Adaptive Design Optimization for Model Discrimination

Jay Myung
Department of Psychology
Ohio State University

Joint work with
Daniel Cavagnaro and Mark Pitt

Annual Interdisciplinary Summer Conference (Bend, OR: July 2010)

Experimentation

- Experimentation is fundamental to the advancement of science, whether one is interested in studying the neuronal basis of a sensory process in cognitive neuroscience or studying attentional biases in working memory task.

Model discrimination

☐ The goal of experimentation is often to discriminate between competing models of a cognitive process under investigation.

Model selection methods:

- Akaike Information Criterion (AIC)
- Minimum Description Length (MDL)
- Bayes Factor

Applied AFTER data have already been collected

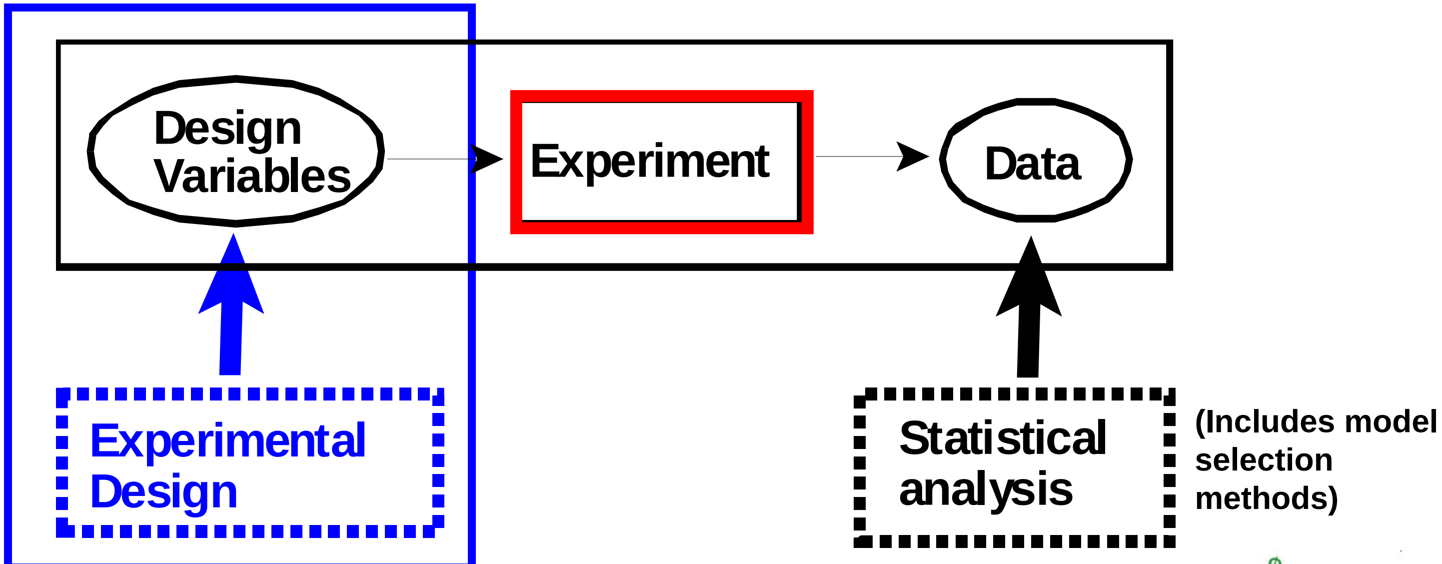
Experimental design

☐ The **design** of an experiment includes **choices made BEFORE data are collected**, and can affect the potential of that experiment to produce discriminating data.

Design variables: number of treatments and stimulus/factor levels and values

Optimal experimental design

- ☐ An **optimal design** maximizes the informativeness of the experiment while being cost effective for the experimenter (Atkinson & Donev, 1992; Myung & Pitt, 2009, *Psy Rev*).



Design Variables

Experiment

Data

Experimental Design

Statistical analysis

(Includes model selection methods)

Adaptive optimal design

- Adaptively designed experiments

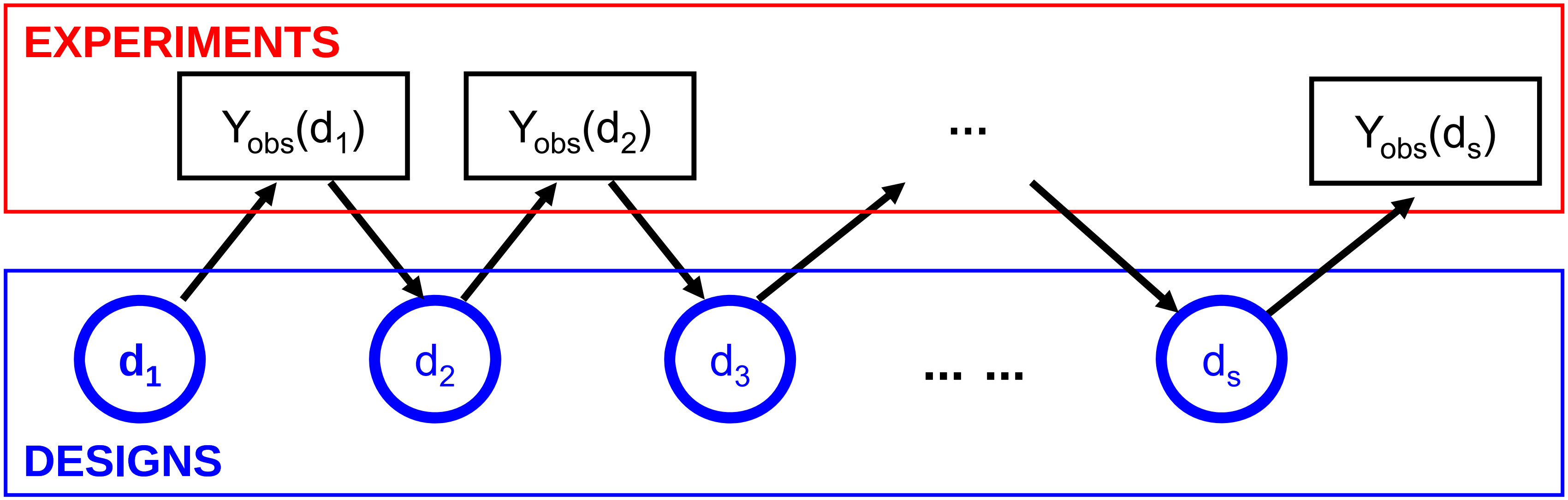
Adaptive optimal design

- Adaptively designed experiments
  - Conduct the full experiment as a sequence of mini-experiments
  - Improve the design of the next mini-experiment using knowledge gained from the previous mini-experiments

Cavagnaro, Myung, Pitt & Kujala (2010) *Neural Computation.*

Adaptive optimal design

- ☐ Sequential design framework



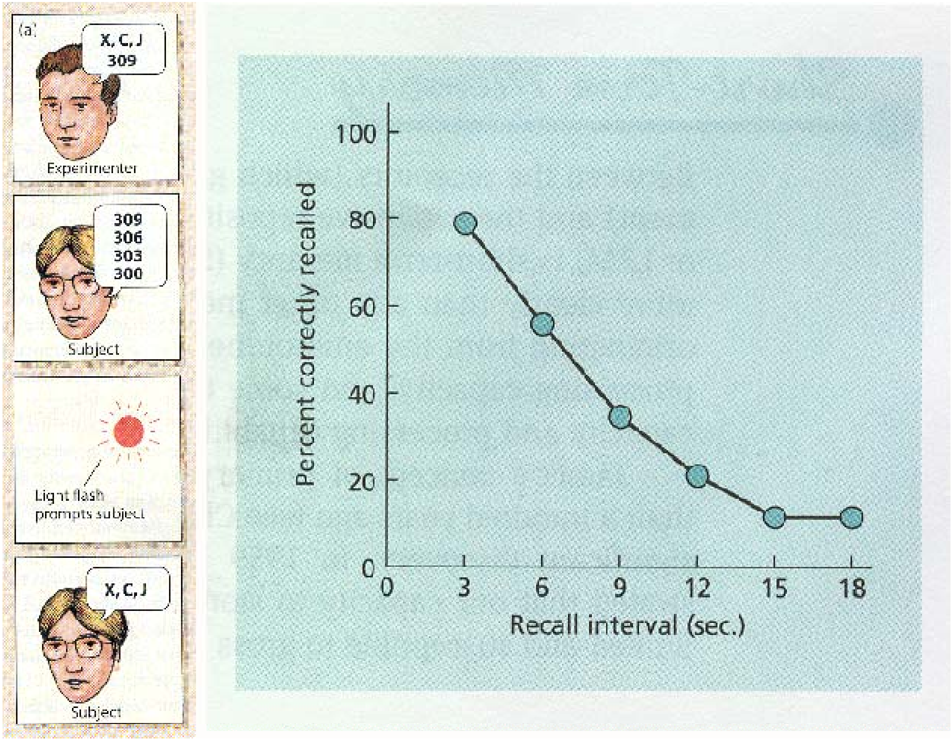
Adapt the design of the next experiment based on the results of preceding experiments

Discriminating
Models of Memory Retention

Designing a Retention Experiment:

What Time Intervals Should be Employed?

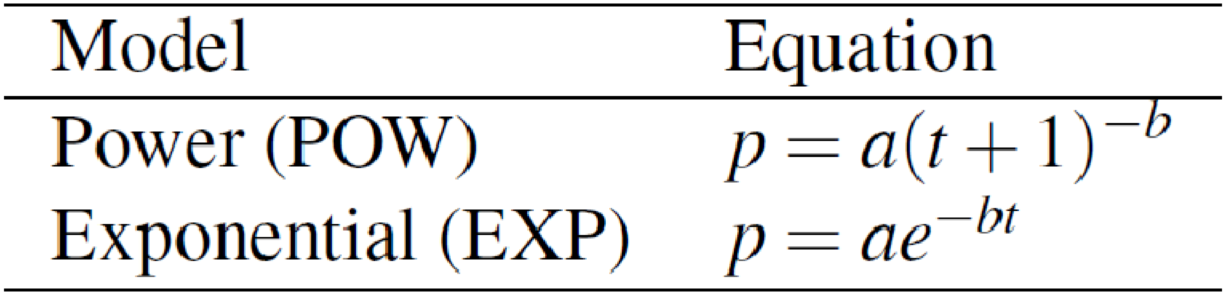
- ☐ Retention: the rate of retrieval failure over time



Designing a Retention Experiment:

What Time Intervals Should be Employed?

- Two models of retention

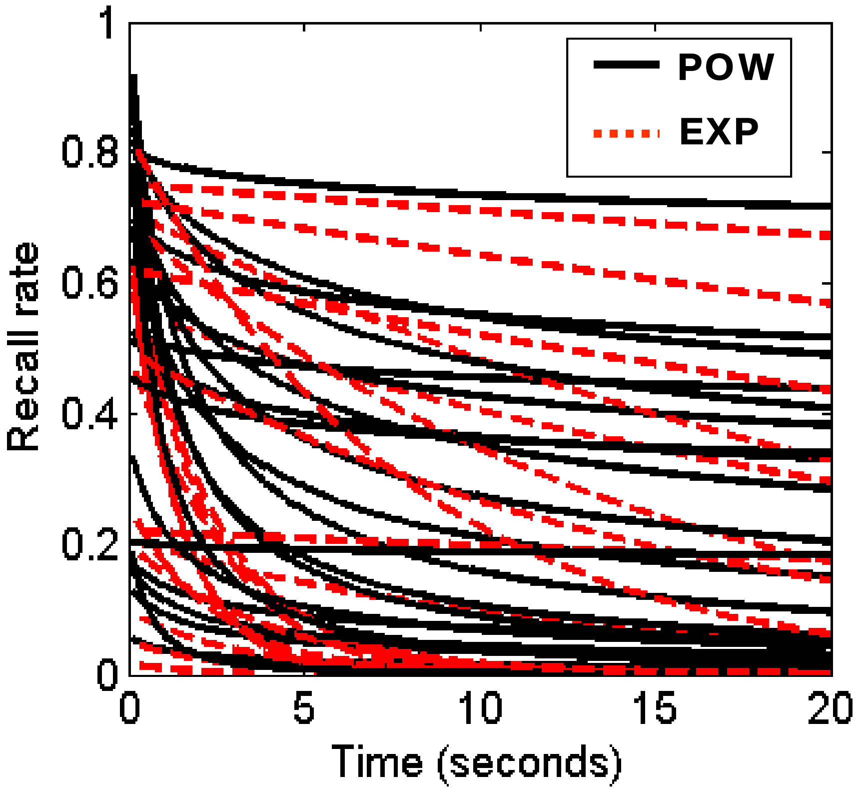


Model	Equation
Power (POW)	$p = a(t+1)^{-b}$
Exponential (EXP)	$p = ae^{-bt}$

Designing a Retention Experiment:

What Time Intervals Should be Employed?

Model mimicry between POW and EXP



Designing a Retention Experiment:

What Time Intervals Should be Employed?

Typical retention experiment:

Study phase:
Target list memorized:
{RXJ, WPW, HJM,...}

→

Retention interval:
Memory decays

→

Recall phase:
Memory assessed

Designing a Retention Experiment:

What Time Intervals Should be Employed?

Typical retention experiment:

Study phase:
Target list memorized:
{RXJ, WPW, HJM,…}

→

Retention interval:
Memory decays

→

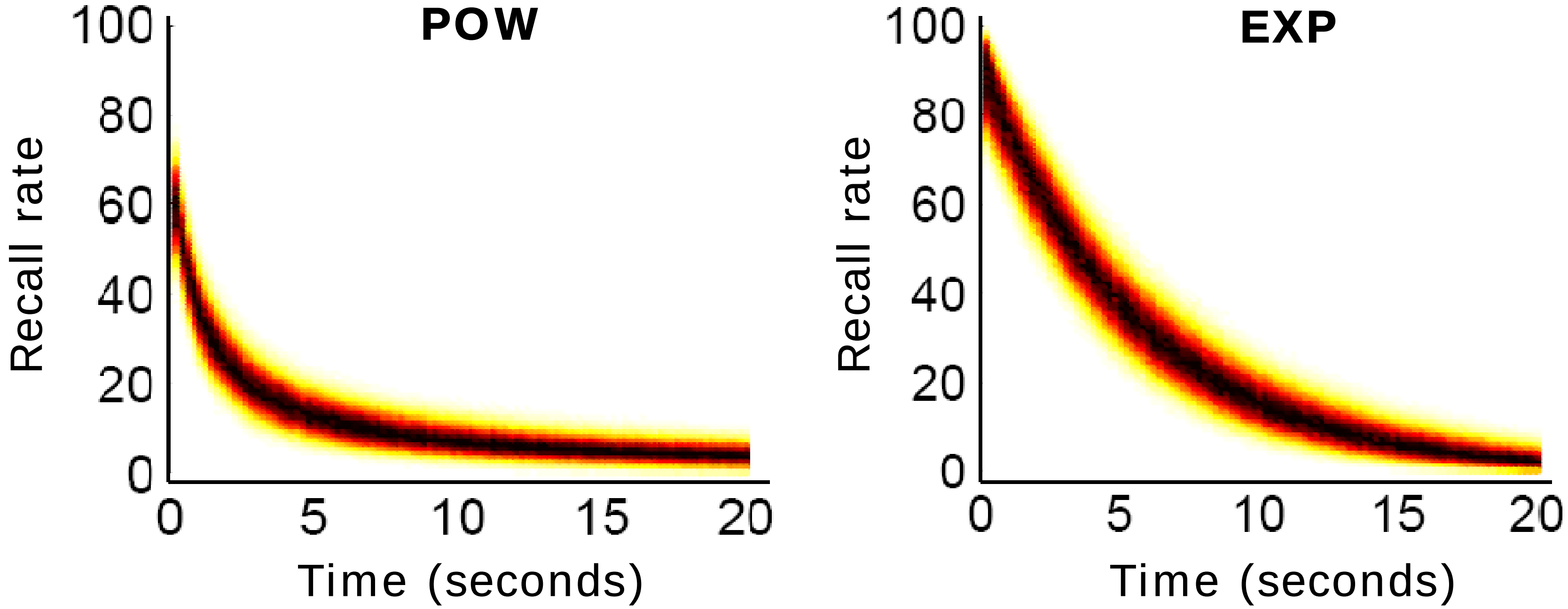
Recall phase:
Memory assessed

Design variable:
How long to wait before
assessing memory in each trial?

Designing a Retention Experiment:

What Time Intervals Should be Employed?

Model predictions for a narrow range of parameters (100 Bernoulli trials)

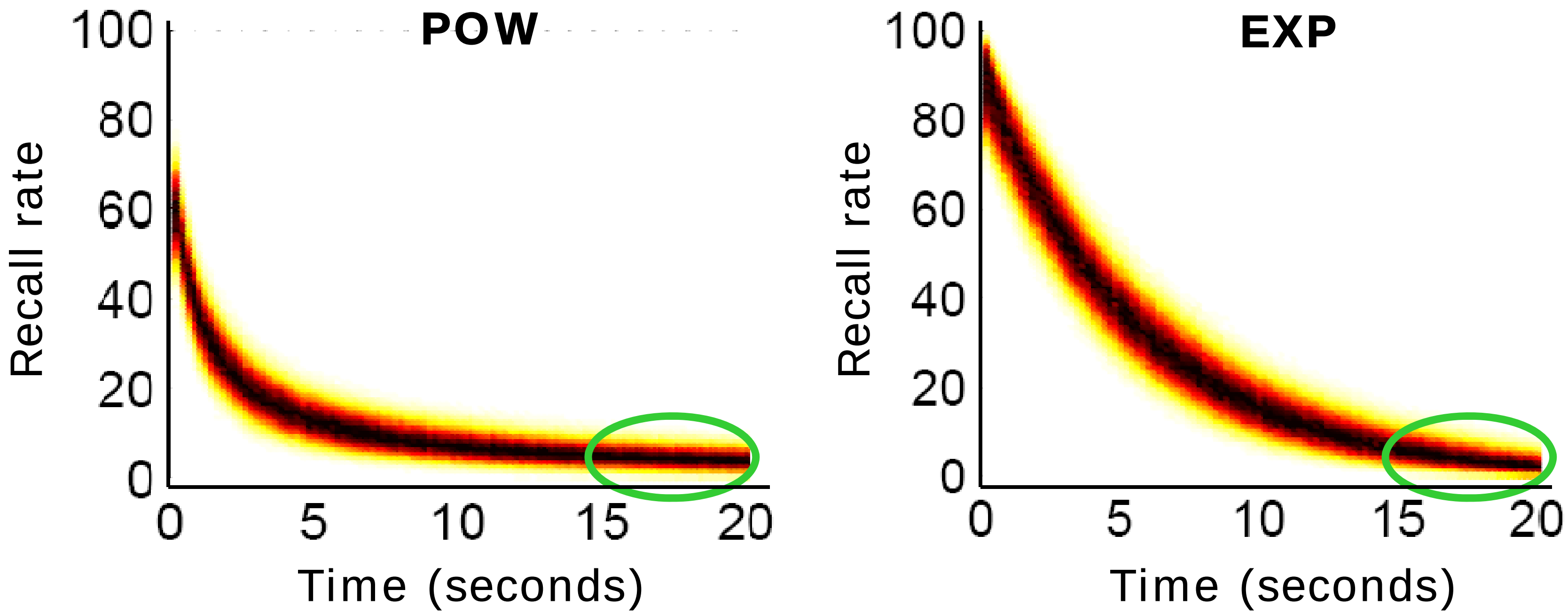


Designing a Retention Experiment:

What Time Intervals Should be Employed?

- ☐ A good design choice can aid in discrimination

15-20 seconds: Bad designs!

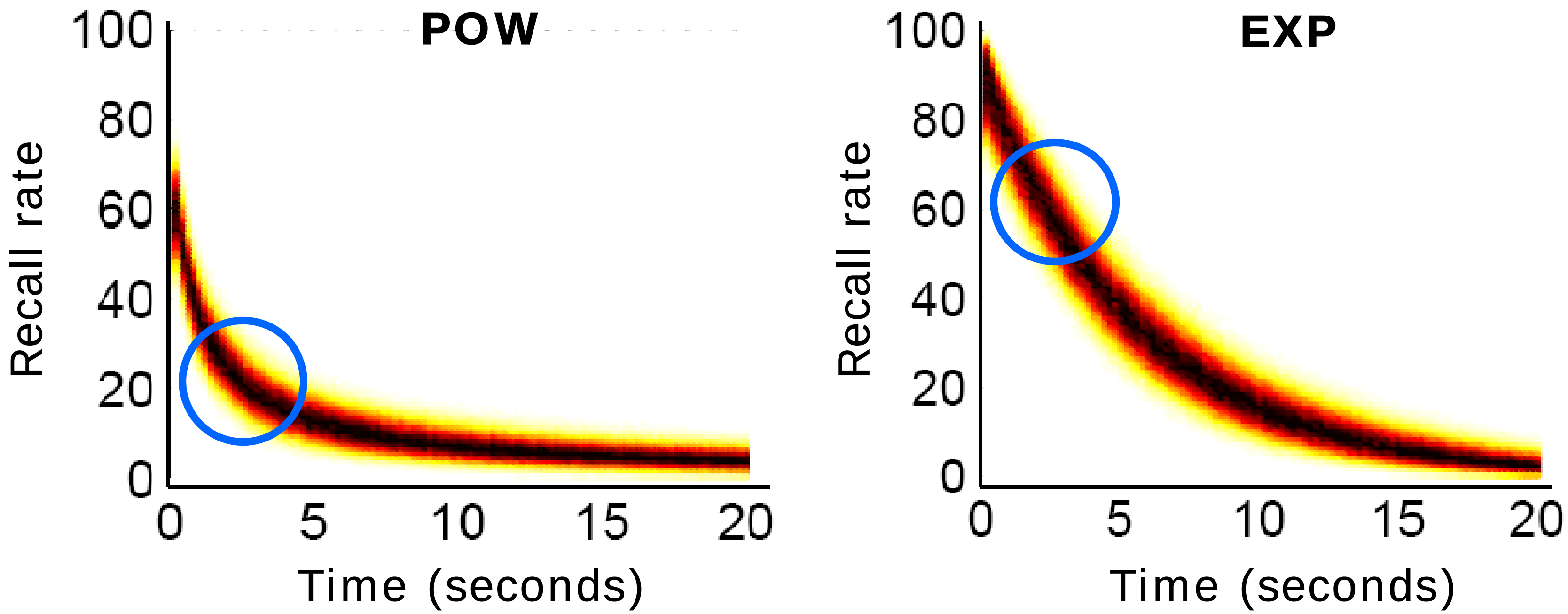


Designing a Retention Experiment:

What Time Intervals Should be Employed?

- ☐ A good design choice can aid in discrimination

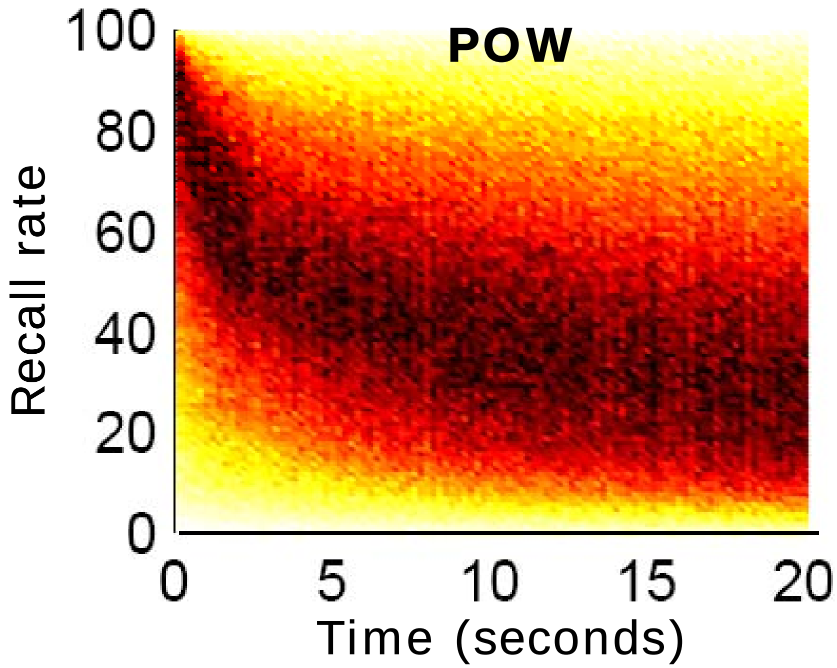
2-4 seconds: Good designs!



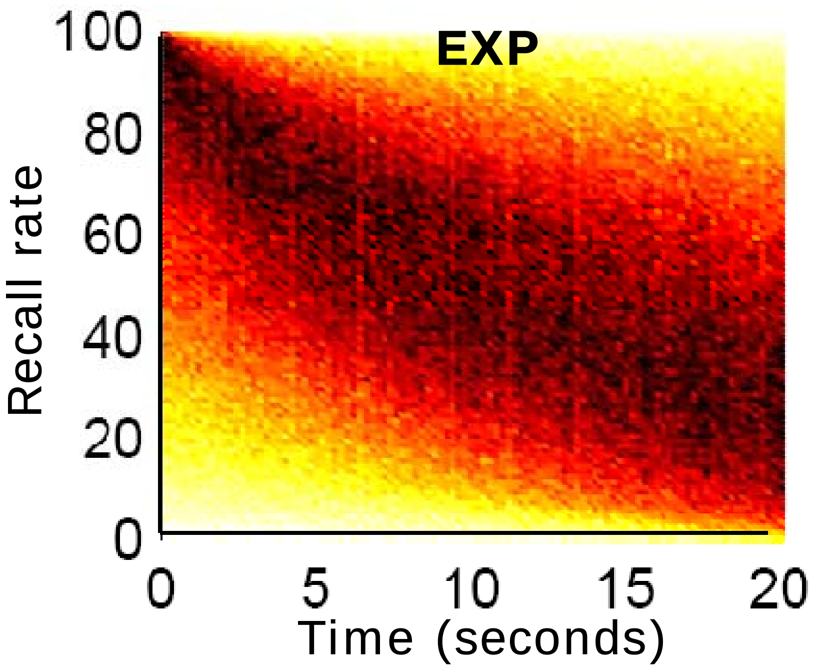
Designing a Retention Experiment:

What Time Intervals Should be Employed?

- More realistic situations require a more principled approach to finding optimal designs.



a~Beta(2,1)
b~Beta(1,4)



a~Beta(2,1)
b~Beta(1,80)

Finding Optimal Designs

Design Optimization

- A principled approach from Bayesian decision theory

Design Optimization

- A principled approach from Bayesian decision theory
  - Treat each possible design as a gamble whose payoff is determined by the outcome of an experiment carried out with that design (Chaloner & Verdinelli, 1995).

Design Optimization

- A principled approach from Bayesian decision theory
  - Treat each possible design as a gamble whose payoff is determined by the outcome of an experiment carried out with that design (Chaloner & Verdinelli, 1995).
  - Compute an ”expected utility” of each design by taking an expectation over models, parameters, and experiment outcomes.

Design Optimization

- A principled approach from Bayesian decision theory
  - Treat each possible design as a gamble whose payoff is determined by the outcome of an experiment carried out with that design (Chaloner & Verdinelli, 1995).
  - Compute an ”expected utility” of each design by taking an expectation over models, parameters, and experiment outcomes.

$$U(d) = \sum_{m=1}^{K} p(m) \int \int u(d, \theta_m, y)\, p(y|\theta_m, d)\, p(\theta_m)\, dy\, d\theta_m$$

Design Optimization

- ☐ A principled approach from Bayesian decision theory
  - Treat each possible design as a gamble whose payoff is determined by the outcome of an experiment carried out with that design.
  - Compute an ”expected utility” of each design by taking an expectation over models, parameters, and experiment outcomes.

$$U(d) = \sum_{m=1}^{K} p(m) \int \int u(d, \theta_m, y)\, p(y|\theta_m, d)\, p(\theta_m)\, dy\, d\theta_m$$

  - The design with the highest expected utility is then chosen as the optimal design

Optimal design: $d^* = \underset{d}{\operatorname{argmax}}\{U(d)\}$

Design Optimization

- A principled approach from Bayesian decision theory
  - Treat each possible design as a gamble whose payoff is determined by the outcome of an experiment carried out with that design.
  - Compute an ”expected utility” of each design by taking an expectation over models, parameters, and experiment outcomes.

$$U(d) = \sum_{m=1}^{K} p(m) \int\int \underbrace{u(d, \theta_m, y)}_{} \, p(y|\theta_m, d) \, p(\theta_m) \, dy \, d\theta_m$$

Value of a hypothetical experiment with design *d* in the case in which the true model is *m*, with parameters θ_m, and outcome is *y* is observed.

Design Optimization

- A principled approach from Bayesian decision theory
  - Treat each possible design as a gamble whose payoff is determined by the outcome of an experiment carried out with that design.
  - Compute an ”expected utility” of each design by taking an expectation over models, parameters, and experiment outcomes.

$$U(d) = \sum_{m=1}^{K} p(m) \int \int u(d, \theta_m, y)\, \underbrace{p(y|\theta_m, d)}_{\substack{\text{Likelihood function;} \\ \text{E.g., binomial in } f(\theta_m)}}\, p(\theta_m)\, dy\, d\theta_m$$

Design Optimization

- A principled approach from Bayesian decision theory
  - Treat each possible design as a gamble whose payoff is determined by the outcome of an experiment carried out with that design.
  - Compute an ”expected utility” of each design by taking an expectation over models, parameters, and experiment outcomes.

$$U(d) = \sum_{m=1}^{K} p(m) \int\int u(d, \theta_m, y)\, p(y|\theta_m, d)\, p(\theta_m)\, dy\, d\theta_m$$

priors

Design Optimization

- A principled approach from Bayesian decision theory
  - Treat each possible design as a gamble whose payoff is determined by the outcome of an experiment carried out with that design.
  - Compute an ”expected utility” of each design by taking an expectation over models, parameters, and experiment outcomes.

$$U(d) = \sum_{m=1}^{K} p(m) \int \int u(d, \theta_m, y)\, p(y|\theta_m, d)\, p(\theta_m)\, dy\, d\theta_m$$

Priors can be updated between stages

Bayesian updating

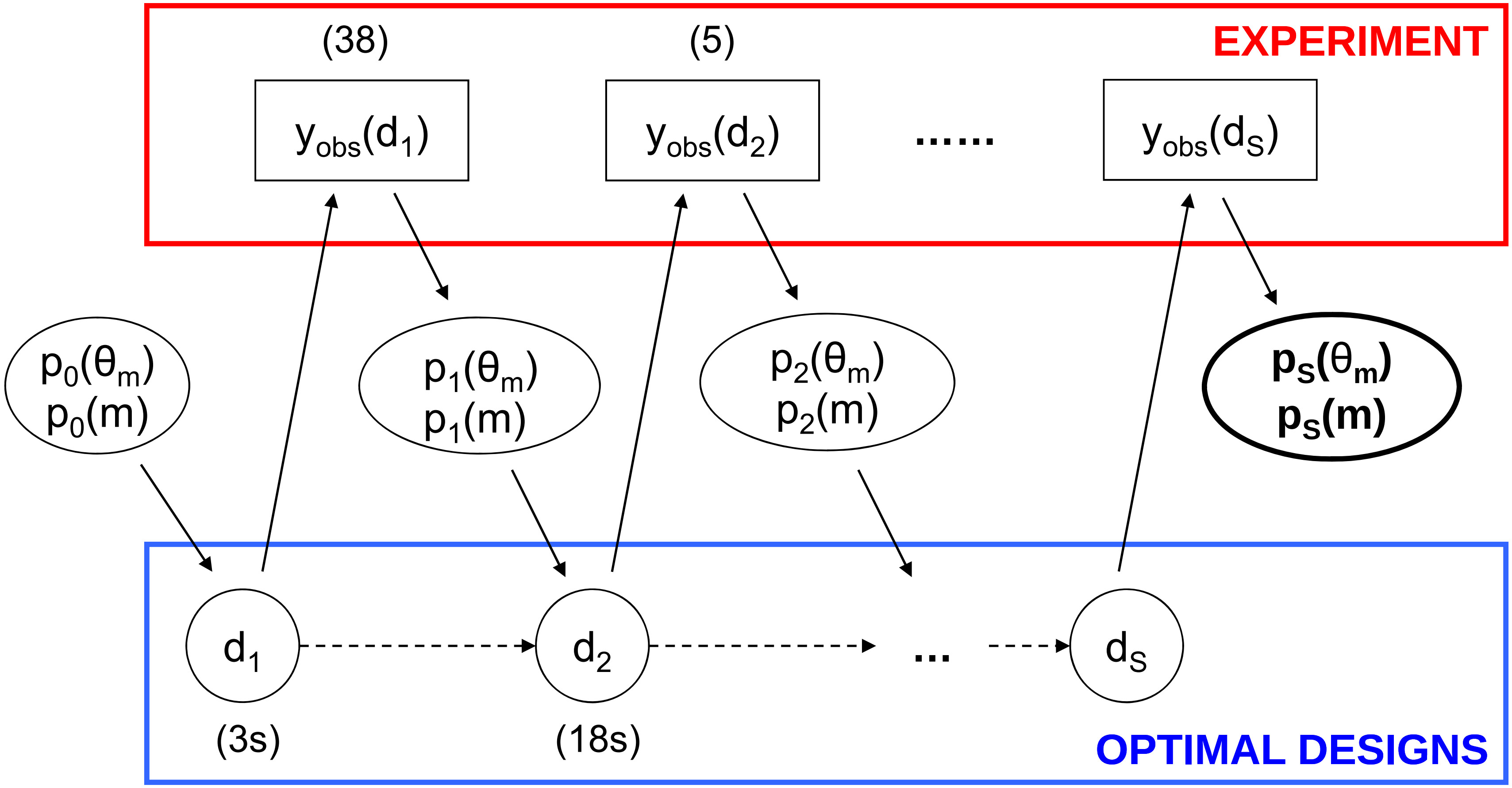
- Model posterior at stage S = 1,2,..
  - Updated from **Bayes factor calculation**

$$p_S(m) = \frac{p_0(m)}{\sum_{k=1}^{K} p_0(k) BF_{(k,m)}(y(d_S))_{p_{S-1}(\theta)}} \qquad (m = 1,\ldots,K)$$

- Parameter posterior at stage S = 1,2,...
  - Updated using **Bayes rule**

$$p_S(\theta_m) = \frac{p(y_S \mid \theta_m, d_S) p_{S-1}(\theta_m)}{\int p(y_S \mid \theta_m, d_S) p_{S-1}(\theta_m) d\theta_m} \qquad (m = 1,\ldots,K)$$

Adaptive Design Optimization (ADO)



Utility function

- Selection of a utility function that adequately captures the goals of the experiment is an integral part of ADO.

$$U(d) = \sum_{m=1}^{K} p(m) \int \int \underline{u(d, \theta_m, y)}\, p(y|\theta_m, d)\, p(\theta_m)\, dy\, d\theta_m$$

Utility function

- Selection of a utility function that adequately captures the goals of the experiment is an integral part of ADO.

Take $U(d)$ to be the mutual information of the random variables $Y|d$ and M

$$U(d) = I(M;Y|d)$$

Random variable over models under consideration

Random variable over outcomes of an experiment with design d

Utility function

- Selection of a utility function that adequately captures the goals of the experiment is an integral part of ADO.

Take $U(d)$ to be the <u>mutual information</u> of the random variables $Y \mid d$ and M

$$\begin{aligned} U(d) &= I(M;Y|d) \\ &= \underbrace{H(M)}_{\text{Entropy of } M} - \underbrace{H(M|Y,d)}_{\substack{\text{Conditional entropy of } M \\ \text{given } Y \text{ and } d}} \end{aligned}$$

Utility function

- Selection of a utility function that adequately captures the goals of the experiment is an integral part of ADO.

Take $U(d)$ to be the mutual information of the random variables $Y|d$ and M

$$\begin{aligned} U(d) &= I(M;Y|d) \\ &= \underbrace{H(M)}_{\text{Entropy of } M} - \underbrace{H(M|Y,d)}_{\substack{\text{Conditional entropy of } M \\ \text{given } Y \text{ and } d}} \end{aligned}$$

Essentially, U(d) measures the amount of information about true model that would be provided by an experiment with design d.

Utility function

Mutual information:

$$I(M;Y\ d) = \sum_{m=1}^{K} p(m) \int\int \log \frac{p(m|y,d)}{p(m)}\, p(y|\theta_m,d)\, p(\theta_m)\ dy\, d\theta_m$$

Utility function

Mutual information:

$$I(M;Y\ d) = \sum_{m=1}^{K} p(m) \int \int \log \frac{p(m|y,d)}{p(m)}\, p(y|\theta_m, d)\, p(\theta_m)\ dy\, d\theta_m$$

Expected utility:

$$U(d) = \sum_{m=1}^{K} p(m) \int \int u(d, \theta_m, y)\, p(y|\theta_m, d)\, p(\theta_m)\, dy\, d\theta_m$$

Utility function

Mutual information:

$$I(M;Y\ d) = \sum_{m=1}^{K} p(m) \int\int \underline{\log \frac{p(m|y,d)}{p(m)}}\ p(y|\theta_m,d)\, p(\theta_m)\ dy\, d\theta_m$$

Expected utility:

$$U(d) = \sum_{m=1}^{K} p(m) \int\int \underline{u(d,\theta_m,y)}\, p(y|\theta_m,d)\, p(\theta_m)\, dy\, d\theta_m$$

So we let $u(d,\theta_m,y) = \log \frac{p(m|y,d)}{p(m)}$

Utility function

- Interpretation of the local utility function:

$$u(d, \theta_m, y) = \log \frac{p(m|y,d)}{p(m)}$$

1. A design that increases our certainty about the true model upon observation of the outcome is more valuable than one that does not.

Utility function

- ☐ Interpretation of the local utility function:

Expanding and simplifying:

$$u(d, \theta_m, y) = -\log \sum_{k=1}^{K} p(k) BF_{(k,m)}(y)$$

1. A design that increases our certainty about the true model upon observation of the outcome is more valuable than one that does not.

2. The designs that are favored by this utility function are those that are expected to produce the most amount of evidence in favor of the true model.

Computation

- Finding an optimal design requires simultaneous optimization and high-dimensional integration (Muller, Sanso & De Iorio, 2004, *JASA*).

$$U(d) = \sum_{m=1}^{K} p(m) \int\int u(d, \theta_m, y)\, p(y|\theta_m, d)\, p(\theta_m)\, dy\, d\theta_m$$

$$d^* = \underset{d}{\operatorname{argmax}}\{U(d)\}$$

- Computation achieved by Sequential Monte Carlo (SMC) particle filtering algorithm with simulated annealing (Amzal, Bois, Parent & Robert, 2006, *JASA*).

ADO in Action

Simulated experiments

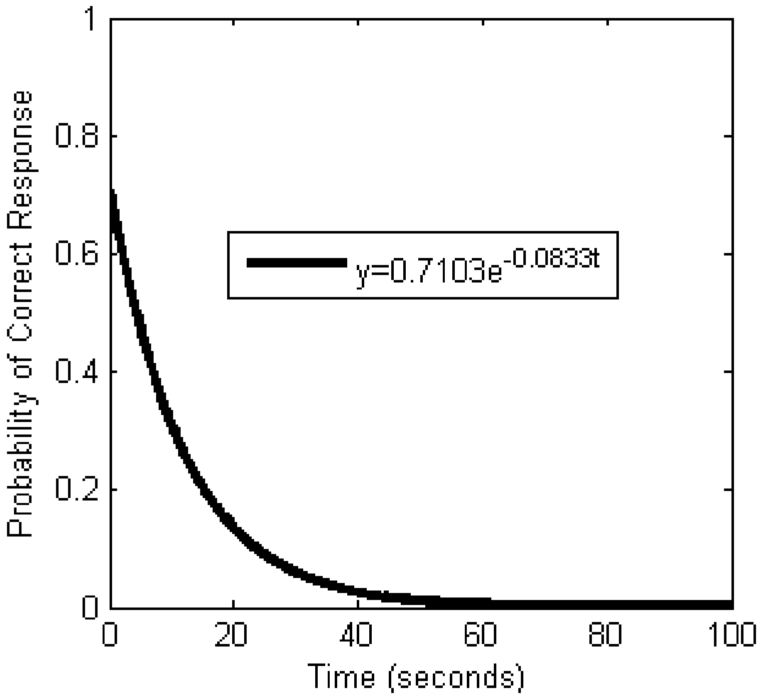
- Generate data at each stage from some true model (e.g. EXP with a=.71 and b=.08) at the optimal retention interval identified by ADO algorithm.

Simulated experiments

- Generate data at each stage from some true model (e.g. EXP with a=.71 and b=.08) at the optimal retention interval identified by ADO algorithm.
- The idea is to see how many stages it takes to accumulate enough data to decisively identify the true model according to some model selection criterion (e.g. Bayes factor)

Simulation results

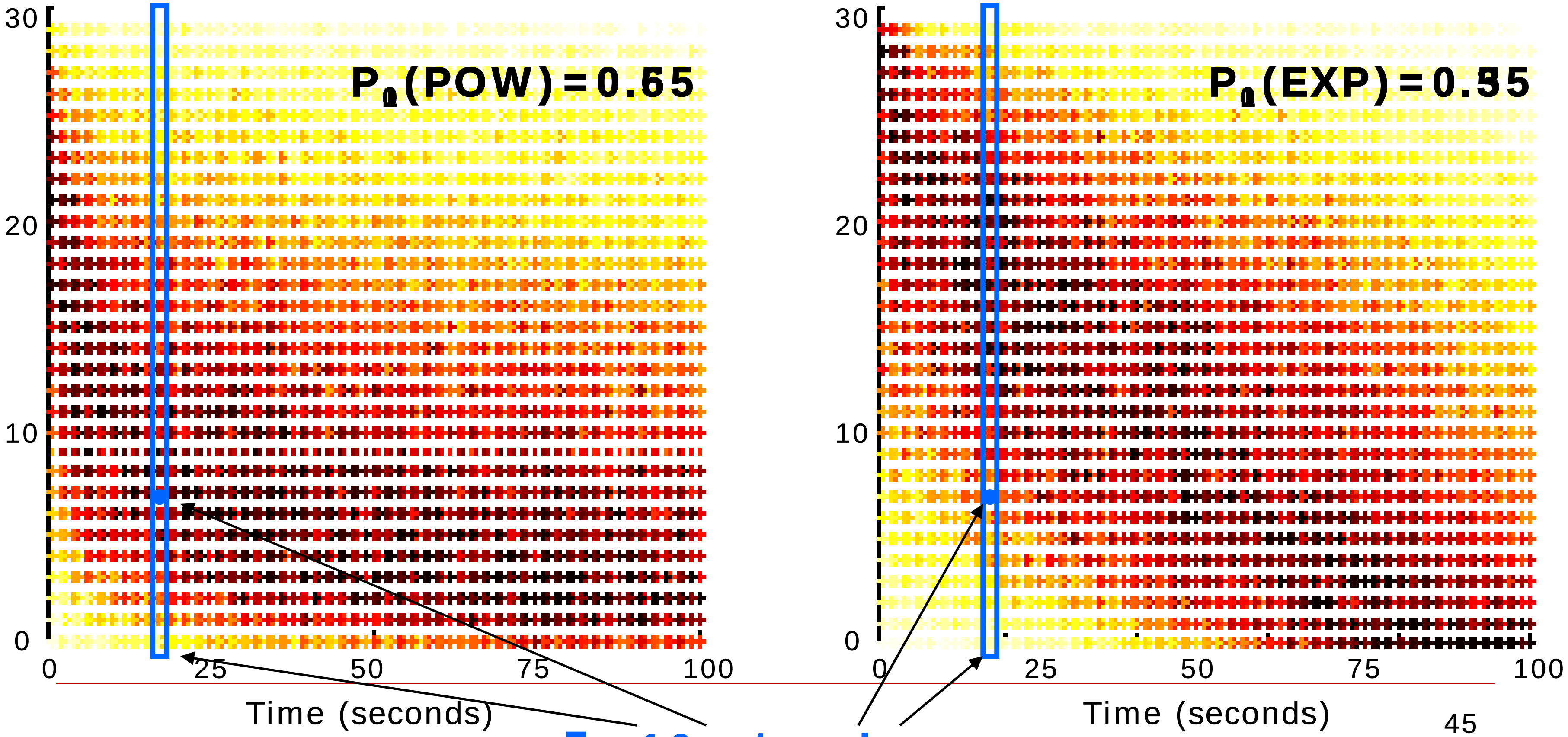
- Data generated from EXP with a=0.71 and b=0.08



Simulation results

- ☐ Stage 1

[illegible] (30 trials)



7 correct responses

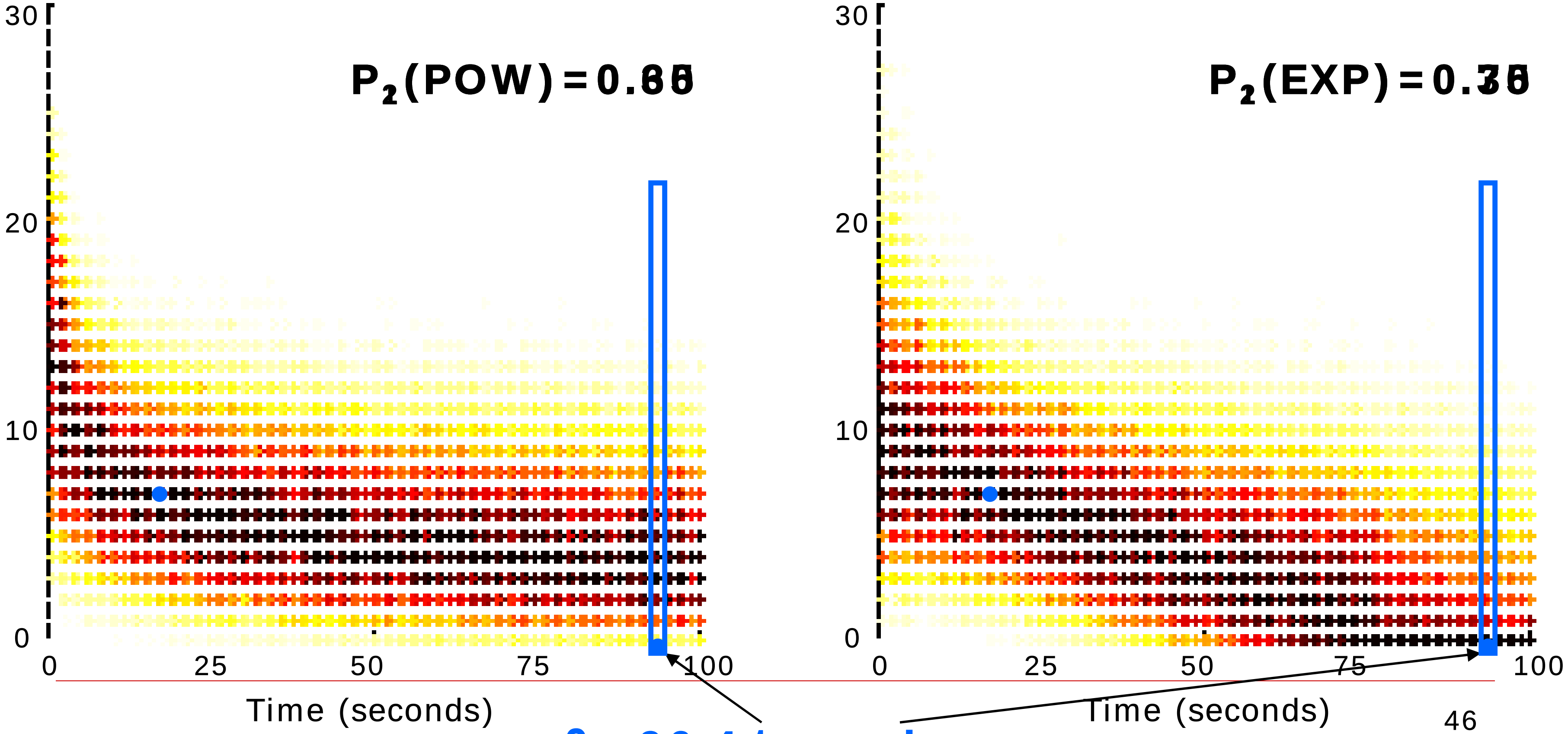
16 seconds

Simulation results

- Stage 2

Collect data (20 trials)

Compute posterior probabilities



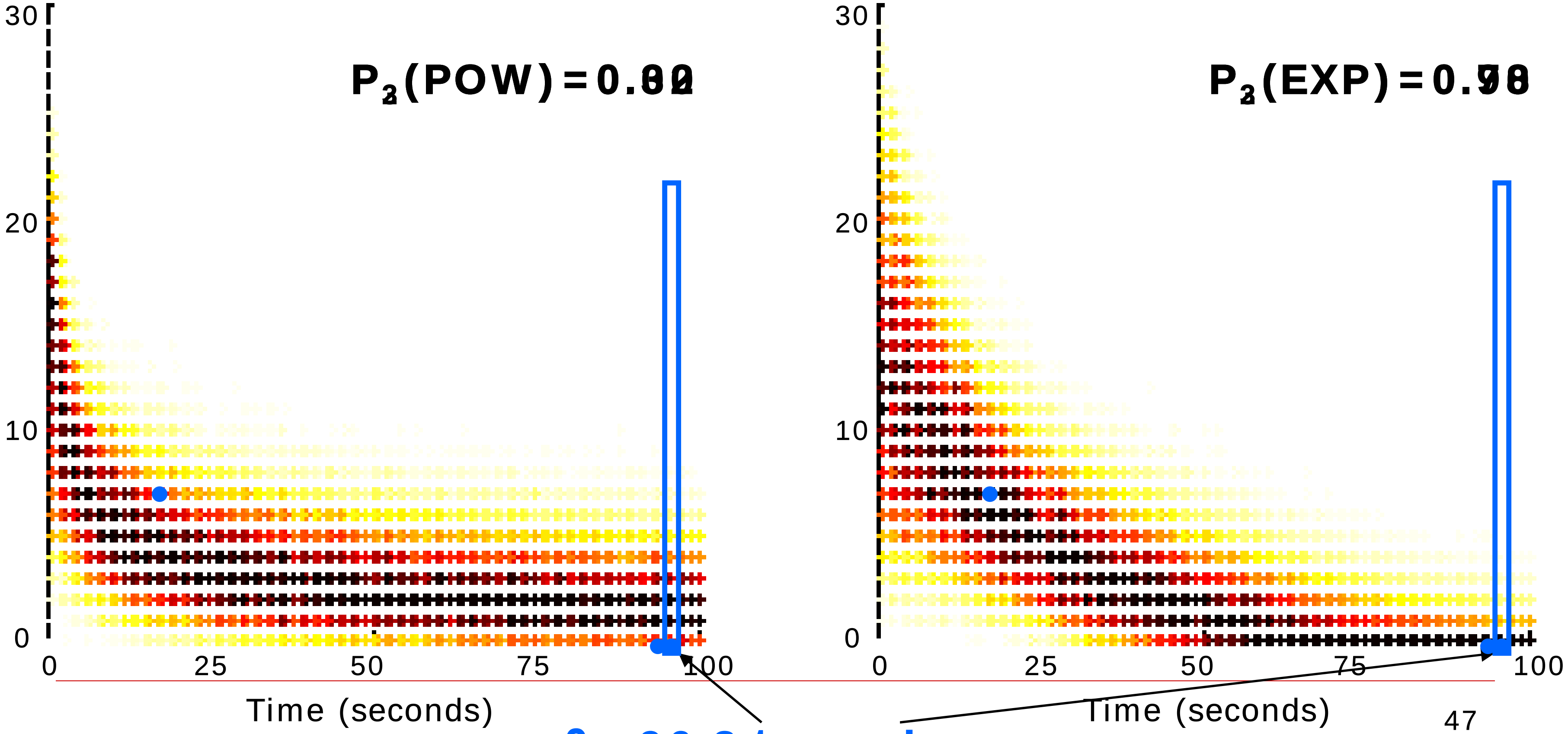
0 correct responses

96.4 seconds

Simulation results

- Stage 3

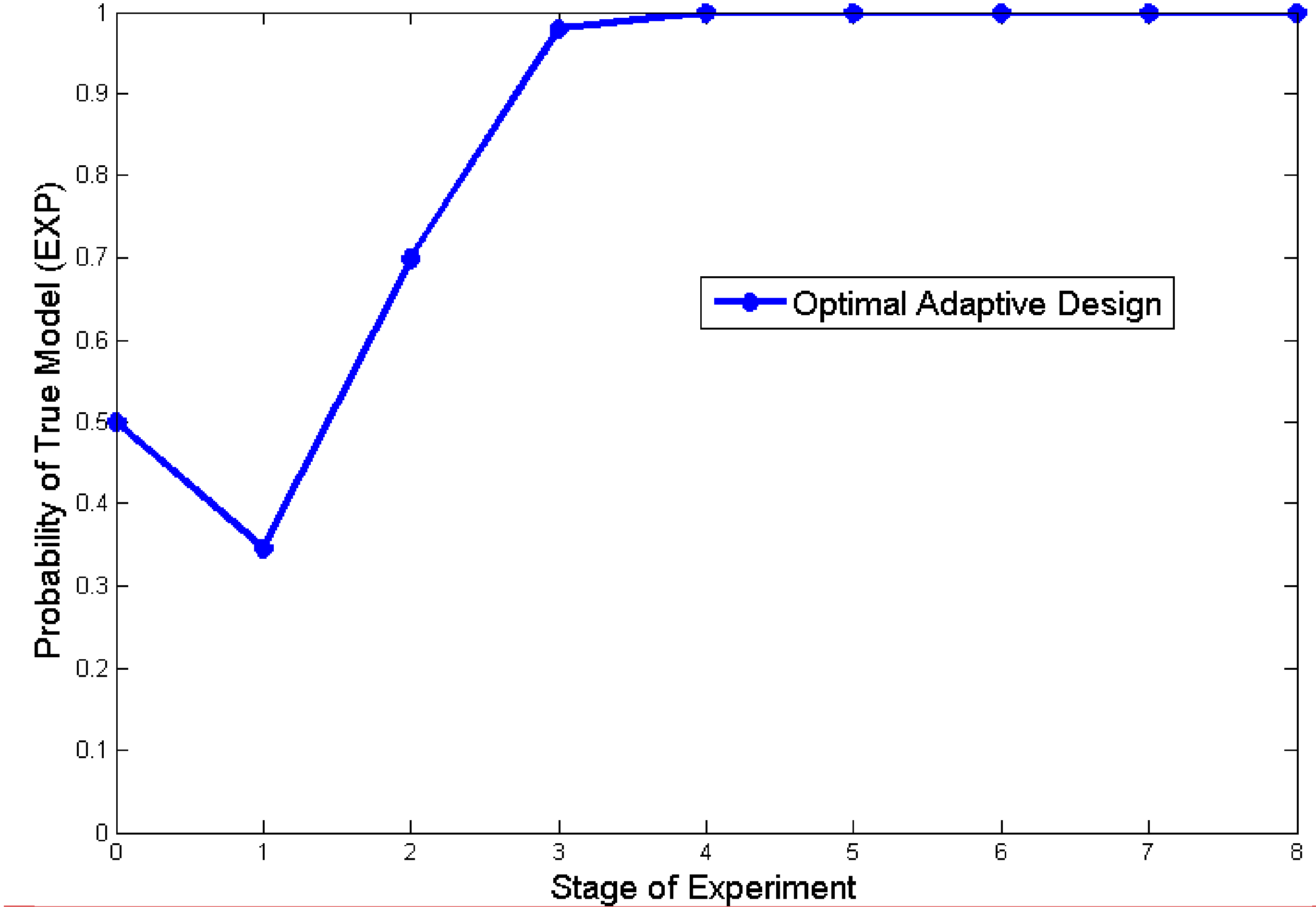
[illegible]



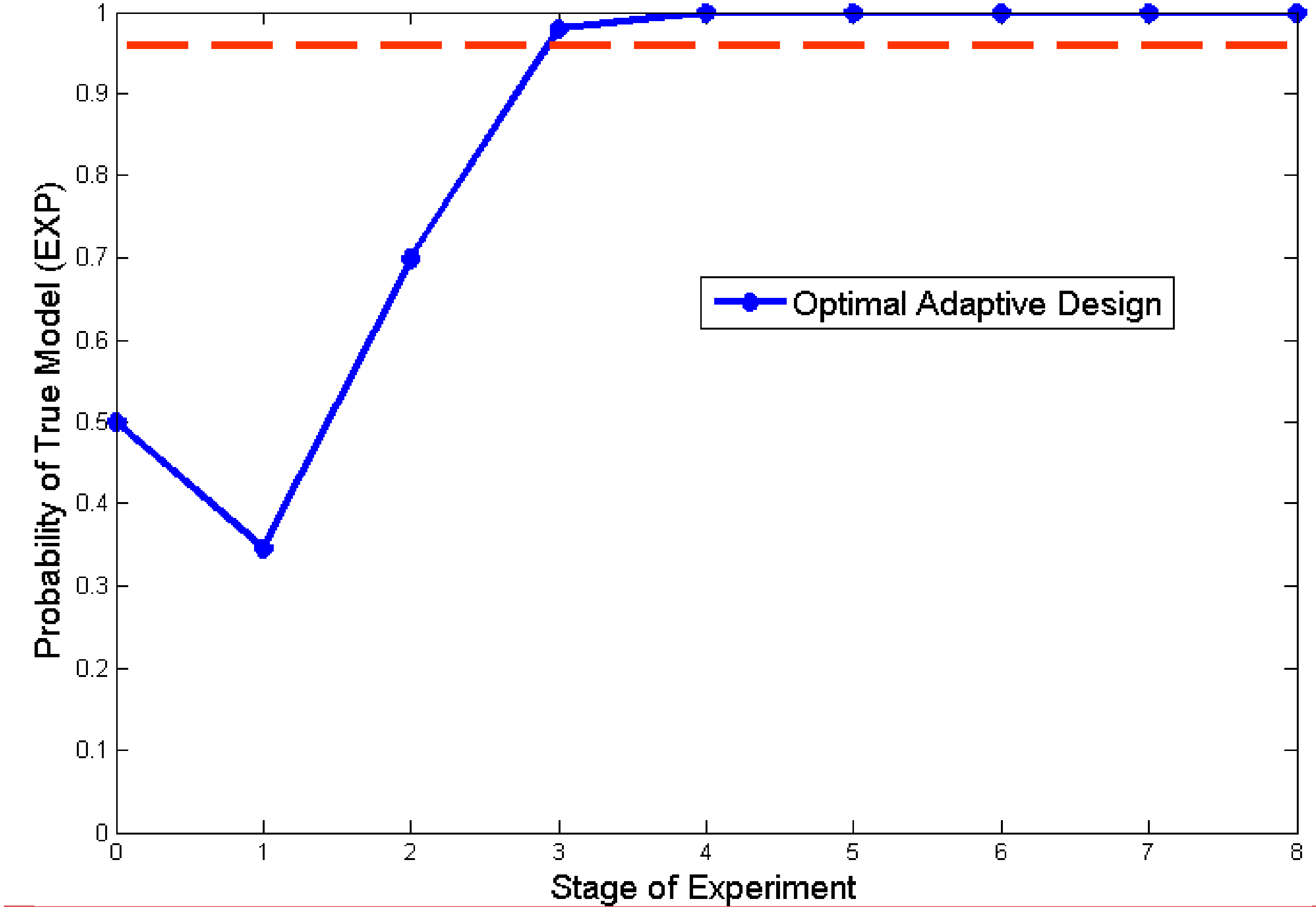
0 correct responses

96.8 seconds

Simulation Results

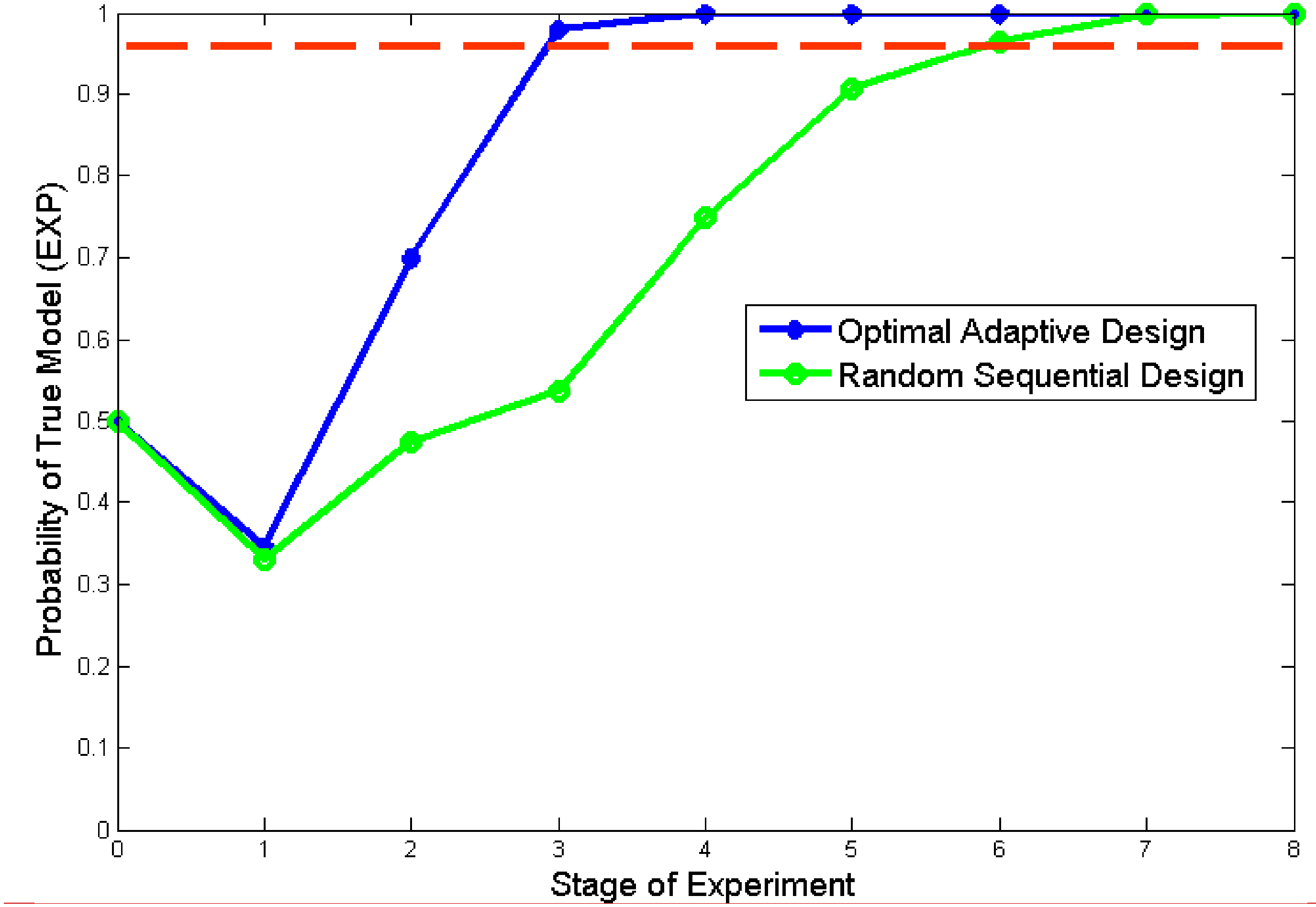


Simulation Results



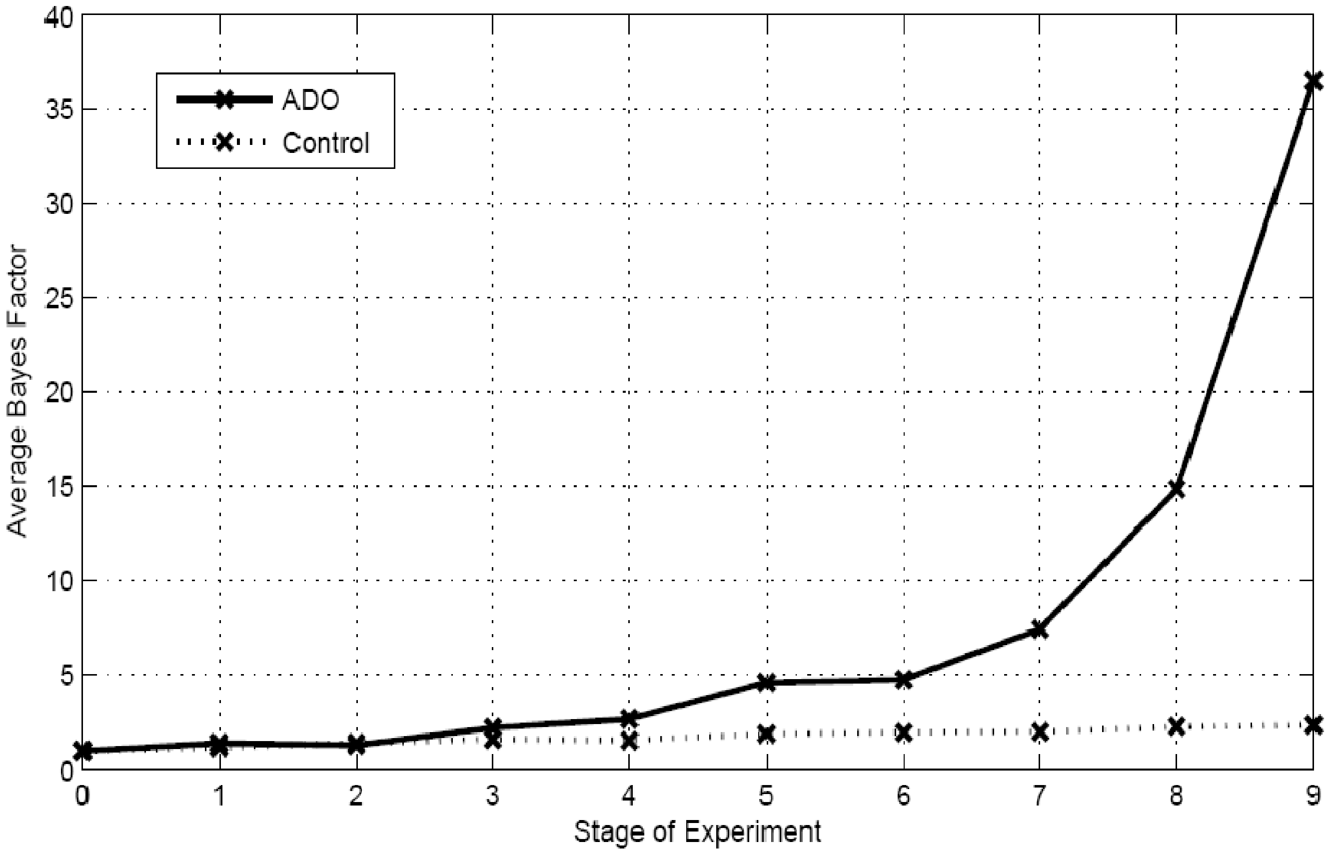
Bayes Factor = 20 (very strong evidence for EXP)

Simulation Results



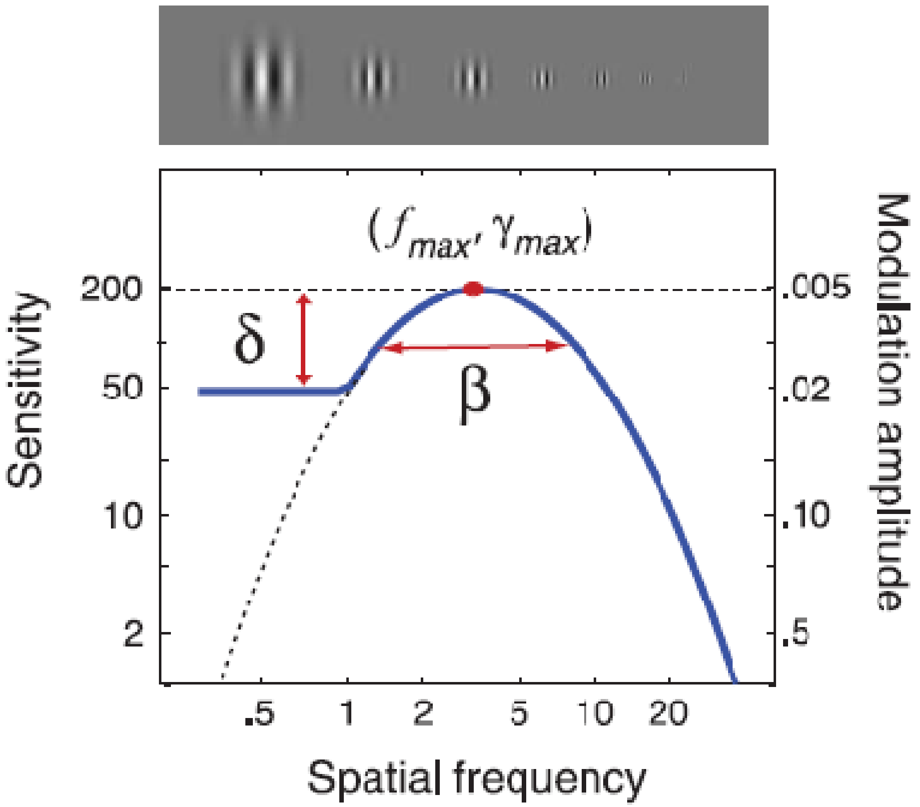
Bayes Factor = 20 (very strong evidence for EXP)

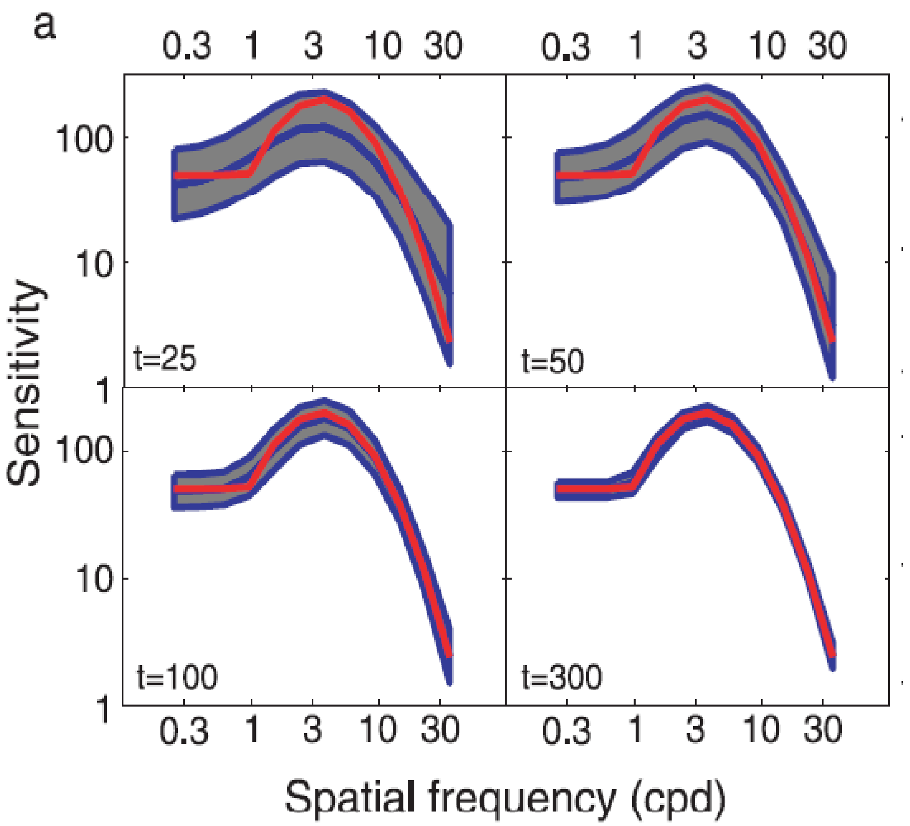
Results with Human Participants



Designing a Visual Psychophysics Experiment

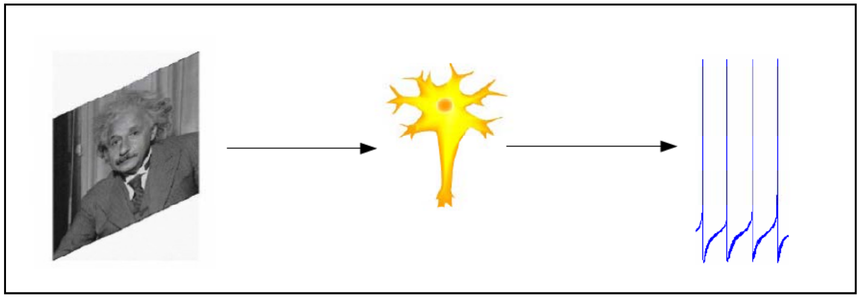
Lesmes, Jeon, Lu & Dosher (2006); Lesmes, Lu, Baek & Albright (2010)

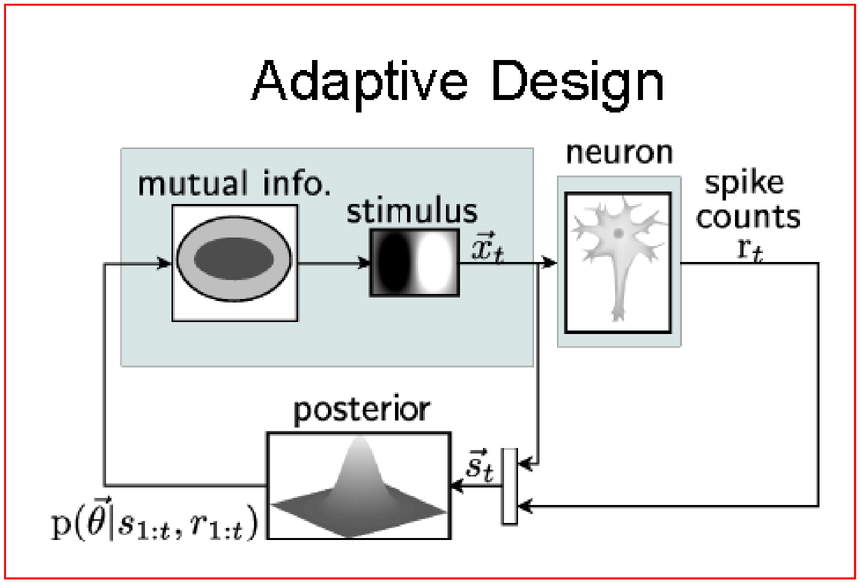



a
0.3
1
3
10
30
0.3
1
3
10
30
100
10
1
100
10
1
Sensitivity
t=25
t=50
t=100
t=300
0.3
1
3
10
30
0.3
1
3
10
30
Spatial frequency (cpd)

Designing a Neurophysiology Experiment

Lewi, Butera & Paninski (2009)





Conclusion

- **Adaptive design optimization (ADO)** is a promising new tool that facilitates efficient collection of data in experiments discriminating formal models.

- Current work
  - **ADO** for discriminating models of decision making under uncertainty.
  - **ADO** for designing numerical estimation experiments with children.

“Much ADO about Nothing”

- William Shakespeare

Thank You