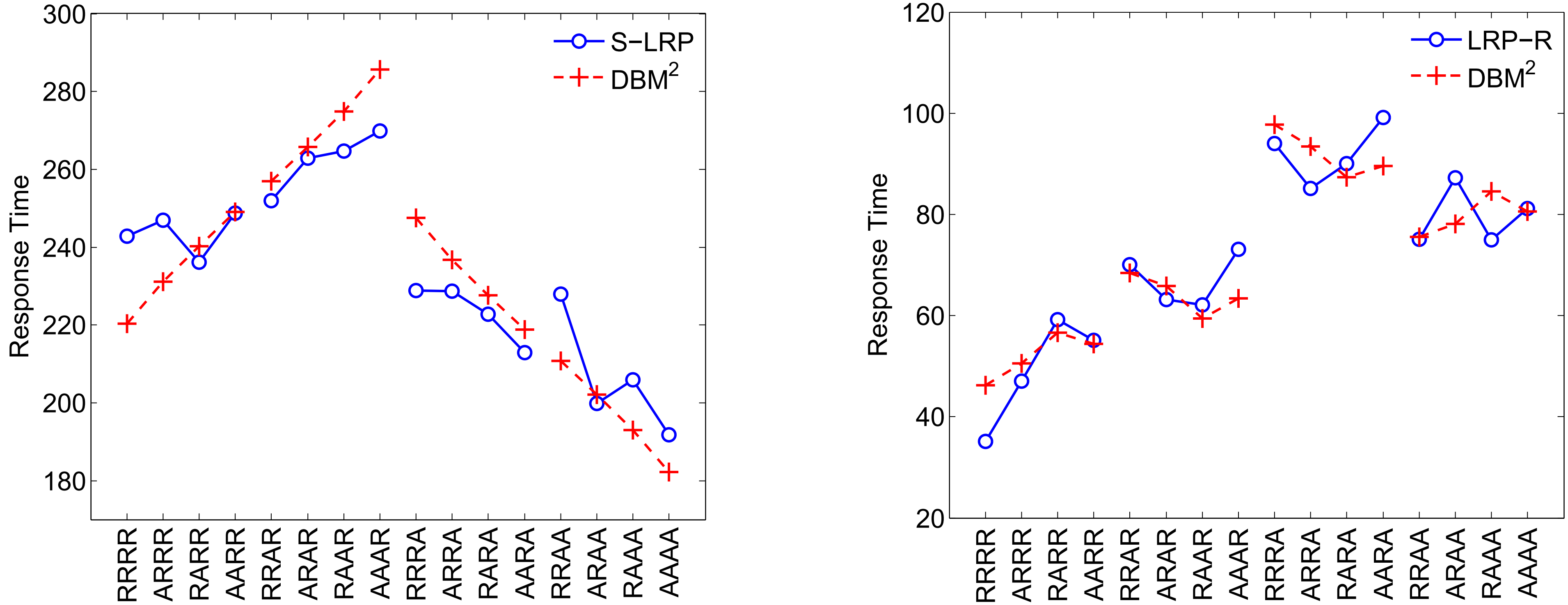
S-LRP and LRP-R roughly breaks total RT into stimulus and response processing components



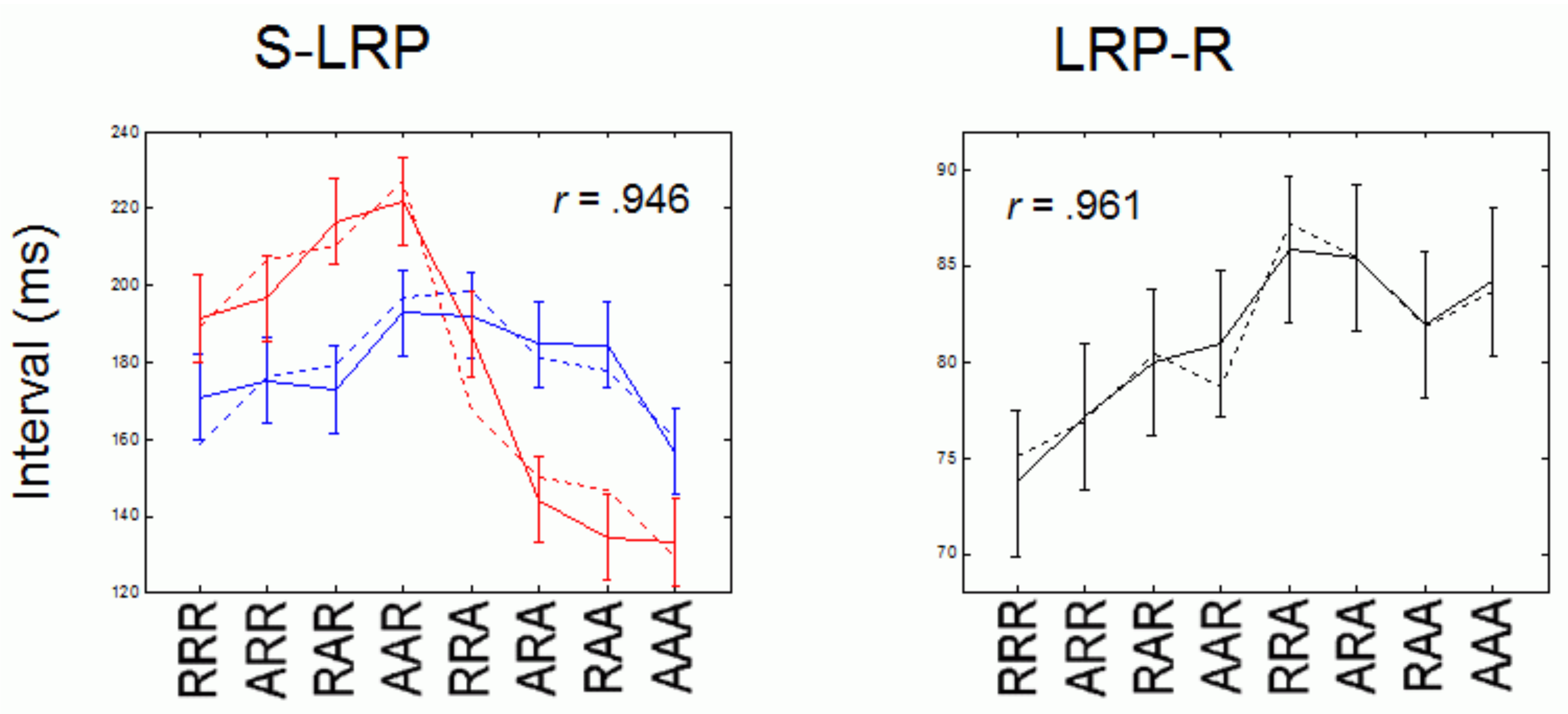
Jentzsch and Sommer (2002)

Fits of stimulus and response processing

model using same parameters as overall RT fits



Jones, Curran, Mozer, & Wilder (2010)



Sequential Effects in Motor Adaptation

Matt Wilder
Department of Computer Science

Alaa Ahmed
Department of Integrative Physiology

Michael Mozer
Department of Computer Science

Matt Jones
Department of Psychology

Reaching Task

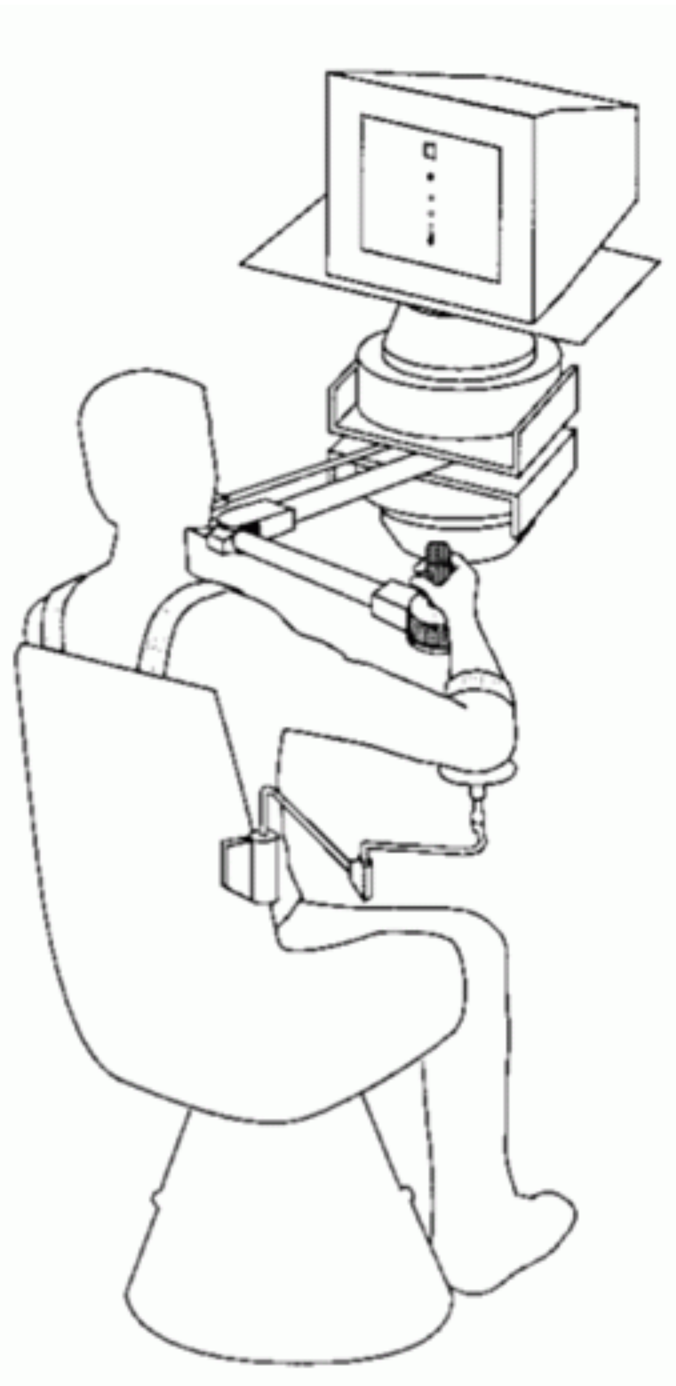
Move robotic arm (manipulandum) straight toward target — 15 cm — and return to starting position

Perpendicular perturbing force applied on each trial, either to the left or the right

Force increases with position for first 5 cm, then constant for last 10

No force on return

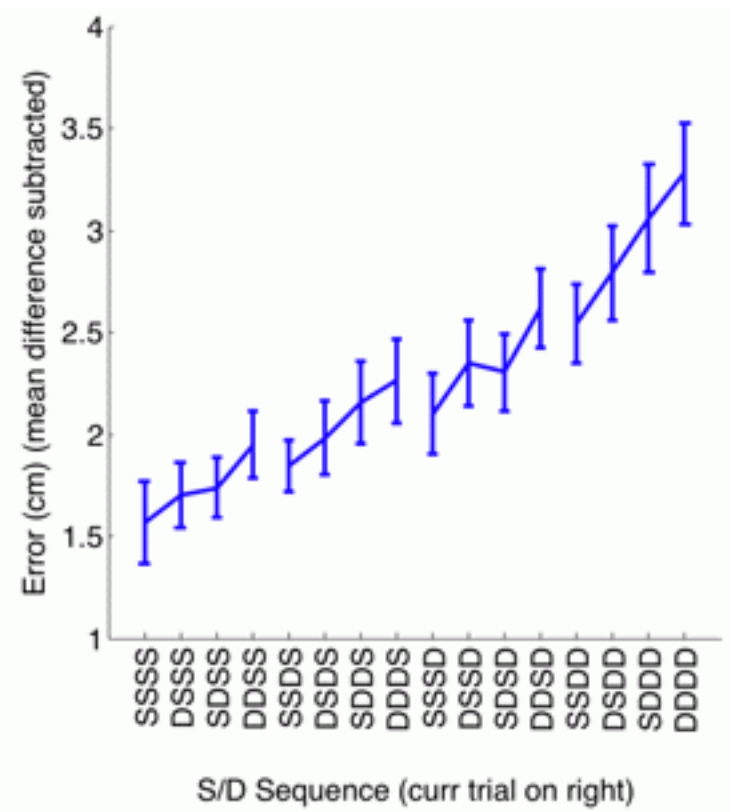
Measure error: maximum deviation from straight path

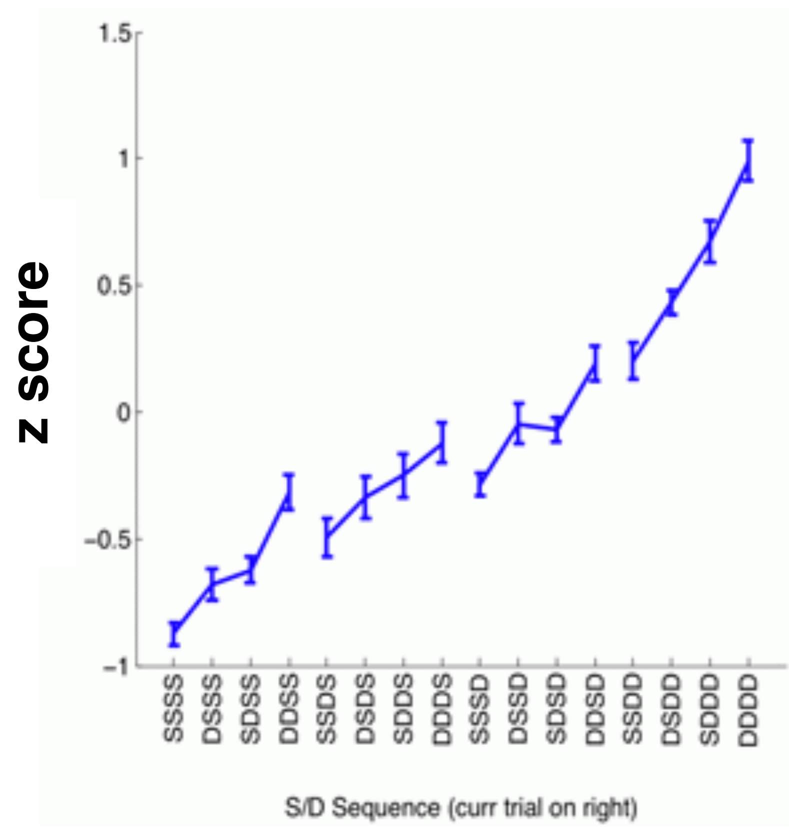


Sequential Effects in Reaching Task

Eight subjects

First-order priming, going back at least four trials





Do Sequential Effects Go Back Further?

For individual subjects, compute:

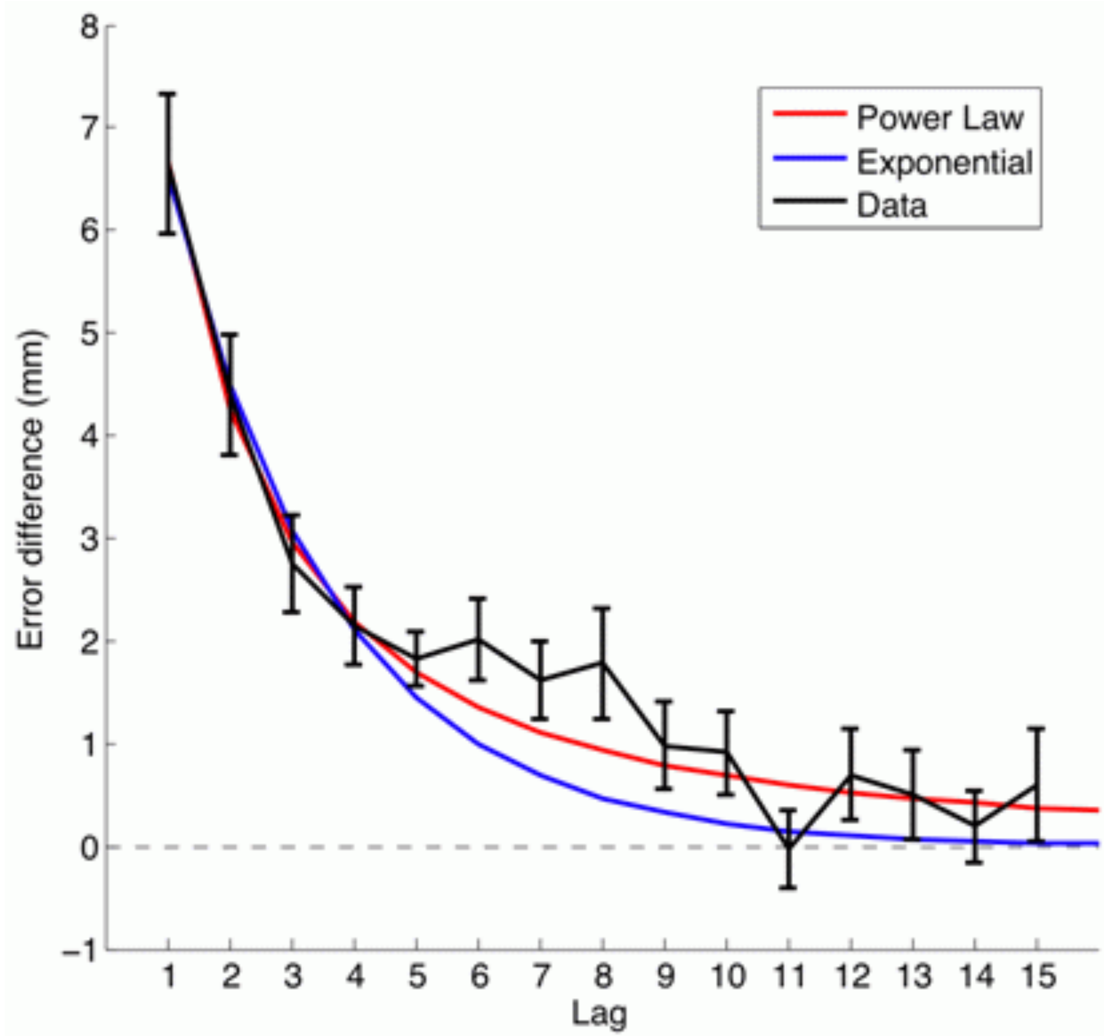
$$\Omega_D(l) = \{t : S_t \neq S_{t-l}\}$$

$$\Omega_S(l) = \{t : S_t = S_{t-l}\}$$

$$e_D(l) = \frac{1}{|\Omega_D(l)|} \sum_{t \in \Omega_D(l)} e_t$$

$$e_S(l) = \frac{1}{|\Omega_S(l)|} \sum_{t \in \Omega_S(l)} e_t$$

$$lag(l) = e_D(l) - e_S(l)$$



Curves are fit based on lags 1-5

Sequential Effects in Driving

Anup Doshi
Cuong Tran
Mohan Trivedi
Department of Electrical Engineering
UCSD

Matt Wilder
Michael Mozer
Department of Computer Science
University of Colorado

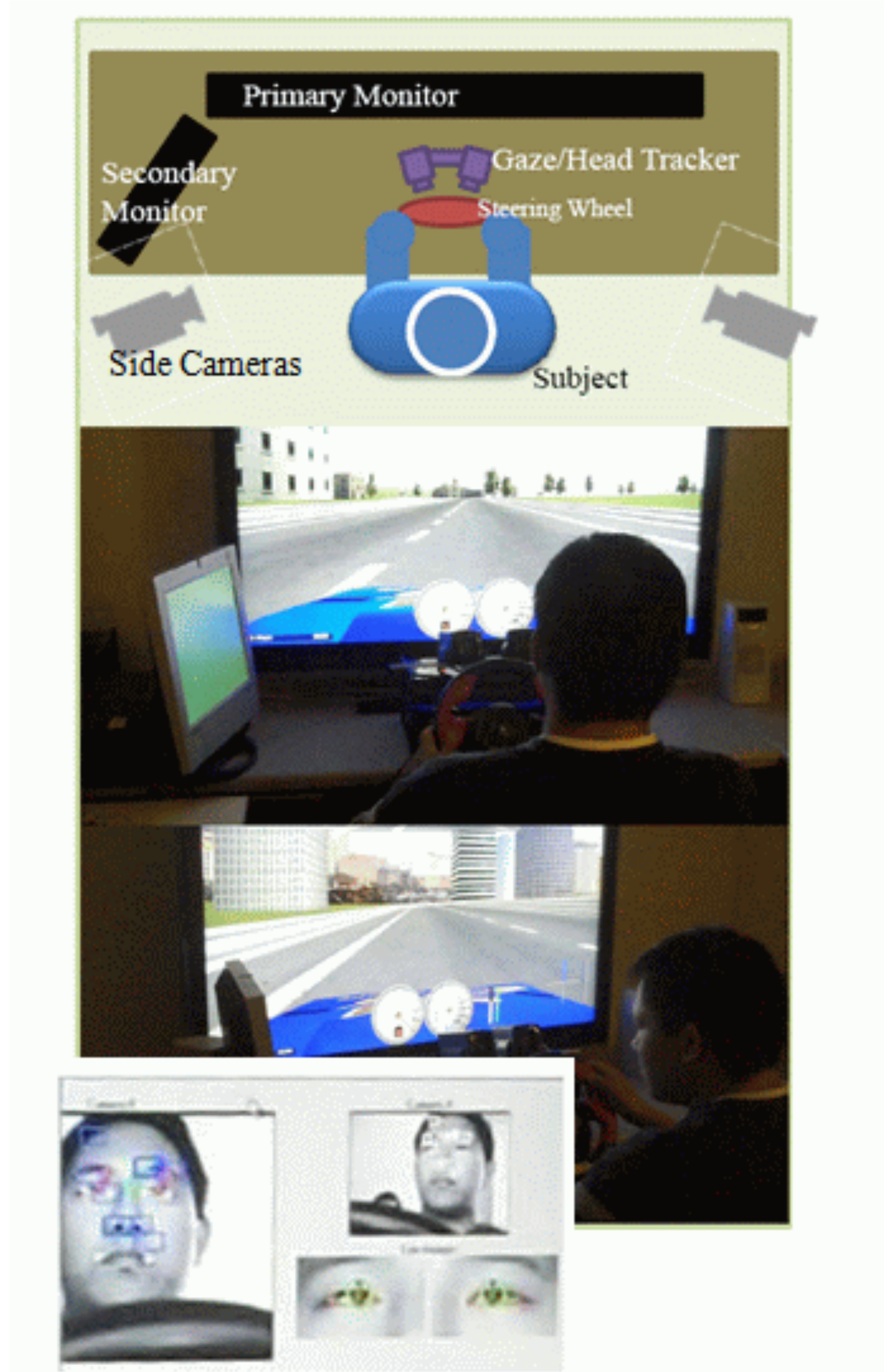
Laboratory for Intelligent & Safe Automobiles (LISA)

Xbox-like driving simulation with realistic physics

Full size steering wheel

Brake, acceleration pedals

Cameras focused on driver’s head and eyes, hands, feet



Task

Drive in simulator

- **Twisty road, constant turns**
- **Driver instructed to stay in middle lane of 3-lane highway**
- **Buildings and objects in the scene**

Occasional cues to brake or accelerate

simulate stop-and-go traffic

guide car: brake lights or kicking up dust

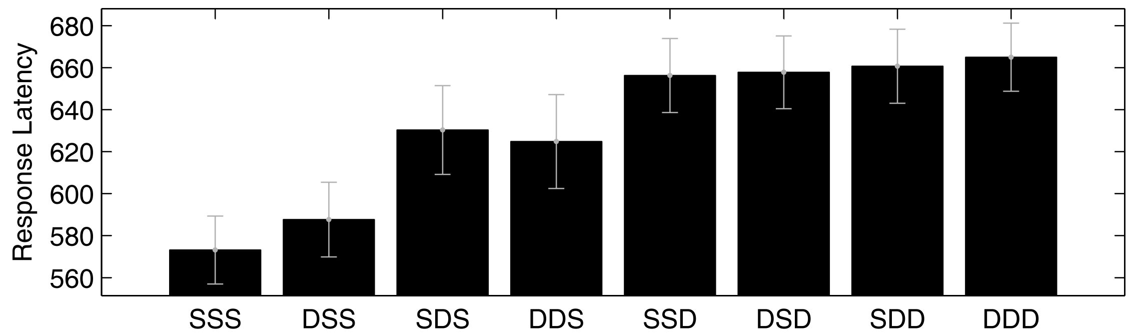
traffic light in windshield

Constant velocity travel when no pedal press

Decomposing The Total Response Time

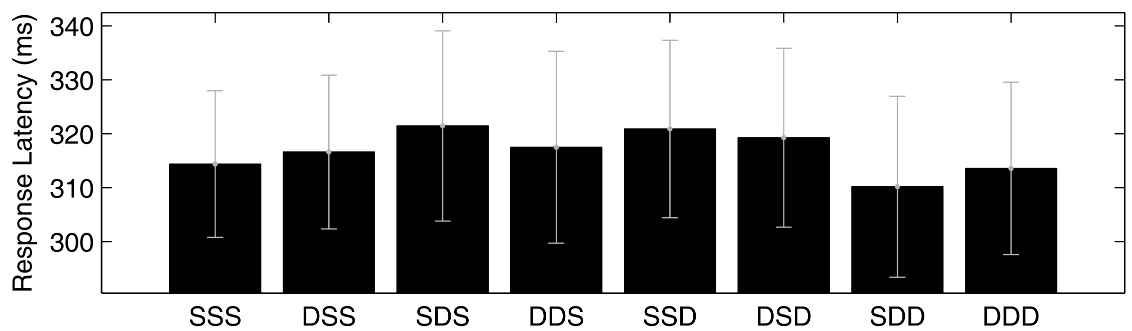
Cameras monitored foot, so we can decompose RT into

total response time



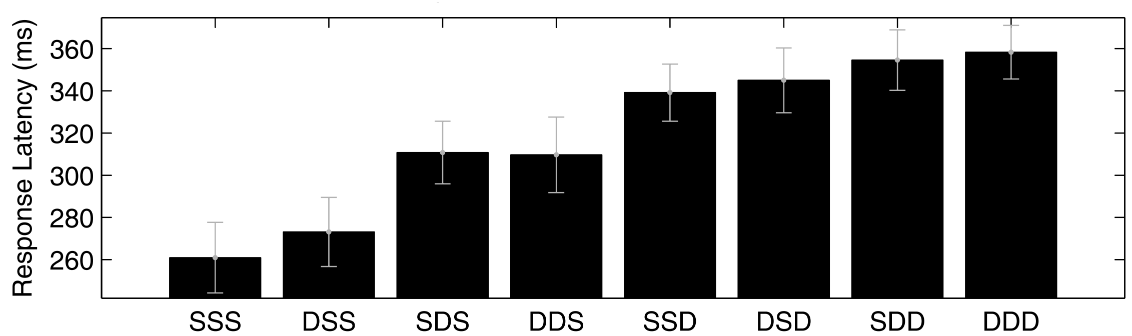
=

time from stimulus onset to foot movement



+

time to move foot to pedal



Summary

Systematic discrepancies in DBM -> elaborated generative model (DBM2)

Summary

Systematic discrepancies in DBM ->
elaborated generative model (DBM2)

Conceptual organization of DBM2 ->
psychological architecture in which first and second order statistics separately represented

Summary

Systematic discrepancies in DBM ->
elaborated generative model (DBM2)

Conceptual organization of DBM2 ->
psychological architecture in which first and second order statistics separately represented

Further fits to data (LRP, intermittent responses) ->
tracked statistics are differentially tied to stimuli and responses

Summary

Systematic discrepancies in DBM -> elaborated generative model (DBM2)

Conceptual organization of DBM2 -> psychological architecture in which first and second order statistics separately represented

Further fits to data (LRP, intermittent responses) -> tracked statistics are differentially tied to stimuli and responses

- First order statistics (a.k.a. baserates, marginal probabilities) of response sequence
- Second order statistics (a.k.a. repetition rates, transition probabilities) of stimulus sequence

Summary

Systematic discrepancies in DBM ->
elaborated generative model (DBM2)

Conceptual organization of DBM2 ->
psychological architecture in which first and second order statistics separately represented

Further fits to data (LRP, intermittent responses) ->
tracked statistics are differentially tied to stimuli and responses

- First order statistics (a.k.a. baserates, marginal probabilities) of response sequence
- Second order statistics (a.k.a. repetition rates, transition probabilities) of stimulus sequence

Sequential effects in other domains

- reaching with perturbations
- driving