Experience Guided Search: A Bayesian Perspective on Cognitive Control

Michael C. Mozer

Computer Science Department and
Institute of Cognitive Science
University of Colorado, Boulder

Collaborators

Karthik Venkatesh
Brett Roads
Bradley Mitchell
Matthew Wilder
C. Ted Wright (UC Irvine, Cognitive Science)
Matt Jones (U. Colorado, Psychology)

Count the number of pennies.

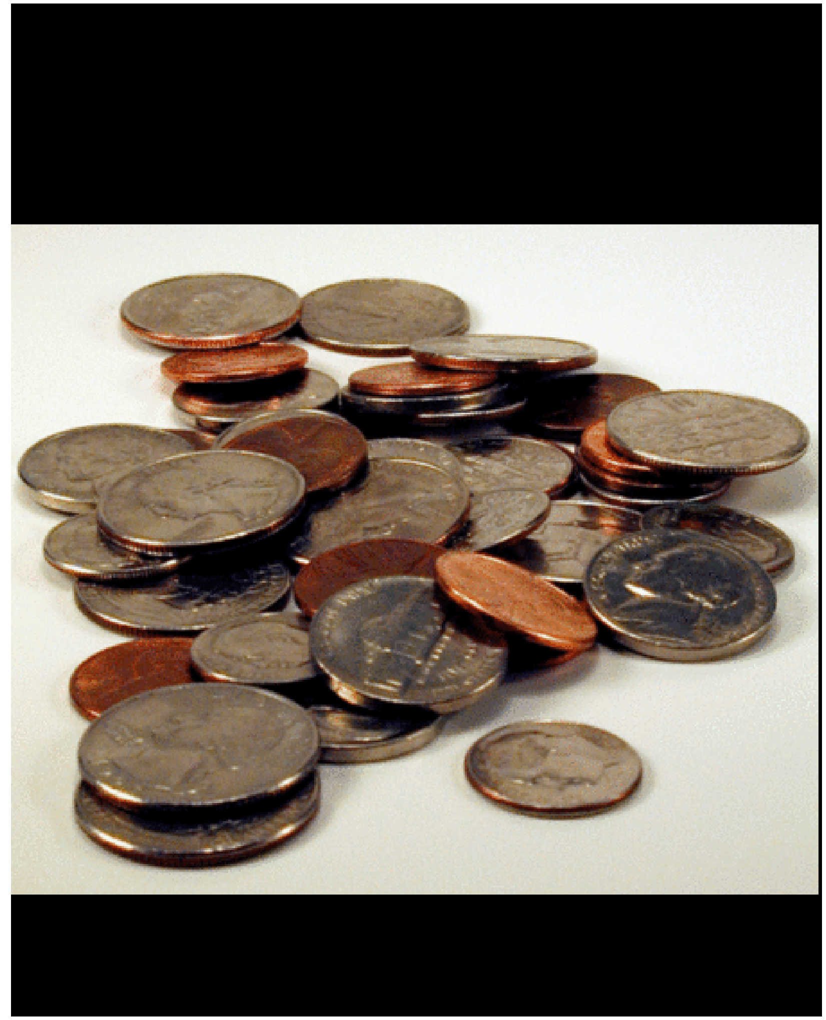
Which coin doesn’t touch the others?

Find Jefferson.

Are there more heads or tails?

Are all the coins U.S.?

Are any coins the wrong size?



Attentional Control

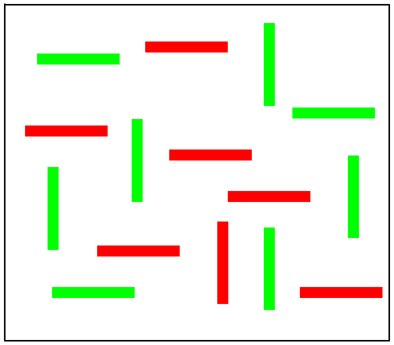
The ability to deploy attention based on task demands

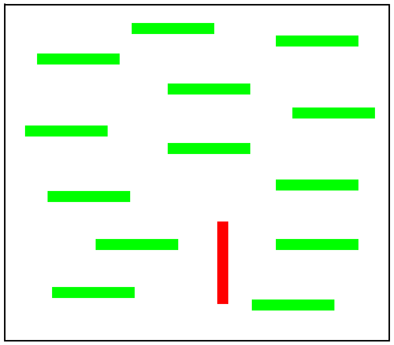
Visual Search

Search for a *target* object among *distractors*.

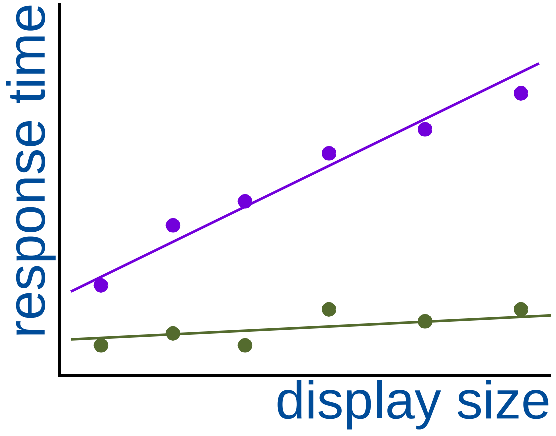
Visual Search in the Lab

e.g., find the red vertical bar





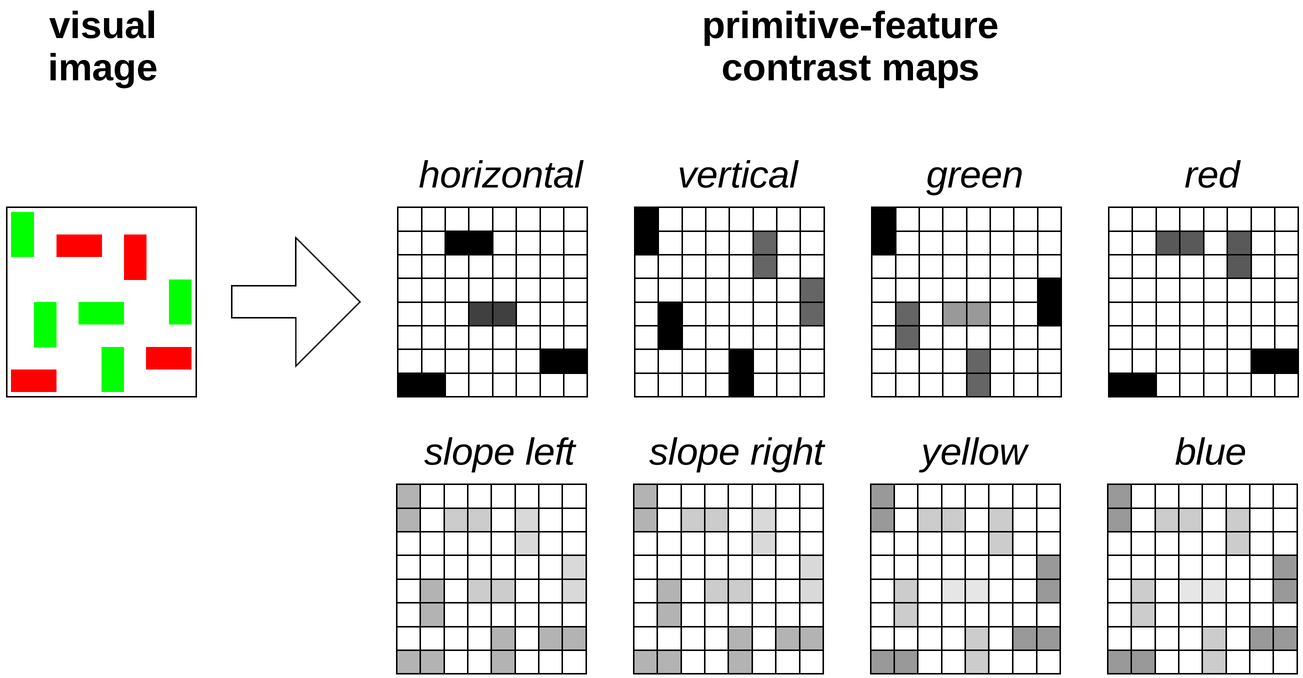
Examine time to detect target presence/absence as a function of display size

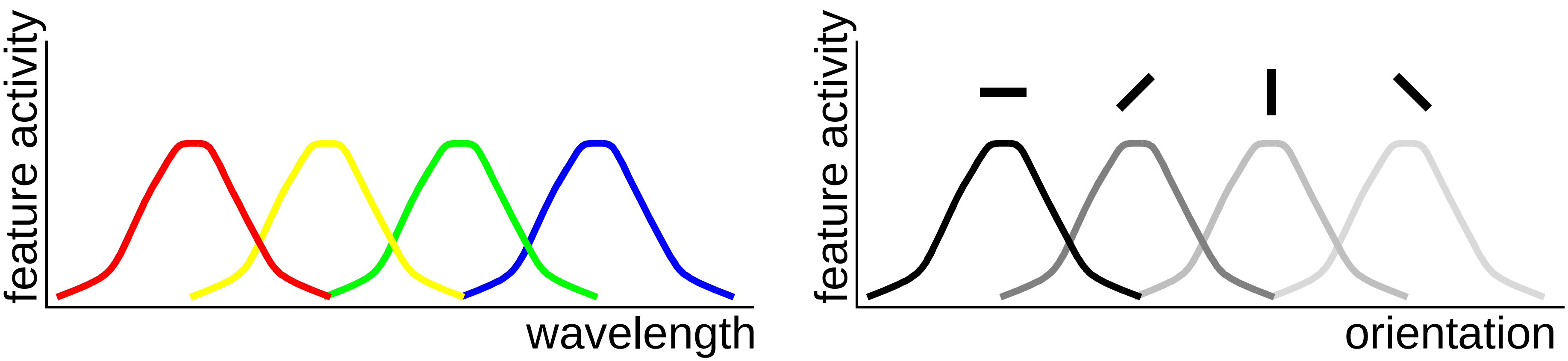


Task difficulty ~ search slope

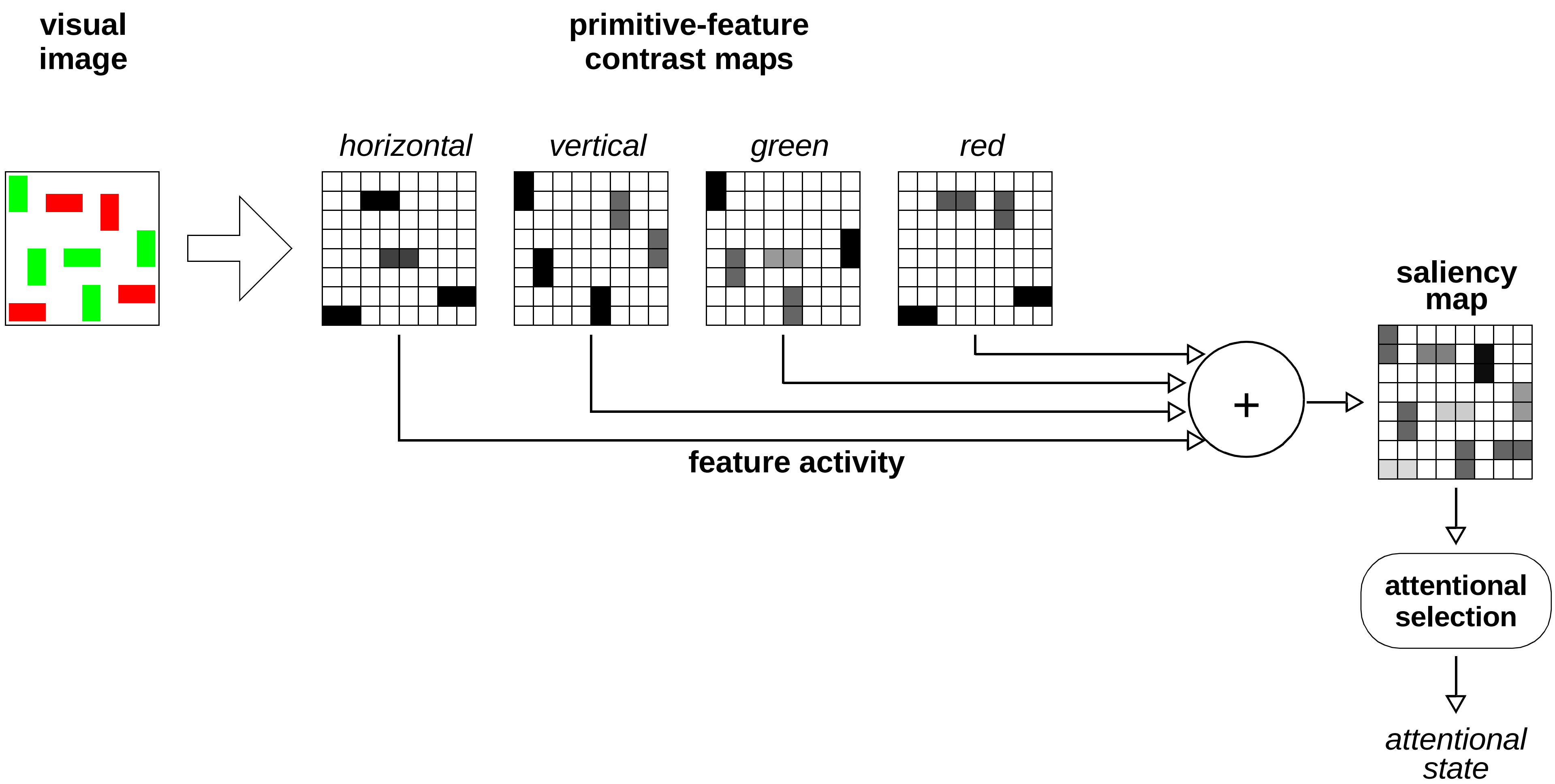
Guided Search Model (Wolfe, 1994, 1997, 2007)

Guided Search Model (Wolfe, 1994, 1997, 2007)





Guided Search Model (Wolfe, 1994, 1997, 2007)

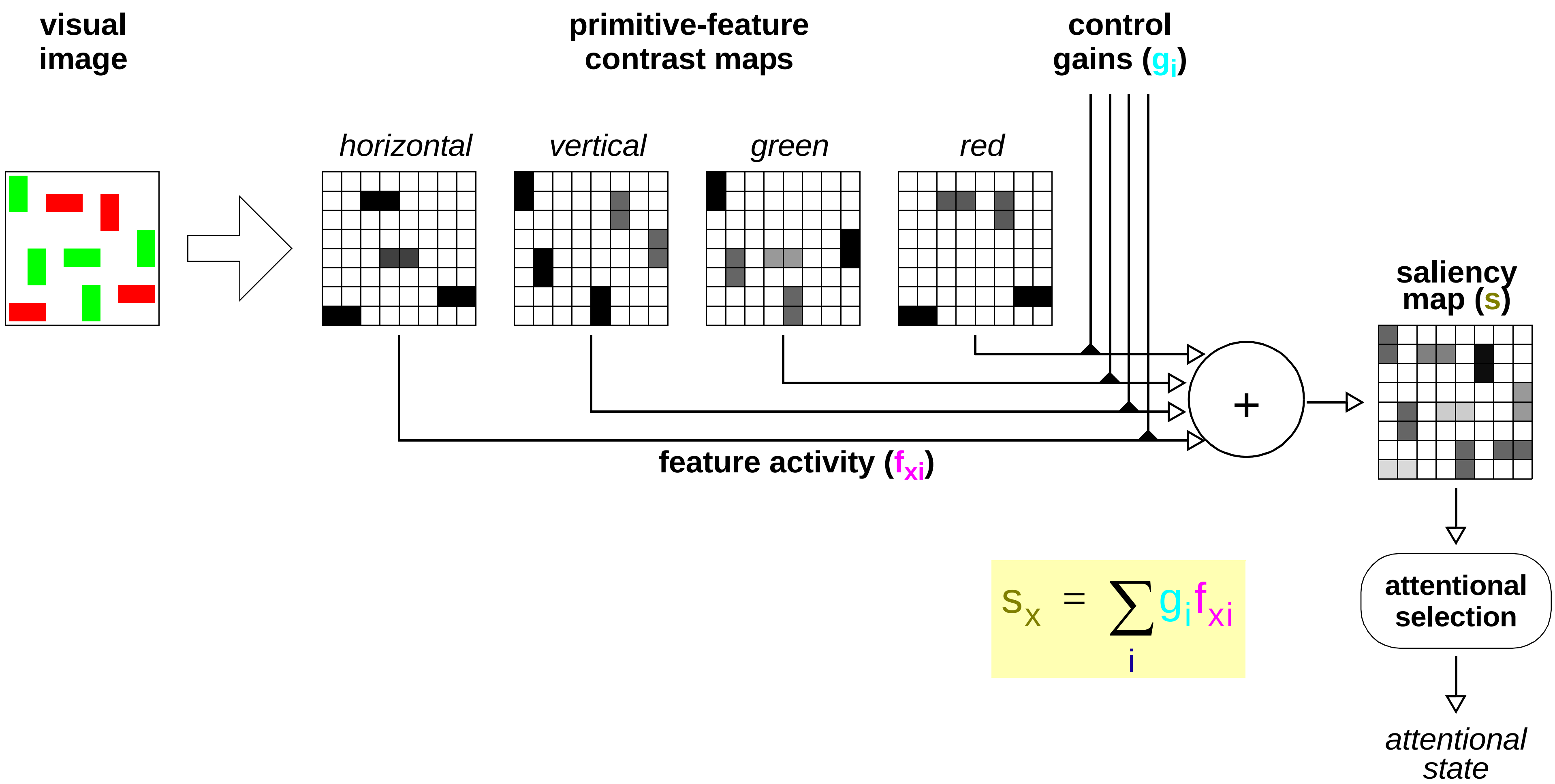


Saliency map prioritizes locations for search.

GS2.0 response rule

Response_time = $\mu_0 + \mu_1$ target_ranking

Guided Search Model (Wolfe, 1994, 1997, 2007)

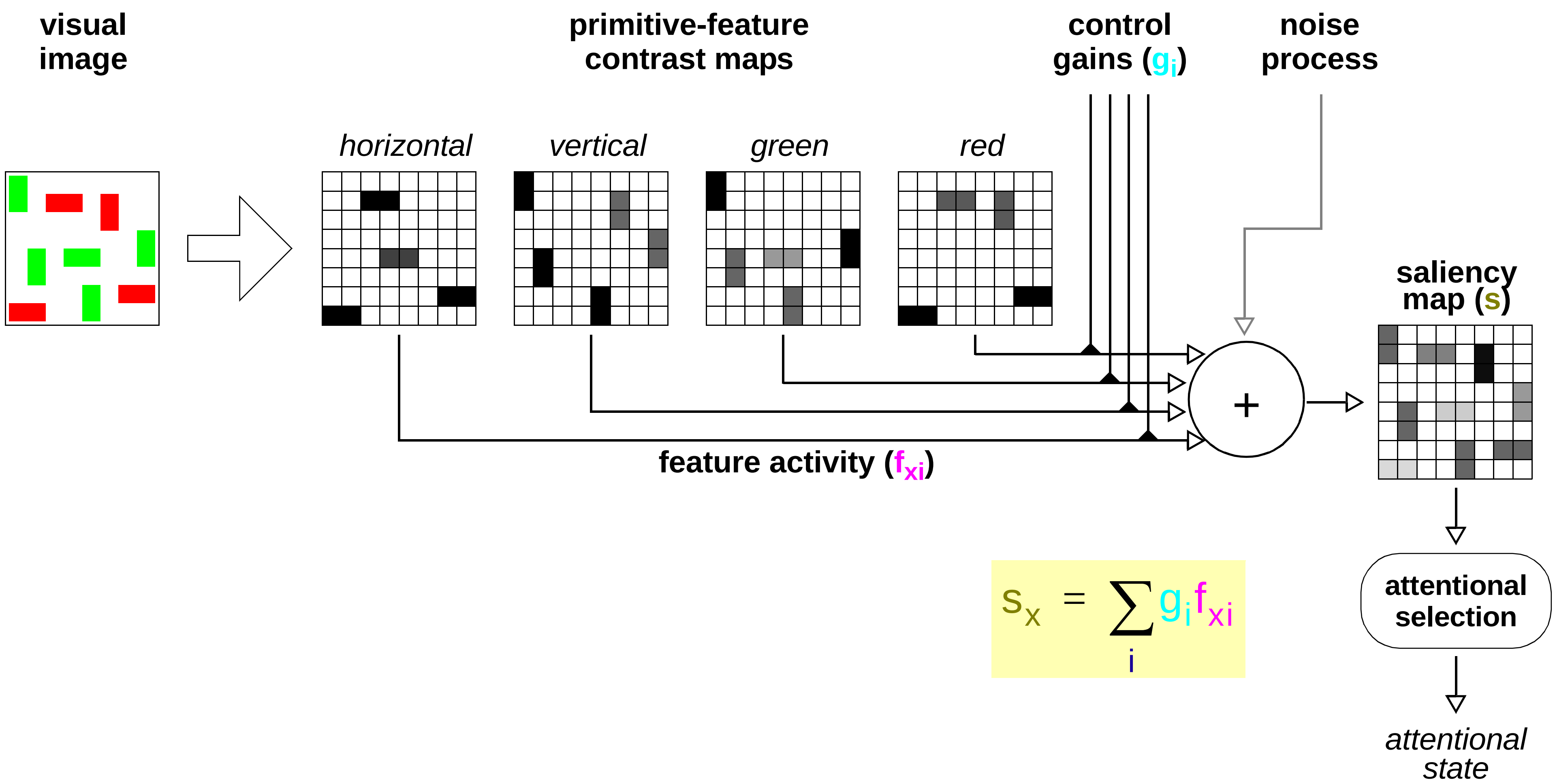


$$s_x = \sum_i g_i f_{xi}$$

Gains guide attention to task-relevant locations.

a.k.a. attentional weights, attentional set

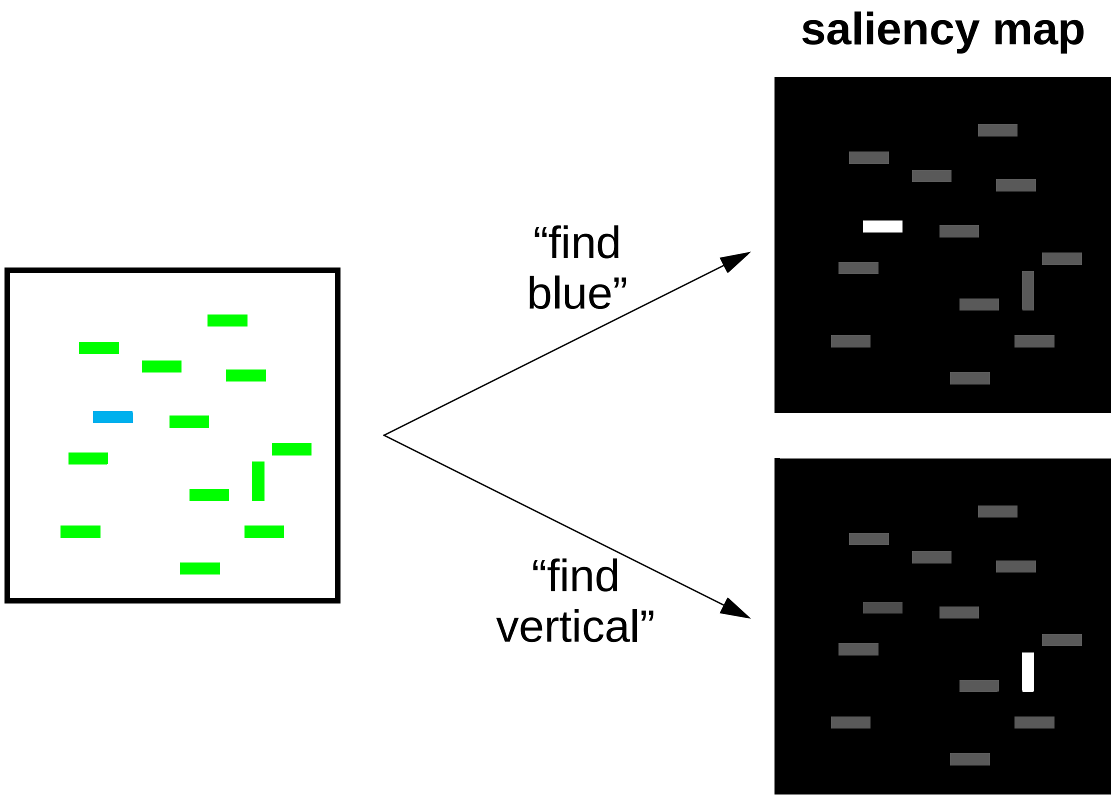
Guided Search Model (Wolfe, 1994, 1997, 2007)



$$s_x = \sum_i g_i f_{xi}$$

Noise corruption of saliency map.

How Guided Search Is Supposed To Work



How Are Gains Determined?

Guided Search doesn't specify

Common intuition

Gain ~ how well feature discriminates targets from nontargets

How Are Gains Determined?

Guided Search doesn't specify

Common intuition

Gain ~ how well feature discriminates targets from nontargets

ρ_{i1}	ρ_{i0}
average activity of feature *i* at locations containing target	**average activity of feature *i* at locations containing nontargets**

How Are Gains Determined?

Guided Search doesn't specify

Common intuition

Gain ~ how well feature discriminates targets from nontargets

ρ_{i1} **>>** ρ_{i0} $\Rightarrow$ **high gain**

average activity of feature *i* at locations containing target

average activity of feature *i* at locations containing nontargets

How Are Gains Determined?

Guided Search doesn't specify

Common intuition

Gain ~ how well feature discriminates targets from nontargets

ρ_{i1} **<<** ρ_{i0} $\Rightarrow$ **low gain**

average activity of feature *i* at locations containing target	**average activity of feature *i* at locations containing nontargets**

Experience-Guided Search

Experience-Guided Search

Adopt same basic architecture as Guided Search

- feature extraction
- contrast enhancement

But frame the model’s objective in probabilistic terms...

$$\textbf{Saliency} \equiv P(T_x \mid \mathbf{F}_x, \rho)$$

task statistics – learned thru experience

feature activity (vector) at location x

target at location x? (1=true, 0=false)

Saliency $\equiv P(T_x|\mathbf{F}_x, \rho)$

task statistics – learned thru experience

feature activity (vector) at location x

target at location x? (1=true, 0=false)

Bayes' Rule

$$P(T_x|\mathbf{F}_x, \rho) = \frac{P(T_x)P(\mathbf{F}_x|T_x, \rho)}{\sum_{t=0}^{1} P(T_x = t)P(\mathbf{F}_x|T_x = t, \rho)}$$

Saliency $\equiv P(T_x | \mathbf{F}_x, \rho)$

task statistics – learned thru experience

feature activity (vector) at location x

target at location x? (1=true, 0=false)

Bayes' Rule

$$P(T_x | \mathbf{F}_x, \rho) = \frac{P(T_x) P(\mathbf{F}_x | T_x, \rho)}{\sum_{t=0}^{1} P(T_x = t) P(\mathbf{F}_x | T_x = t, \rho)}$$

Saliency $\equiv P(T_x | \mathbf{F}_x, \rho)$

task statistics – learned thru experience

feature activity (vector) at location x

target at location x? (1=true, 0=false)

Bayes' Rule

$$P(T_x | \mathbf{F}_x, \rho) = \frac{P(T_x)P(\mathbf{F}_x | T_x, \rho)}{\sum_{t=0}^{1} P(T_x = t)P(\mathbf{F}_x | T_x = t, \rho)}$$

Saliency $\equiv P(T_x \mid \mathbf{F}_x, \rho)$

task statistics – learned thru experience

feature activity (vector) at location x

target at location x? (1=true, 0=false)

Bayes' Rule

$$P(T_x \mid \mathbf{F}_x, \rho) = \frac{P(T_x) P(\mathbf{F}_x \mid T_x, \rho)}{\sum_{t=0}^{1} P(T_x = t) P(\mathbf{F}_x \mid T_x = t, \rho)}$$

$P(\mathbf{F}_x \mid T_x, \rho)$ **is a task-specific model of the environment.**

Indicates visual system response ($\mathbf{F}_x$) for targets (T_x=1) vs. nontargets (T_x=0)

Modeling game

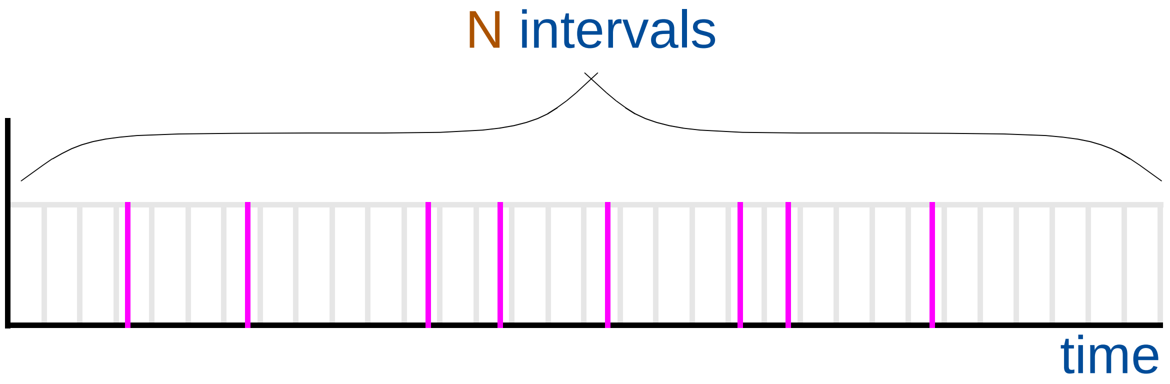
Specify a set of qualitative assumptions that define the environment model, and explore the consequences.

Key Assumptions

1. Feature responses are conditionally independent

$$P(\mathbf{F}_x | T_x, \rho) = \prod_i P(F_{xi} | T_x, \rho)$$

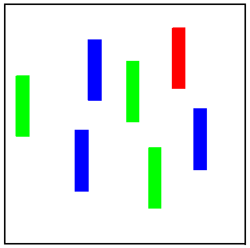
2. Feature detection is carried out by rate-coded spiking neuron



F_{xi}: count of the number of spikes observed for feature i at location x

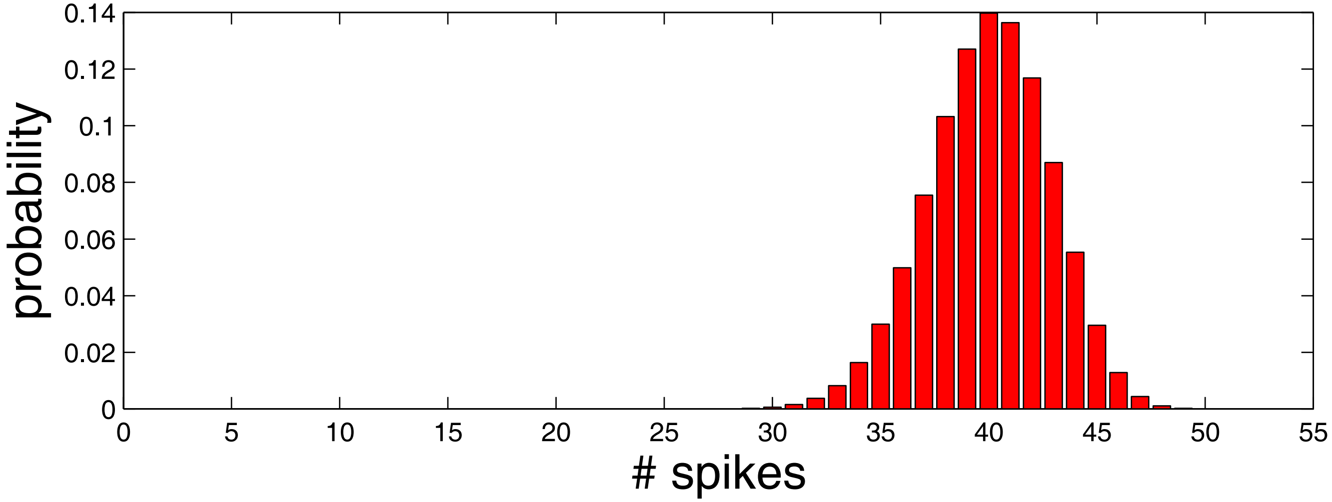
ρ_{it}: spike rate for feature *i* if the target is of type *t*

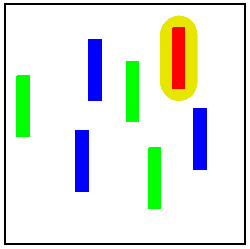
Consider search for a red object among non-red.
What will the response of the red feature detector be?



Consider search for a red object among non-red. What will the response of the red feature detector be?

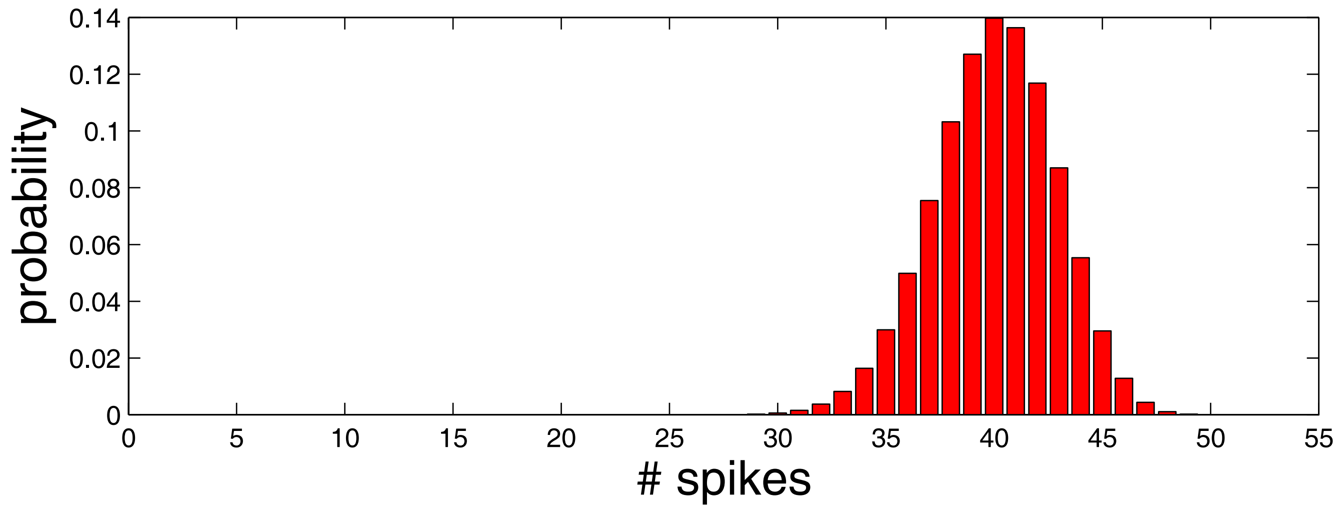
$P(F_{red} | T = 1, \rho_{red,1} = 0.8)$ **for N=50**

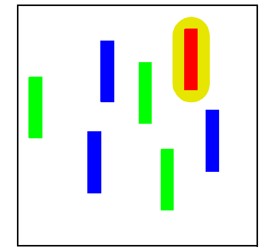




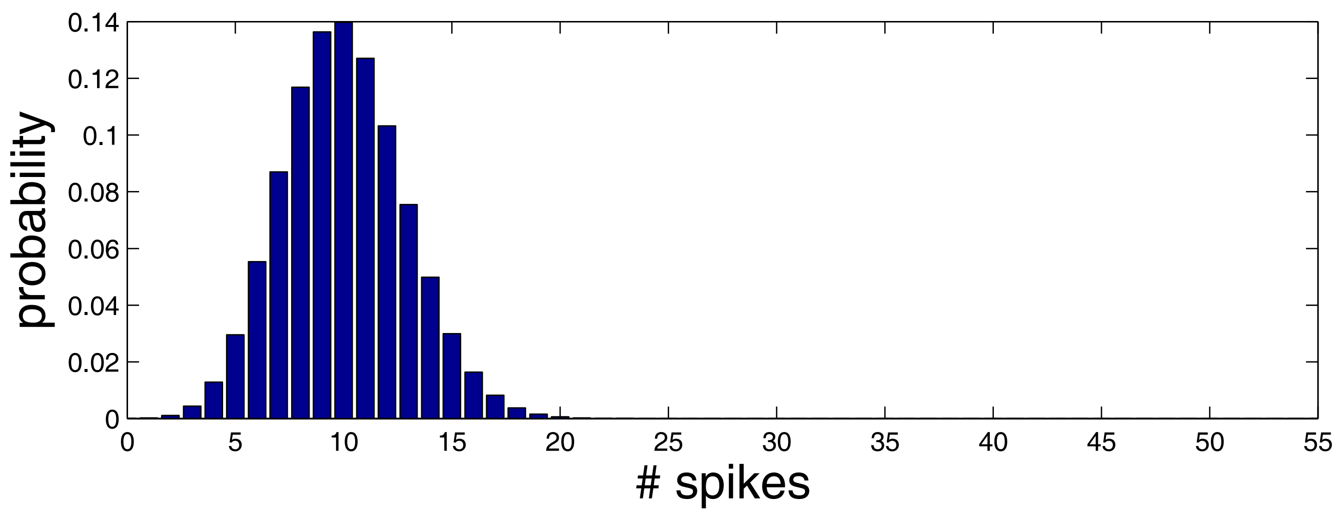
Consider search for a red object among non-red.
What will the response of the red feature detector be?

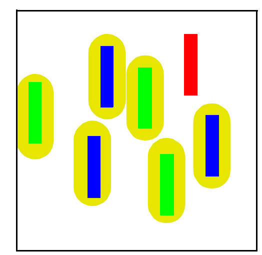
$P(F_{red} | T = 1, \rho_{red,1} = 0.8)$ **for N=50**





$P(F_{red} | T = 0, \rho_{red,0} = 0.2)$ **for N=50**





$$P(F_{xi} \mid T_x, \rho) \sim \mathrm{Binomial}(\rho_{it}, N)$$

number of time intervals

spiking rate of feature i
for target (t=1) or distractor (t=0)

$$P(F_{xi} \mid T_x, \rho) \sim \cancel{\text{Binomial}(\rho_{it}, N)}$$

number of time intervals

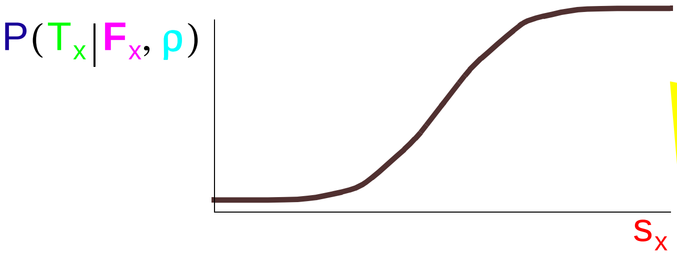
spiking rate of feature i
for target (t=1) or distractor (t=0)

$$\sim \text{Gaussian}(N\rho_{it}, N\rho_{it}(1-\rho_{it}))$$

These assumptions lead to...

$$P(T_x \mid \mathbf{F}_x, \rho) = \frac{1}{1 + e^{-(c_1 + c_2 s_x)}}$$

$$P(T_x | \mathbf{F}_x, \rho) = \frac{1}{1 + e^{-(c_1 + c_2 s_x)}}$$



Because attentional priority depends on relative saliency, we can substitute s_x for $P(T_x | \mathbf{F}_x, \rho)$.

$$P(T_x \mid \mathbf{F}_x, \rho) = \frac{1}{1 + e^{-(c_1 + c_2 s_x)}}$$

response of feature i
in location x

$$s_x = \sum_i \sum_{t=0}^{1} \frac{1-2t}{\rho_{it}(1-\rho_{it})} (f_{xi} - \rho_{it})^2$$

$$s_x = \sum_i \sum_{t=0}^{1} \frac{1-2t}{\rho_{it}(1-\rho_{it})} \left(f_{xi} - \rho_{it}\right)^2$$

$$s_x = \sum_i c_i f_{xi} + \tilde{c}_i f_{xi}^2$$

$$s_x = \sum_i \sum_{t=0}^{1} \frac{1-2t}{\rho_{it}(1-\rho_{it})}(f_{xi}-\rho_{it})^2$$

$$\frac{2(\rho_{i1}-\rho_{i0})}{(1-\rho_{i0})(1-\rho_{i1})} \qquad \frac{1}{\rho_{i0}(1-\rho_{i0})} - \frac{1}{\rho_{i1}(1-\rho_{i1})}$$

$$s_x = \sum_i c_i f_{xi} + \tilde{c}_i f_{xi}^2$$

$$\frac{2(\rho_{i1} - \rho_{i0})}{(1-\rho_{i0})(1-\rho_{i1})} \qquad \frac{1}{\rho_{i0}(1-\rho_{i0})} - \frac{1}{\rho_{i1}(1-\rho_{i1})}$$

Experience-Guided Search

$$s_x = \sum_i c_i f_{xi} + \tilde{c}_i f_{xi}^2$$

Guided Search

$$s_x = \sum_i c_i f_{xi}$$

Differences Between EGS and GS

1. EGS includes terms quadratic in f_{xi}

2. GS determines gains via heuristics or optimization;
   in EGS, gains are function of environment/task statistics

$$\frac{2(\rho_{i1}-\rho_{i0})}{(1-\rho_{i0})(1-\rho_{i1})} \qquad \frac{1}{\rho_{i0}(1-\rho_{i0})}-\frac{1}{\rho_{i1}(1-\rho_{i1})}$$

Experience-Guided Search

$$s_x = \sum_i c_i f_{xi} + \tilde{c}_i f_{xi}^2$$

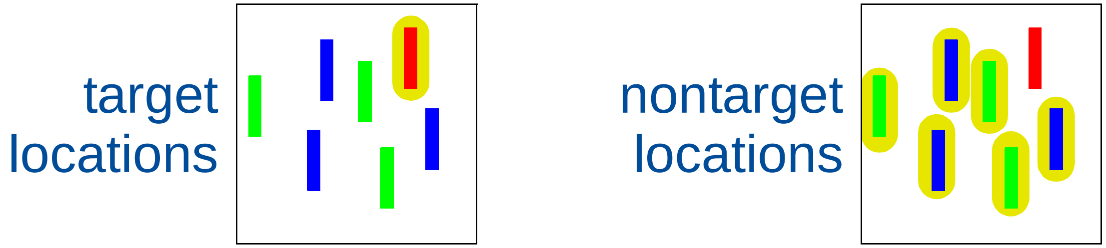
Guided Search

$$s_x = \sum_i c_i f_{xi}$$

How Are Environment Statistics (ρ) Learned?

How Are Environment Statistics (ρ) Learned?

As locations are inspected during a trial, a supervisory process labels each element as target or nontarget.



Given these observations, update (learn) the ρ_{it} via Bayesian inference.

Assumptions Underlying Learning

Assumptions Underlying Learning

1. **Prior (bias) that all features are considered relevant in the absence of experience**

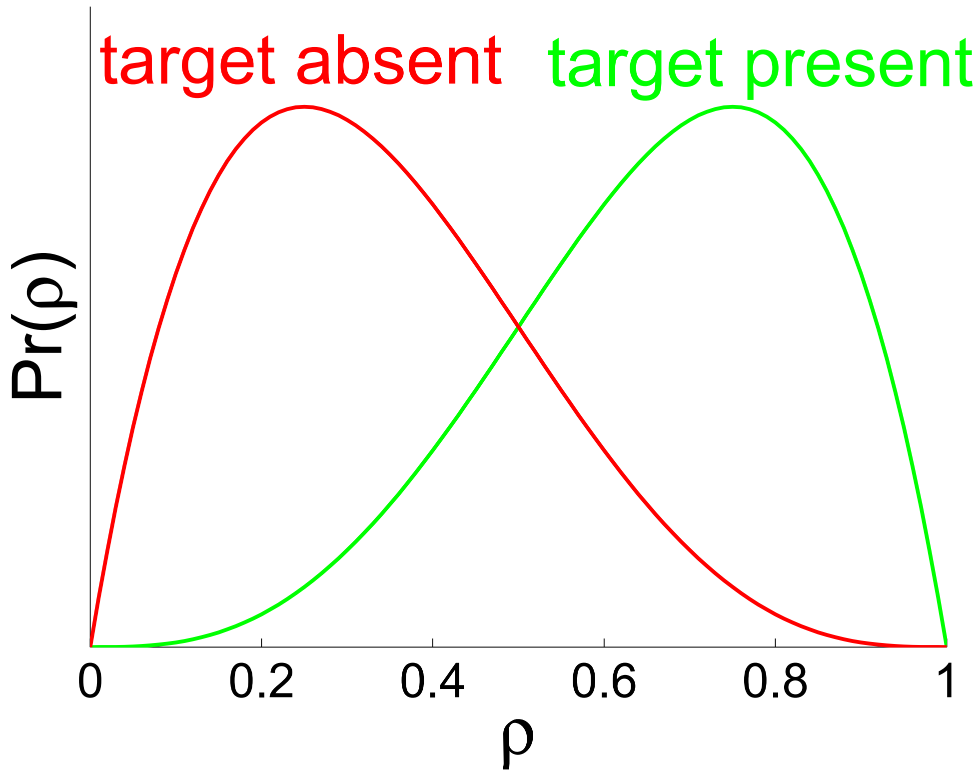
   Achieved by treating ρ as a Beta random variable with imaginary-count prior

Assumptions Underlying Learning

1. Prior (bias) that all features are considered relevant in the absence of experience

~~Achieved by treating ρ as a Beta random variable with imaginary-count prior~~

Encode not just the most likely value of ρ, but uncertainty distribution.



Assumptions Underlying Learning

1. **Prior (bias) that all features are considered relevant in the absence of experience**

   ~~Achieved by treating ρ as a Beta random variable with imaginary-count prior~~

   Encode not just the most likely value of ρ, but uncertainty distribution.

2. **Environment is nonstationary**

   With probability λ, environment and/or task can change.

Assumptions Underlying Learning

1. Bias that all features are considered relevant in the absence of experience

~~Achieved by treating ρ as a Beta random variable with imaginary-count prior~~

Encode not just the most likely value of ρ, but uncertainty distribution.

2. Environment is nonstationary

With probability λ, environment and/or task can change.

From these two claims, we have *three* free parameters total.

Qualitative performance does not depend on parameters as long as $\lambda > 0$ and $E[\rho_{i0}] < E[\rho_{i1}]$

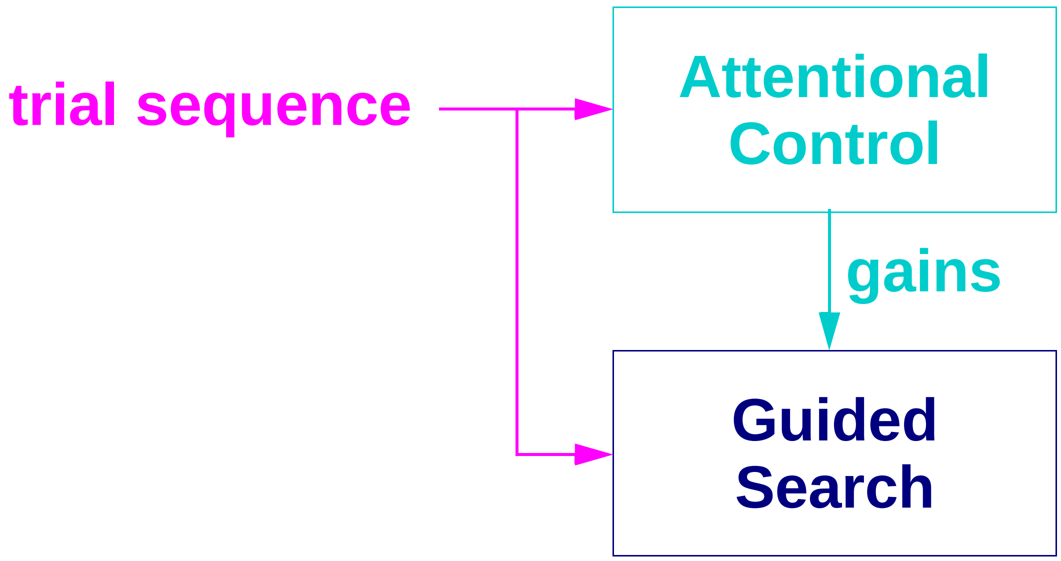
Summary So Far

Under a probabilistic generative model of the task environment, we obtain

- an expression for saliency given feature activations and task statistics
- an inference rule for updating task statistics following each trial

Three free parameters in model

- bias that all features are task relevant (2 parameters)
- environmental change probability (1 parameter)
- + a few leftover parameters of GS (e.g., RT scaling)



What It Boils Down To

What It Boils Down To

- Generate stimulus sequence corresponding to experiment.

What It Boils Down To

- **Generate stimulus sequence corresponding to experiment.**
- **Initialize task statistics**

$$\alpha_{i1} = \beta_{i0} = \varphi$$
$$\alpha_{i0} = \beta_{i1} = \theta$$

where $\bar{\rho}_{it} = \dfrac{\alpha_{it}}{(\alpha_{it} + \beta_{it})}$

What It Boils Down To

- Generate stimulus sequence corresponding to experiment.
- Initialize task statistics
- On each trial, perform feature extraction on display.

What It Boils Down To

- **Generate stimulus sequence corresponding to experiment.**
- **Initialize task statistics**
- **On each trial, perform feature extraction on display.**
- **Compute saliency at each location x**

$$s_x = \sum_i \frac{2(\rho_{i1} - \rho_{i0})}{(1-\rho_{i0})(1-\rho_{i1})}\tilde{f}_{xi} + \left[\frac{1}{\rho_{i0}(1-\rho_{i0})} - \frac{1}{\rho_{i1}(1-\rho_{i1})}\right]\tilde{f}_{xi}^2$$

What It Boils Down To

- **Generate stimulus sequence corresponding to experiment.**
- **Initialize task statistics**
- **On each trial, perform feature extraction on display.**
- **Compute saliency at each location x**
- **Determine response time based on ranking**

  $\text{ResponseTime} = \mu_0 + \mu_1 \text{ SaliencyRankingOfTarget}$

What It Boils Down To

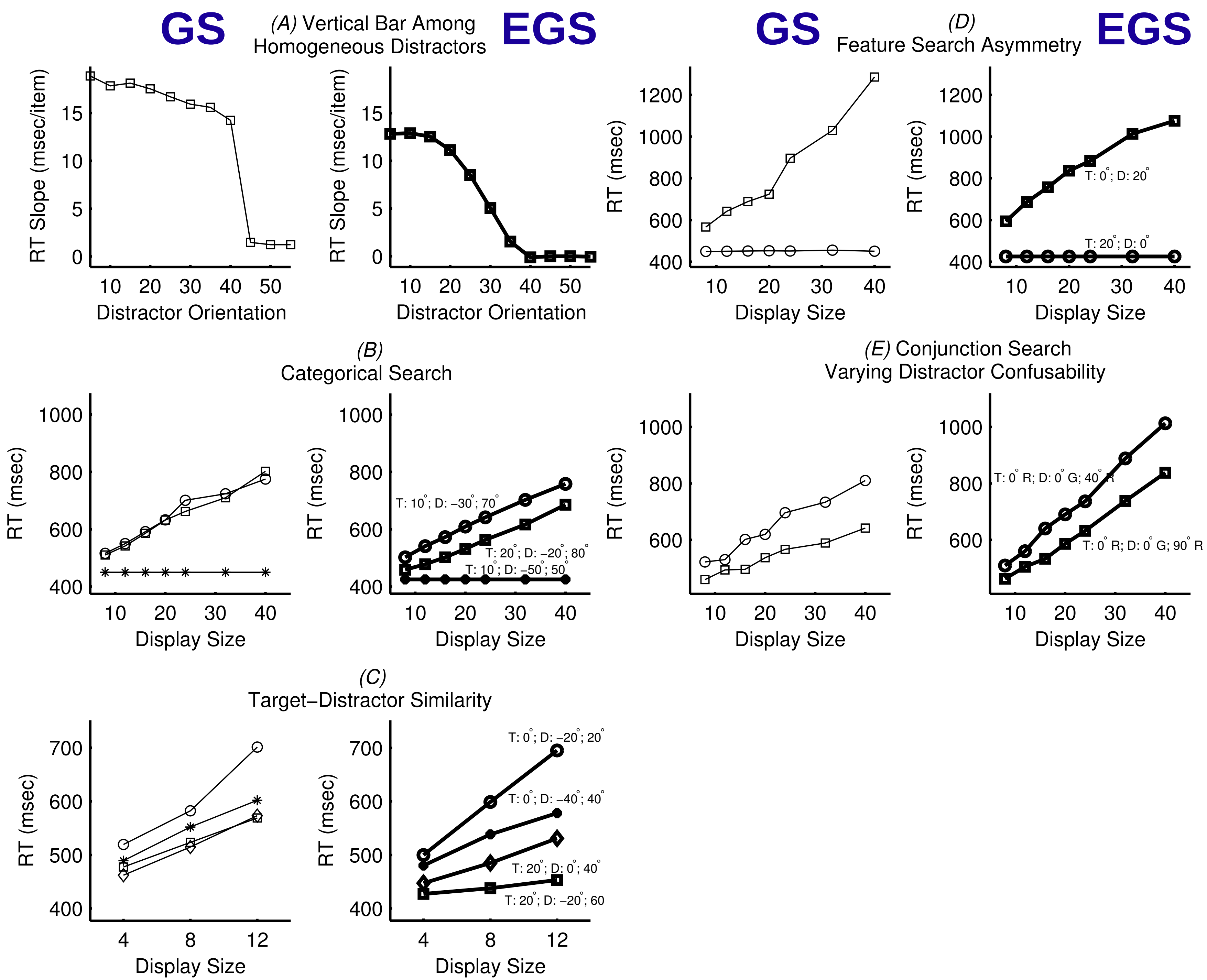
- **Generate stimulus sequence corresponding to experiment.**
- **Initialize task statistics**
- **On each trial, perform feature extraction on display.**
- **Compute saliency at each location x**
- **Determine response time based on ranking**
- **Update task statistics based on current trial feature activity**

$$\alpha_{it} \leftarrow \lambda \alpha_{it}^{0} + (1-\lambda)\left(\alpha_{it} + \sum_{x \in \chi_t} \tilde{f}_{xi}\right)$$

$$\beta_{it} \leftarrow \lambda \beta_{it}^{0} + (1-\lambda)\left(\beta_{it} + \sum_{x \in \chi_t} 1 - \tilde{f}_{xi}\right)$$

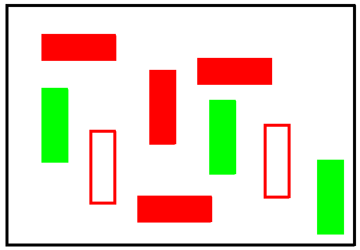
Approximate inference, but excellent approximation

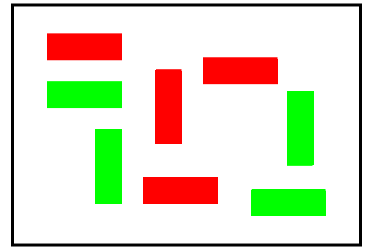
EGS Replicates GS

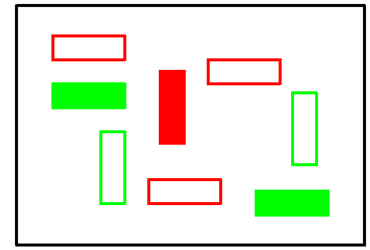


Guiding Search (Wolfe, Cave, & Franzel, 1989)

target







ratio of number of features defining target to number of features shared with distractors

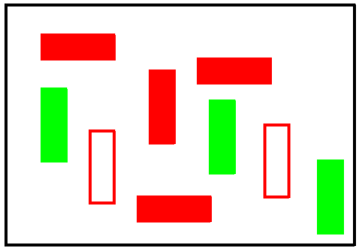
3:2

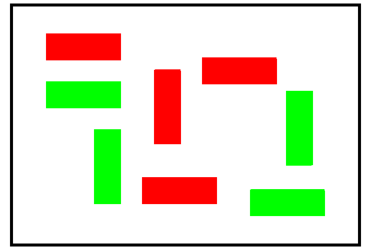
2:1

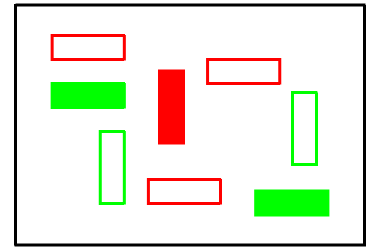
3:1

Guiding Search (Wolfe, Cave, & Franzel, 1989)

target





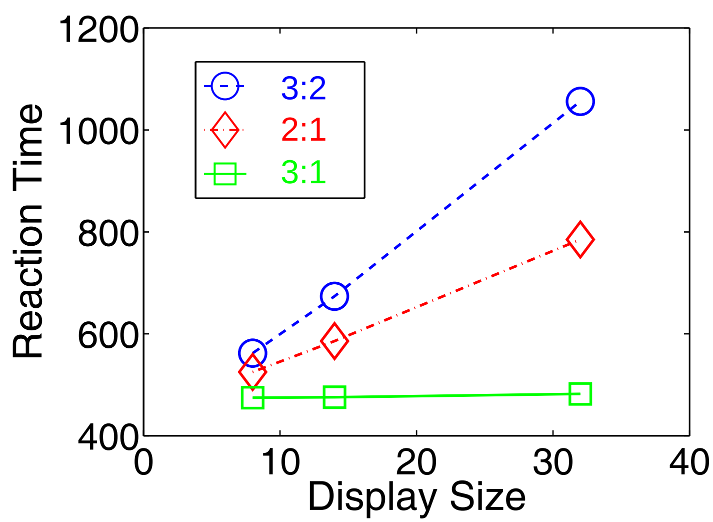


ratio of number of features defining target to number of features shared with distractors

3:2

2:1

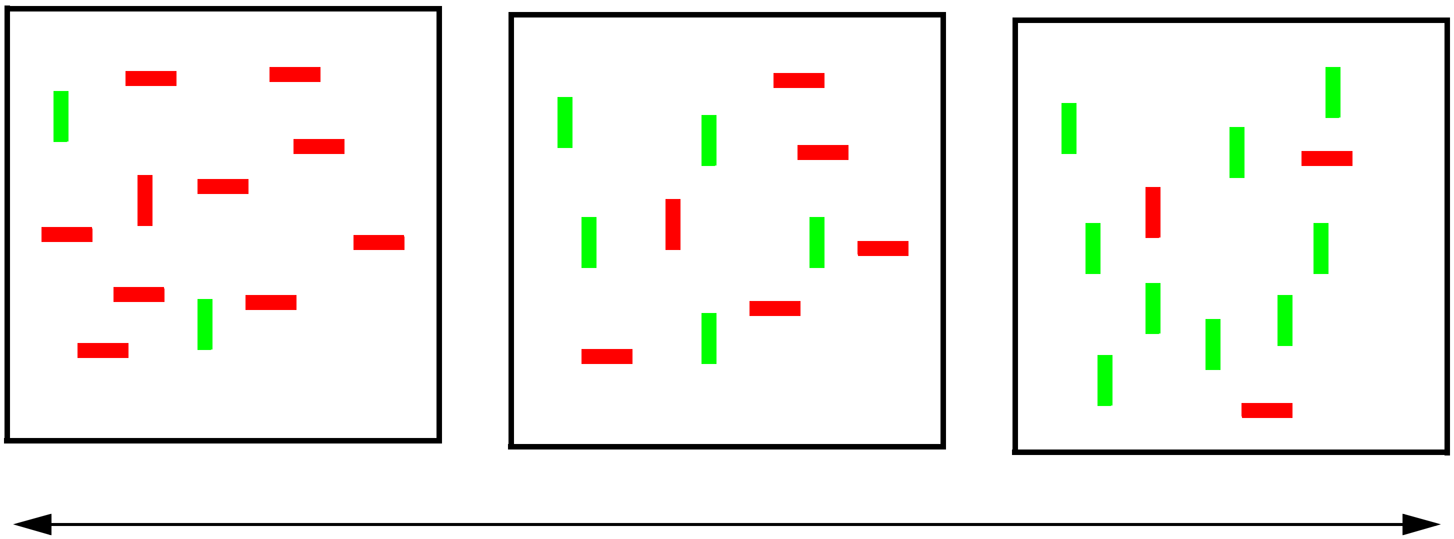
3:1



cf. Shiffrin talk

Varying Distractor Proportion

(Egeth, Virzi, & Garbart, 1984; Poisson & Wilkinson, 1992; Zohary & Hochstein, 1989)



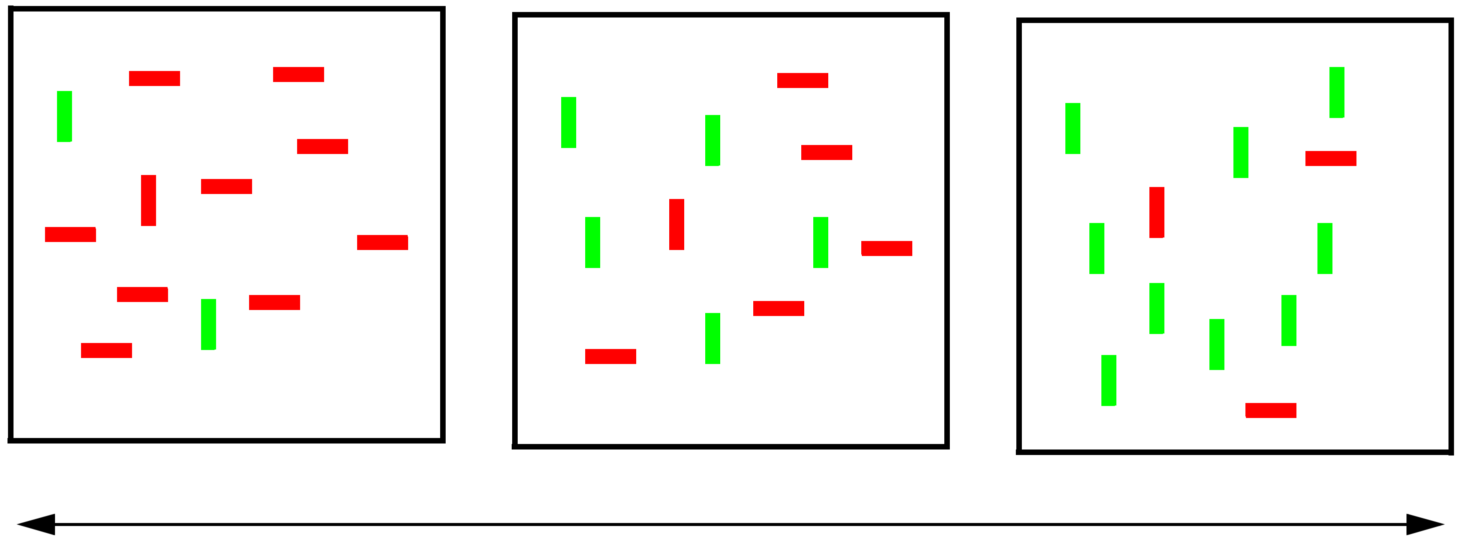
mostly red distractors

equal number red and vertical

mostly vertical distractors

Varying Distractor Proportion

(Egeth, Virzi, & Garbart, 1984; Poisson & Wilkinson, 1992; Zohary & Hochstein, 1989)

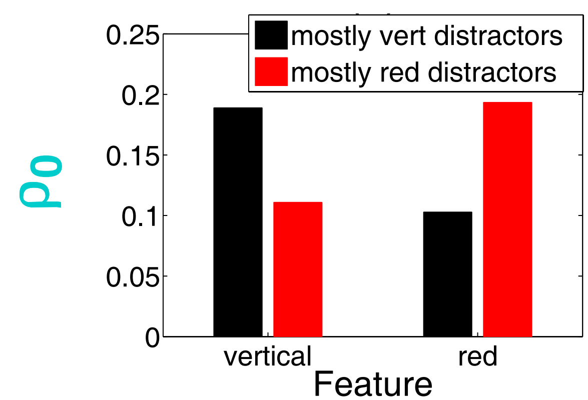


mostly red distractors | equal number red and vertical | mostly vertical distractors

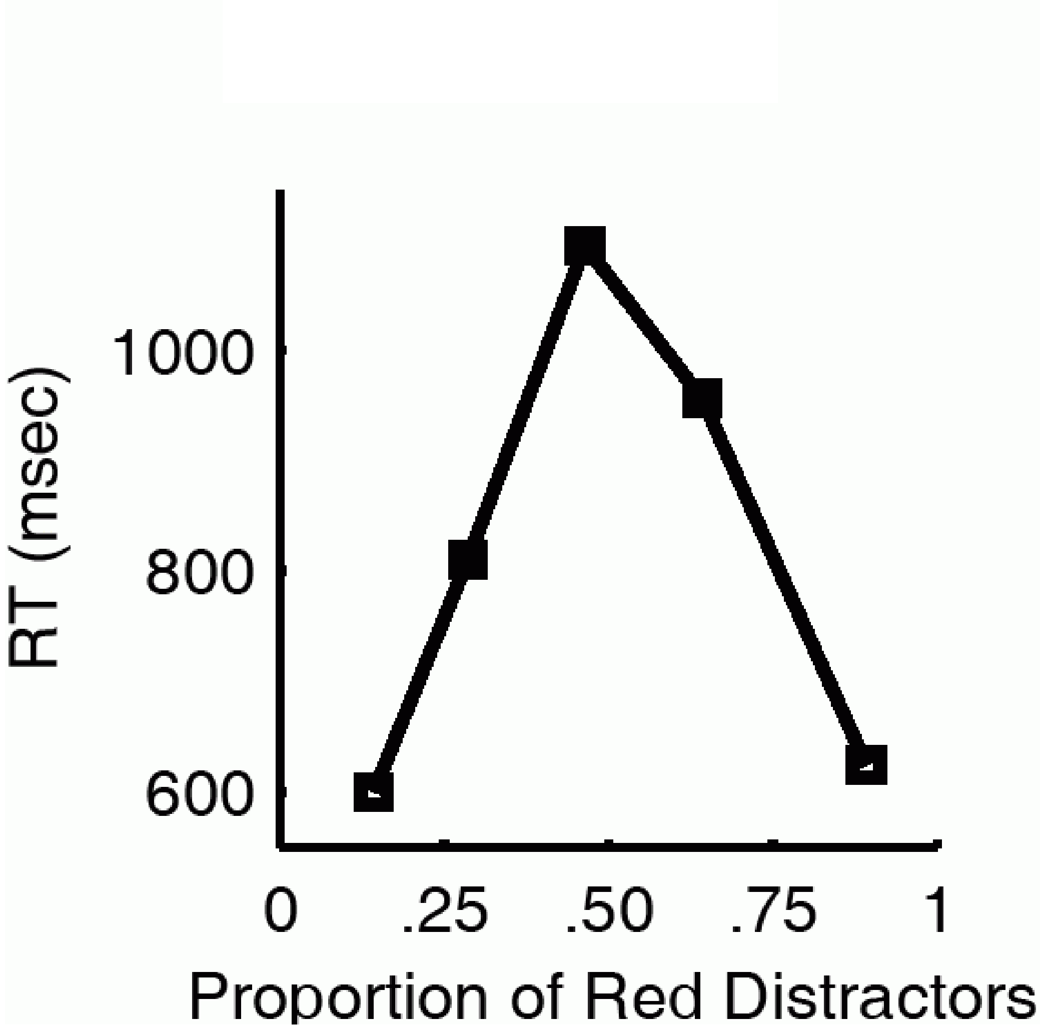
Blocked trials ->

environment statistics make one feature a more discriminable cue ->

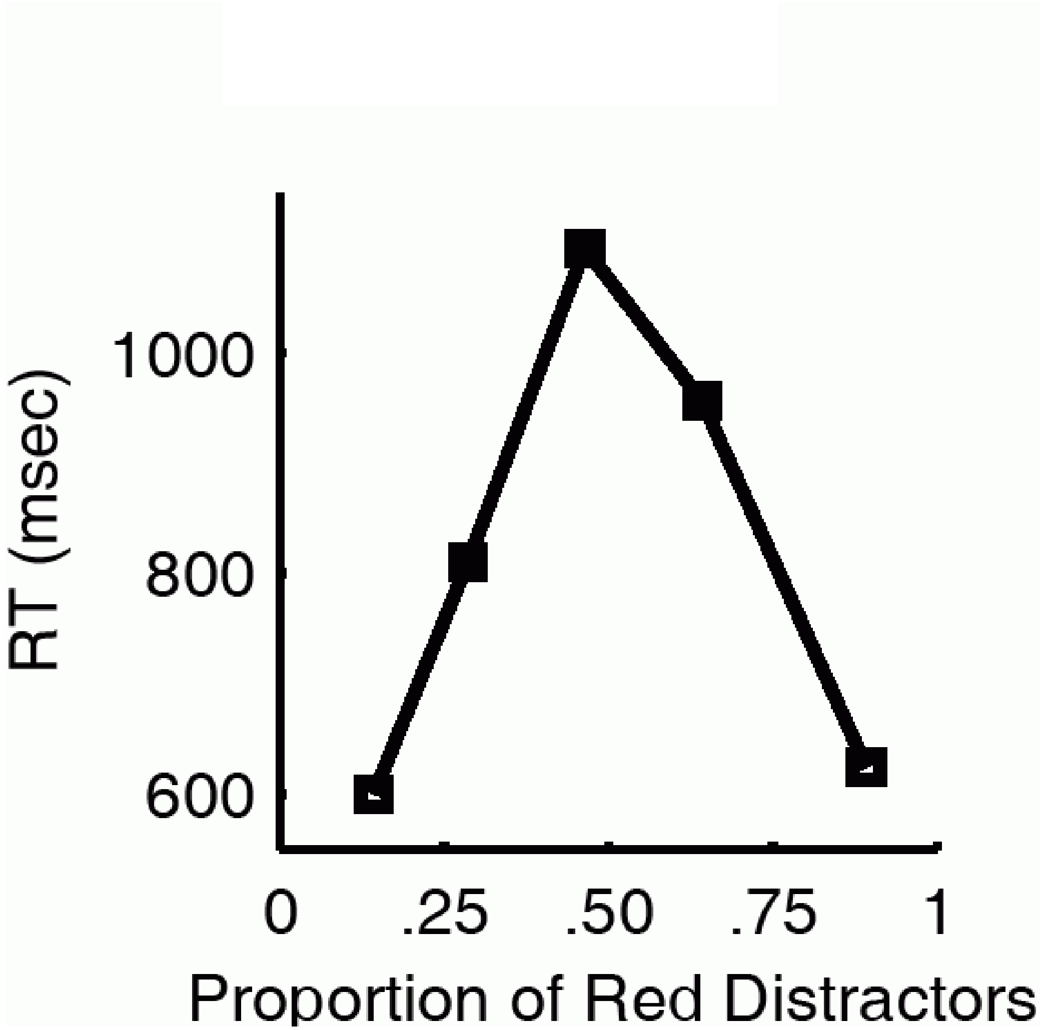
EGS gives it greater weighting.



Varying Distractor Proportion



Varying Distractor Proportion



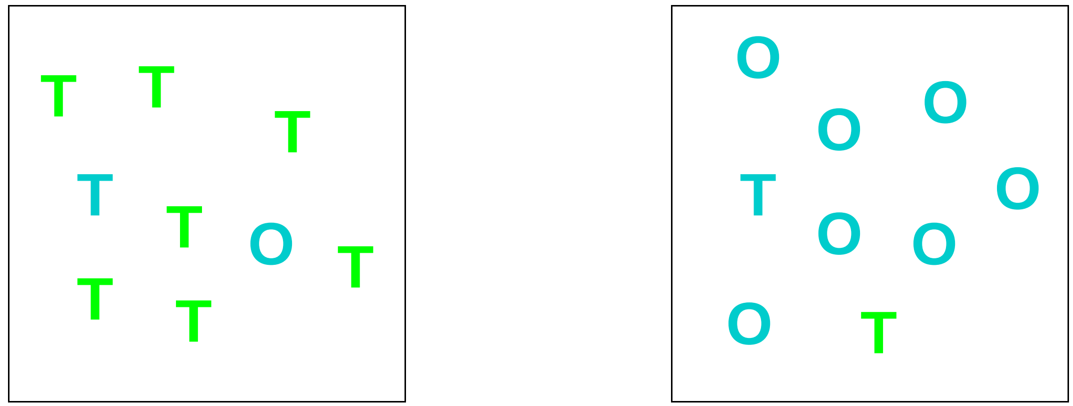
In these studies, distractor proportion is blocked.

Is efficiency achieved when proportion varies within block?

Wright and Main (unpublished)

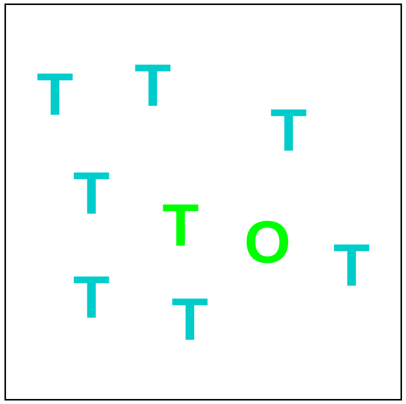
Wright and Main (unpublished)

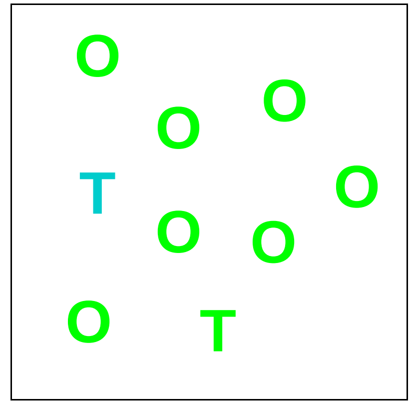
Search for T among T and O



Wright and Main (unpublished)

Search for T among T and O

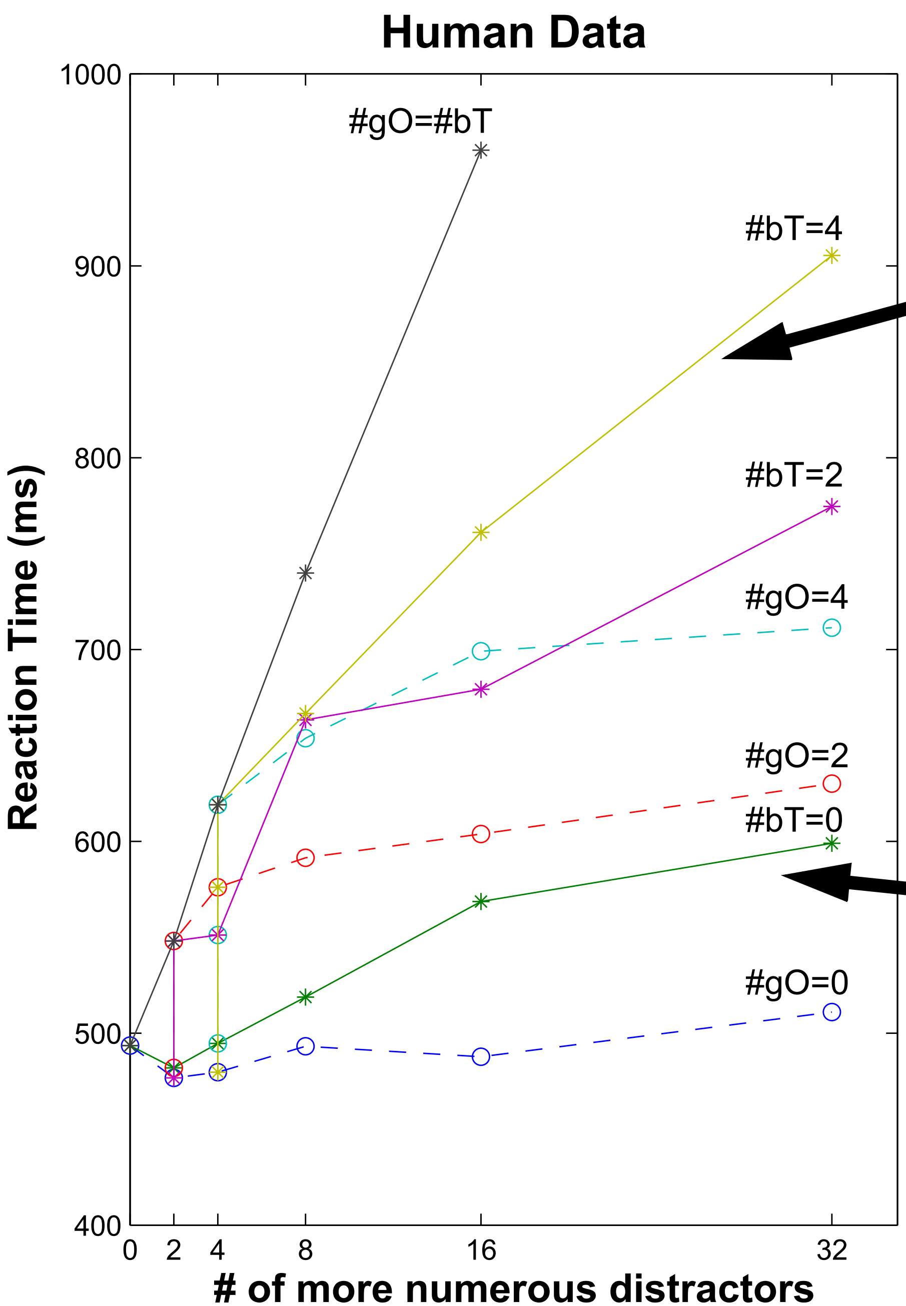




Number of distractors of each type varies trial to trial

Number of O	Number of T: 0	2	4	8	16	32
0	X	X	X	X	X	X
2	X	X	X	X	X	X
4	X	X	X	X	X	X
8	X	X	X	X		
16	X	X	X		X	
32	X	X	X			

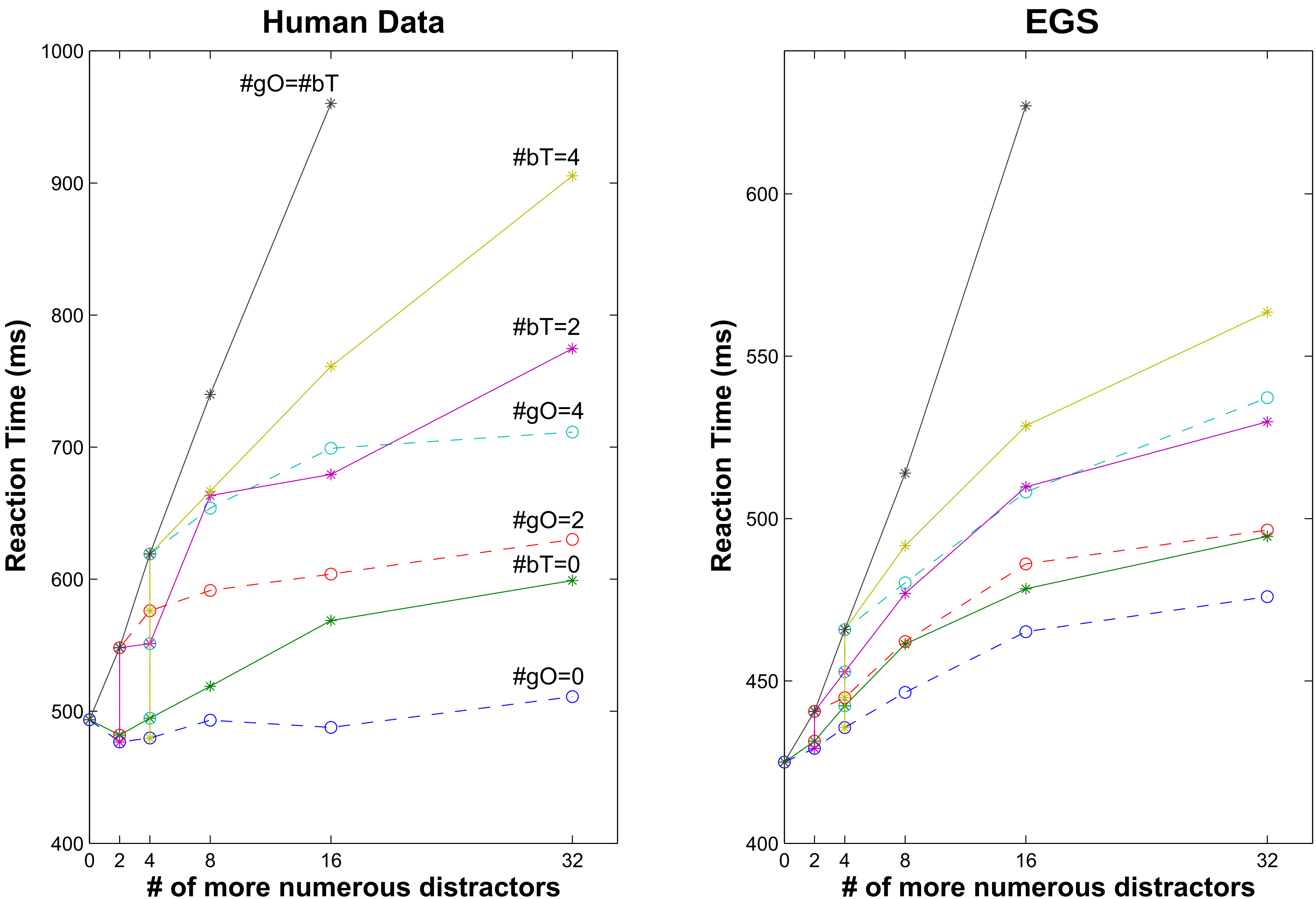
Wright and Main (unpublished)



inefficient exploitation of distractor ratio

pop out display!

Wright and Main (unpublished)



EGS makes strong predictions about intertrial priming.

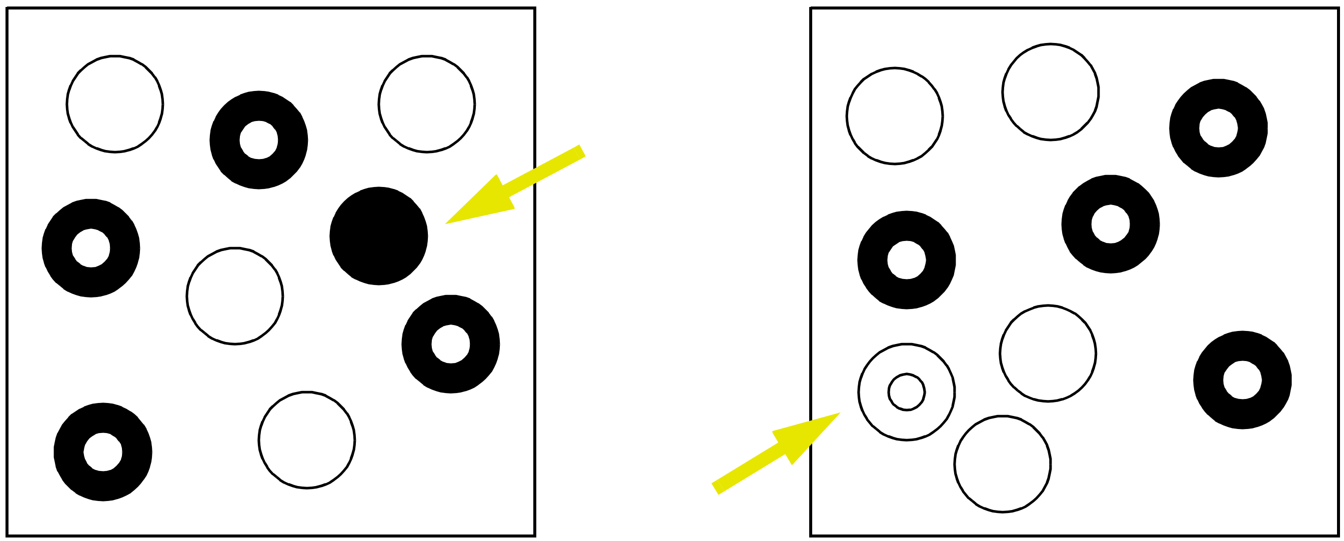
Attentional gains on current trial depends on statistics of recent trials.

Had originally hoped to use Wright data to look at how performance on trial *n* influenced by trial *n–1.*

Insufficient data

Kristjánsson and Driver (2008)

Intertrial priming with oddball conjunction search

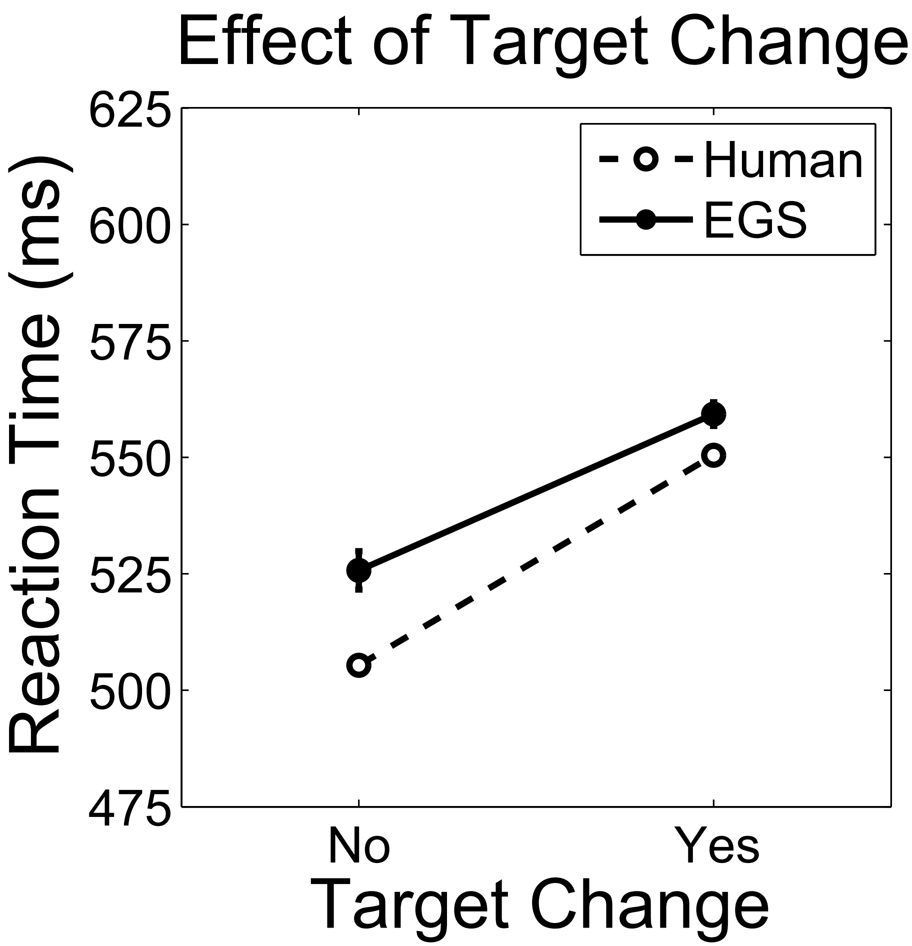


target(t–1)
distractors(t–1)

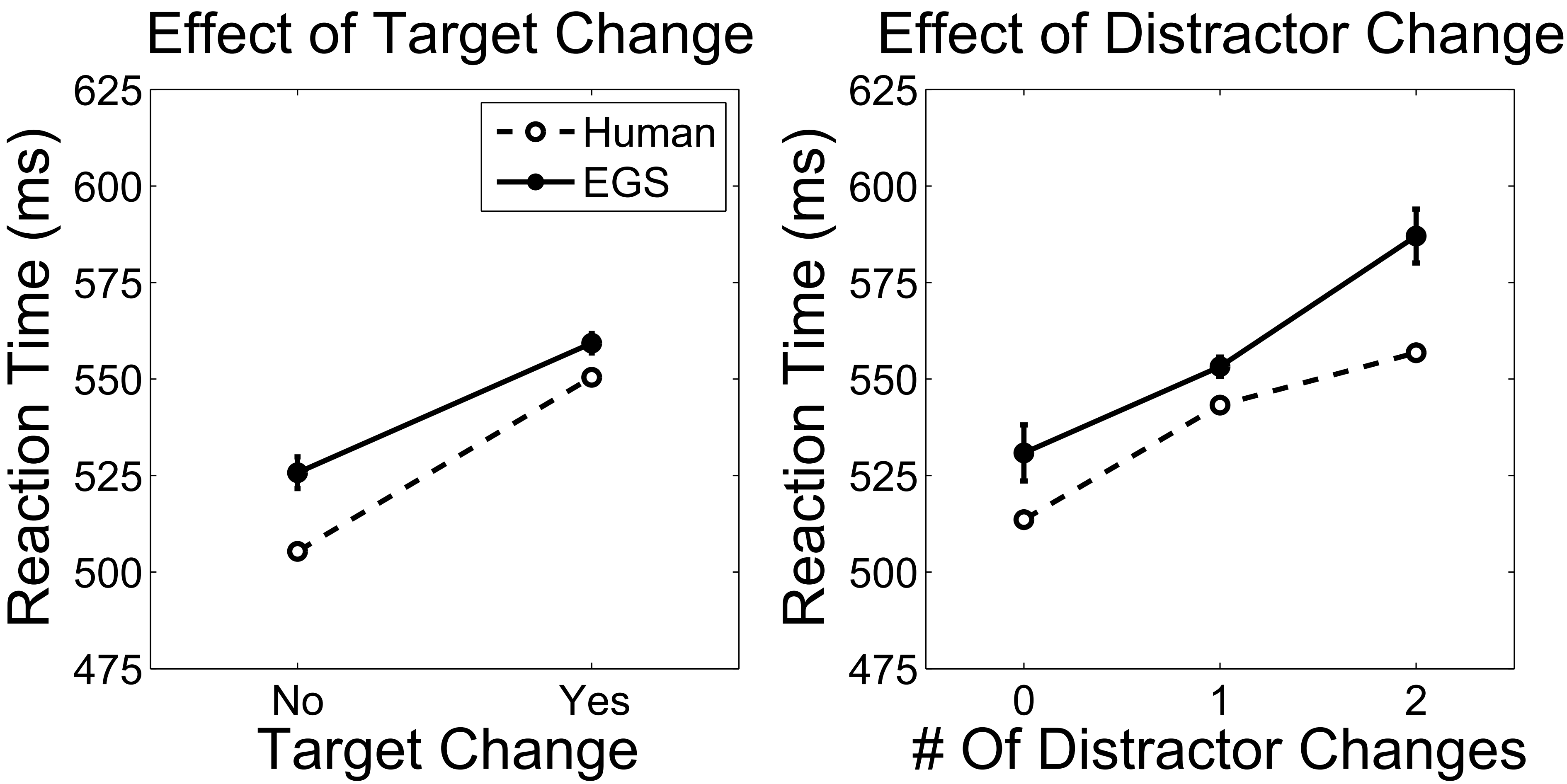
target(t)
distractors(t)

Sequential effects in RT based on relationship of target and distractor on trial *t–1* to target and distractor on trial *t*

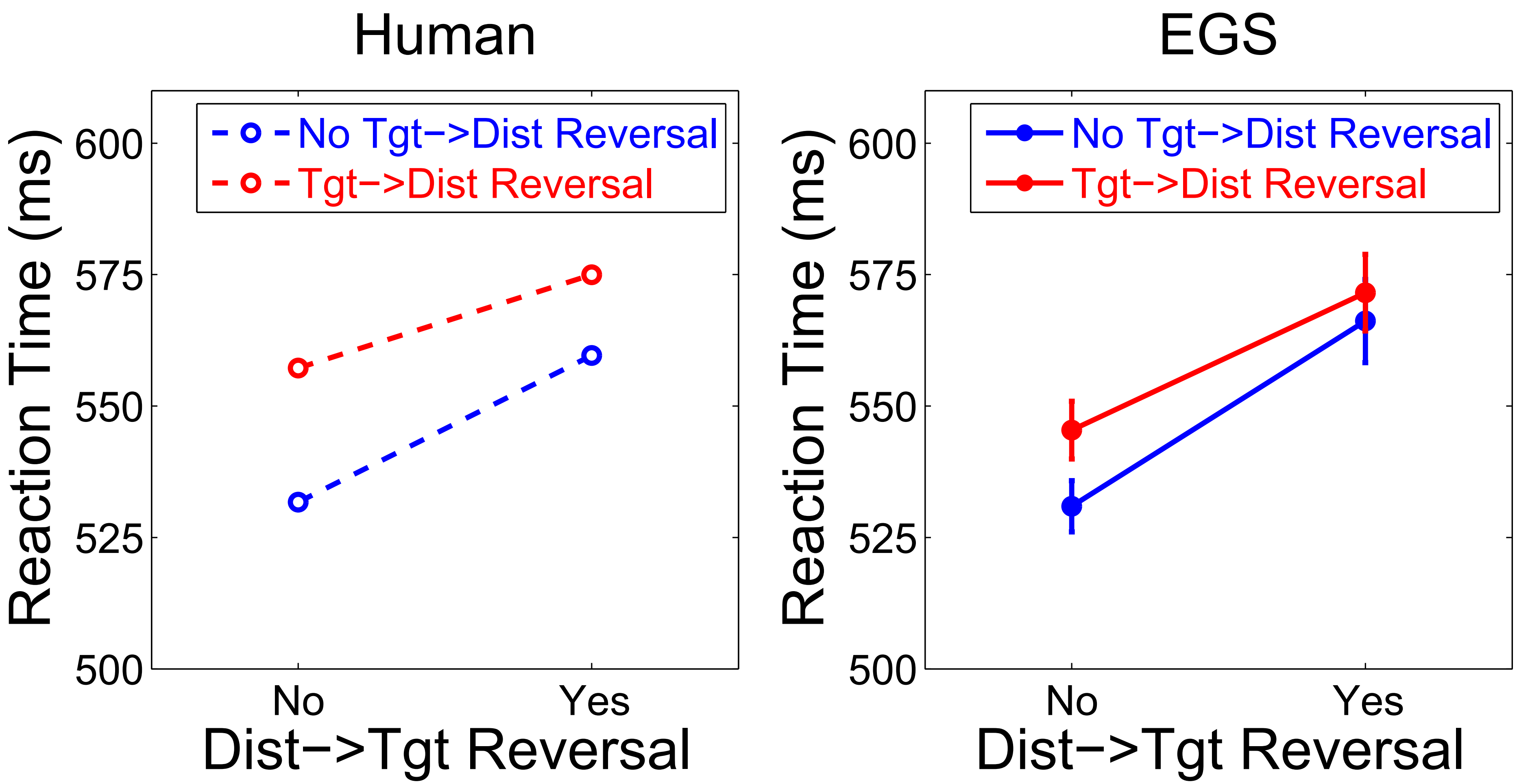
Kristjánsson and Driver (2008)



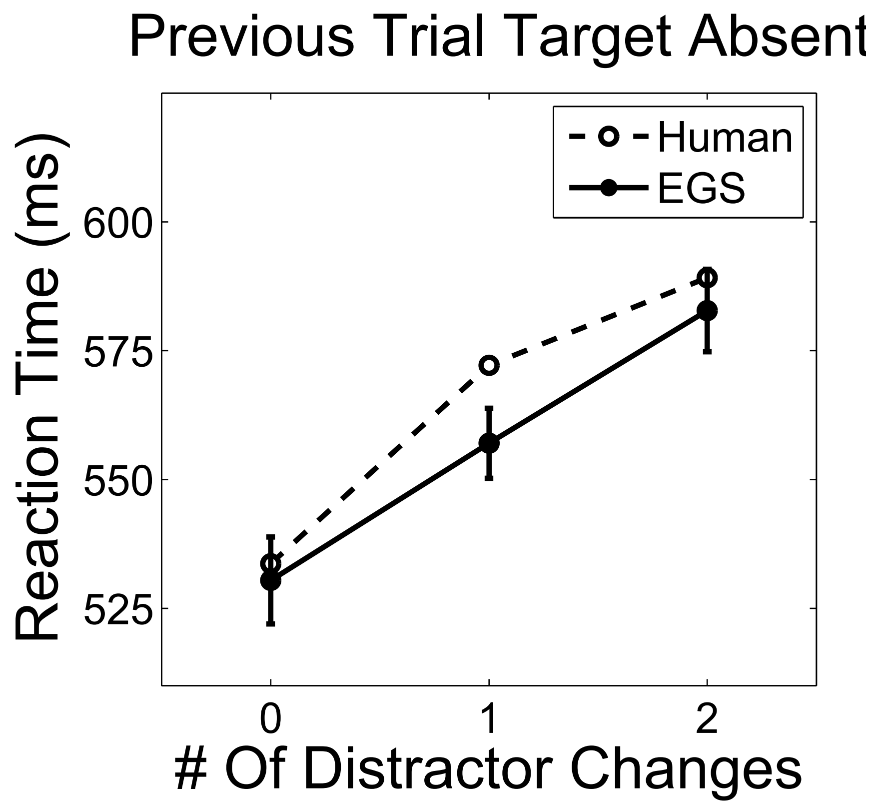
Kristjánsson and Driver (2008)



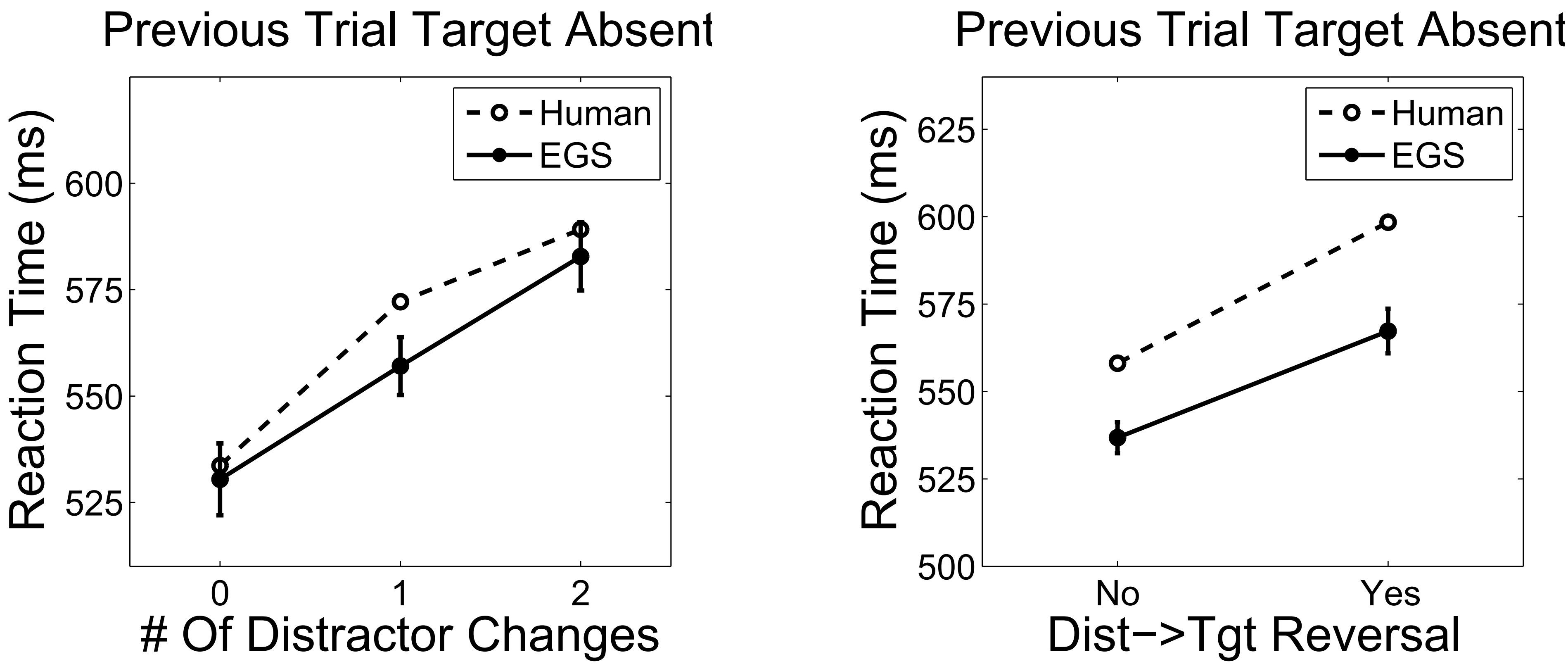
Kristjánsson and Driver (2008)



Kristjánsson and Driver (2008)

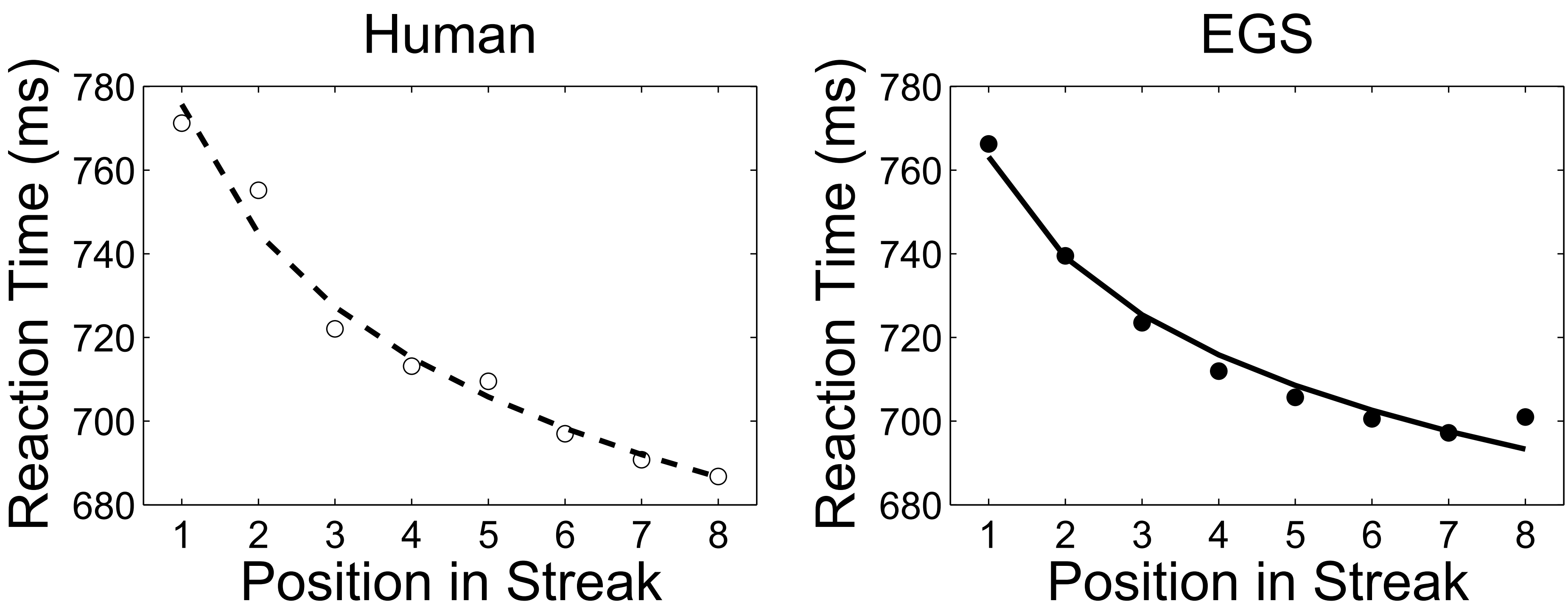


Kristjánsson and Driver (2008)

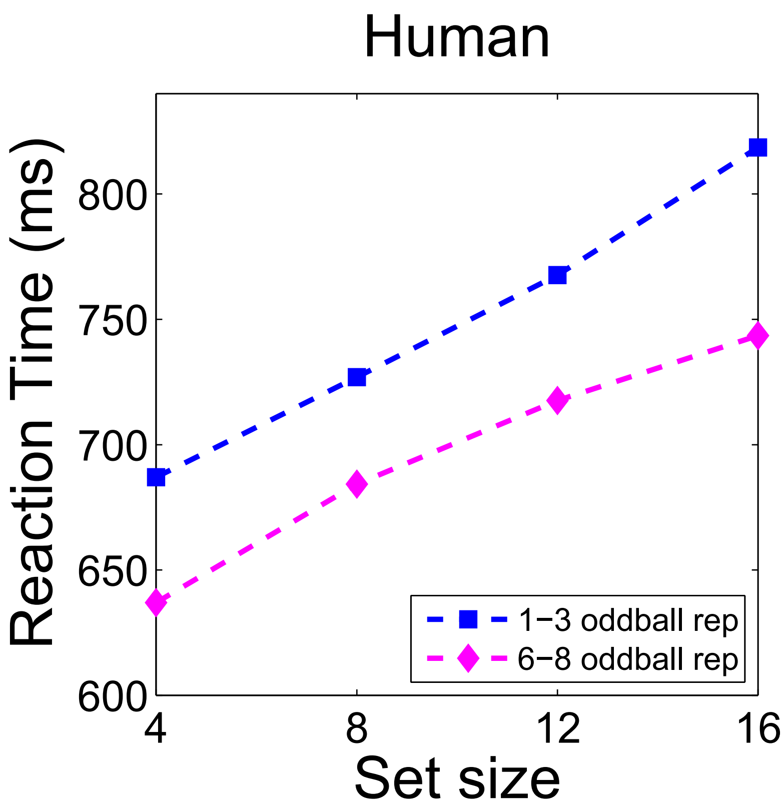


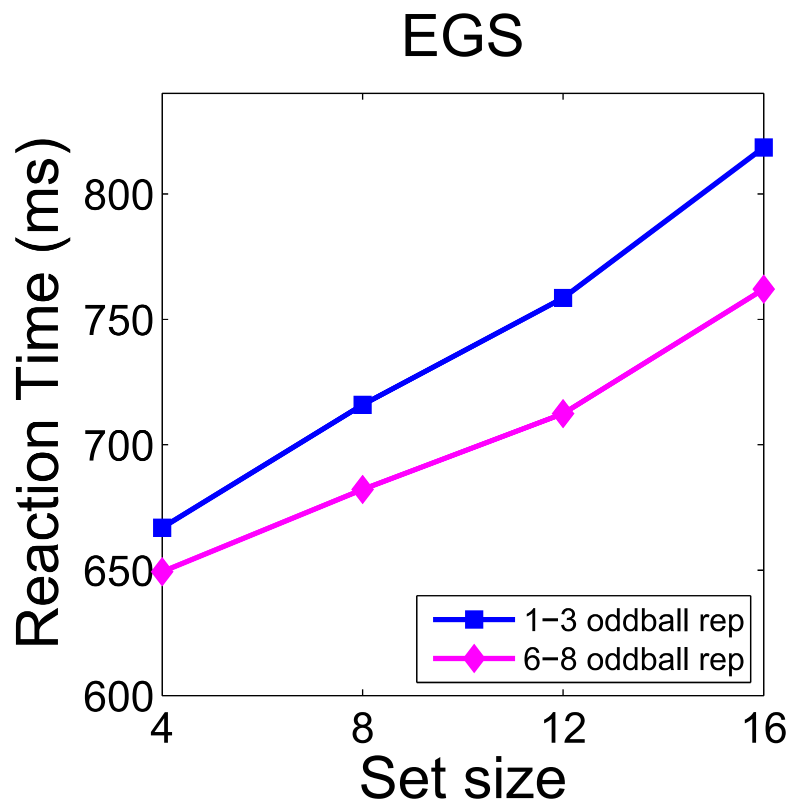
Kristjánsson, Wang, and Nakayama (2002)

Similar oddball study with streaks of repeated target/distractor combinations

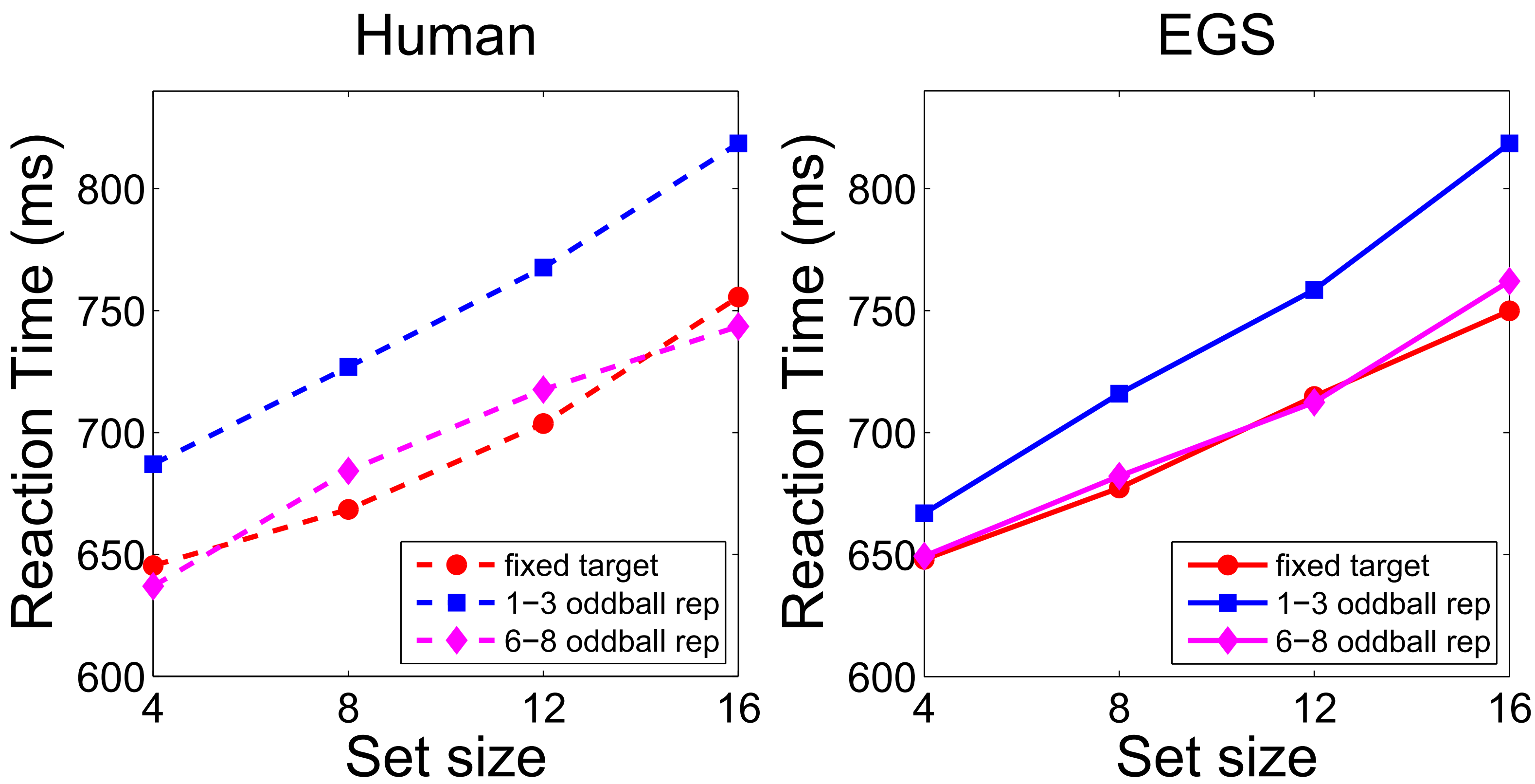


Kristjánsson, Wang, and Nakayama (2002)



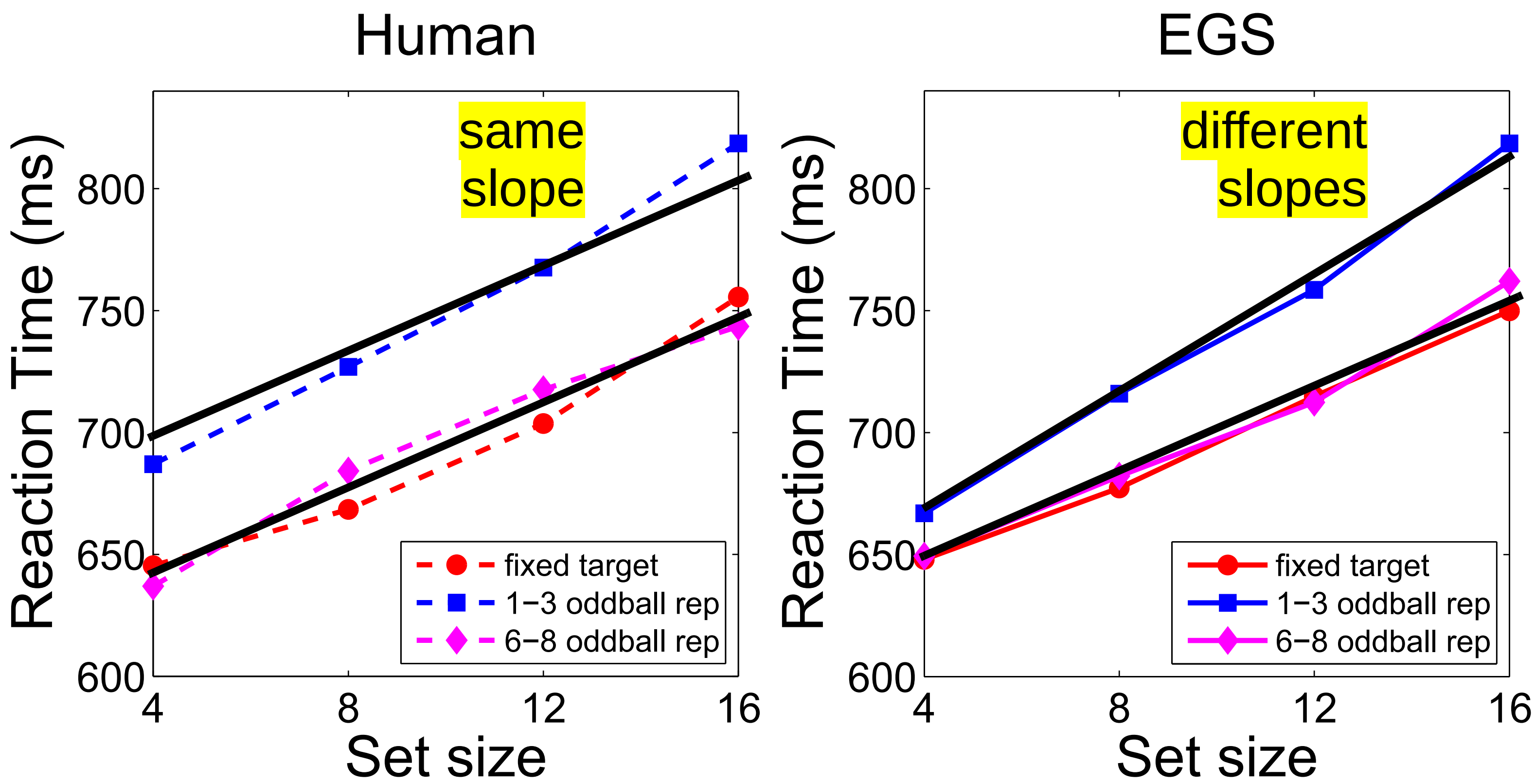


Kristjánsson, Wang, and Nakayama (2002)



Control processes involved in specifying fixed target are qualitatively the same as those involved in sequential effects.

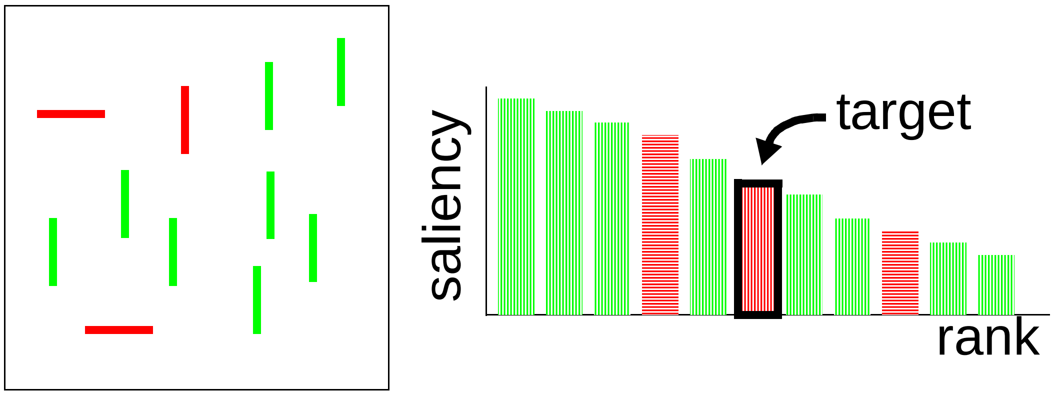
Modeler Sleaze



Other findings in literature that priming does not affect search slopes (Kunar et al., 2007; Lamy, 2012; Mozer & Roads, in preparation).

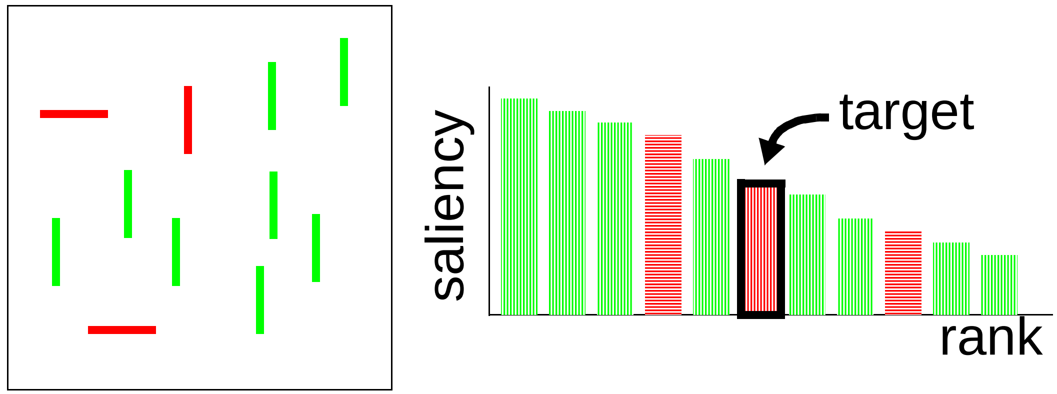
Long Story Short

EGS with serial search prioritized by saliency rank
⇒ search-slope reduction.



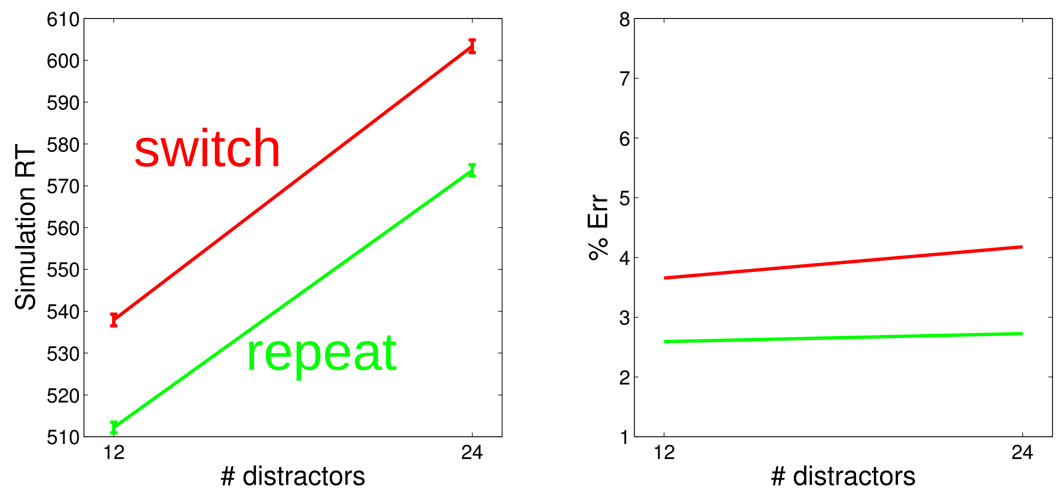
Long Story Short

EGS with serial search
prioritized by saliency rank
⇒ search-slope reduction.



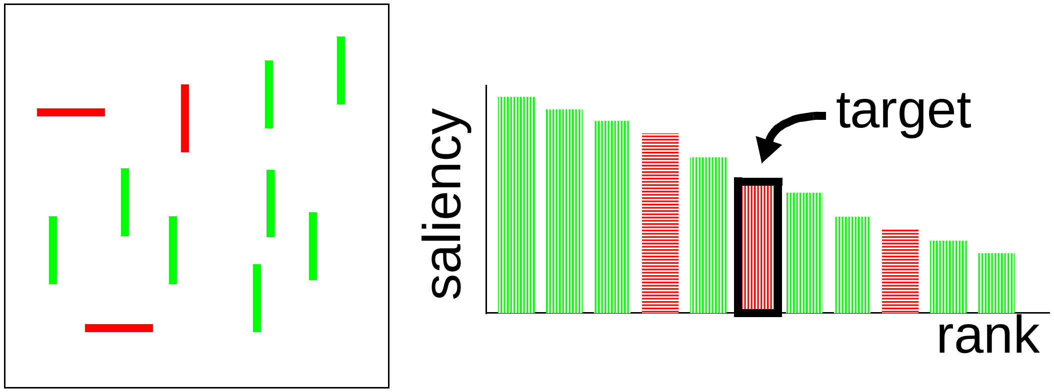
EGS with stochastic selection in which elements compete proportional to saliency
⇒ no search-slope reduction.

e.g., one-step accumulator model



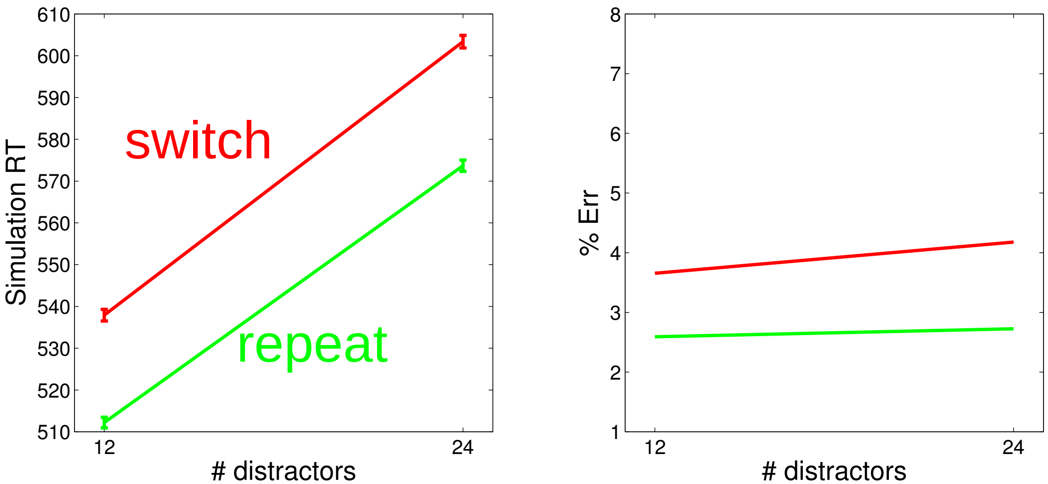
Long Story Short

EGS with serial search prioritized by saliency rank ⇒ search-slope reduction.



EGS with stochastic selection in which elements compete proportional to saliency ⇒ no search-slope reduction.

e.g., one-step accumulator model



Lamy talk

"no interaction ⇒ intertrial priming does not affect attentional prioritization"

Modeling provides existence proof otherwise.

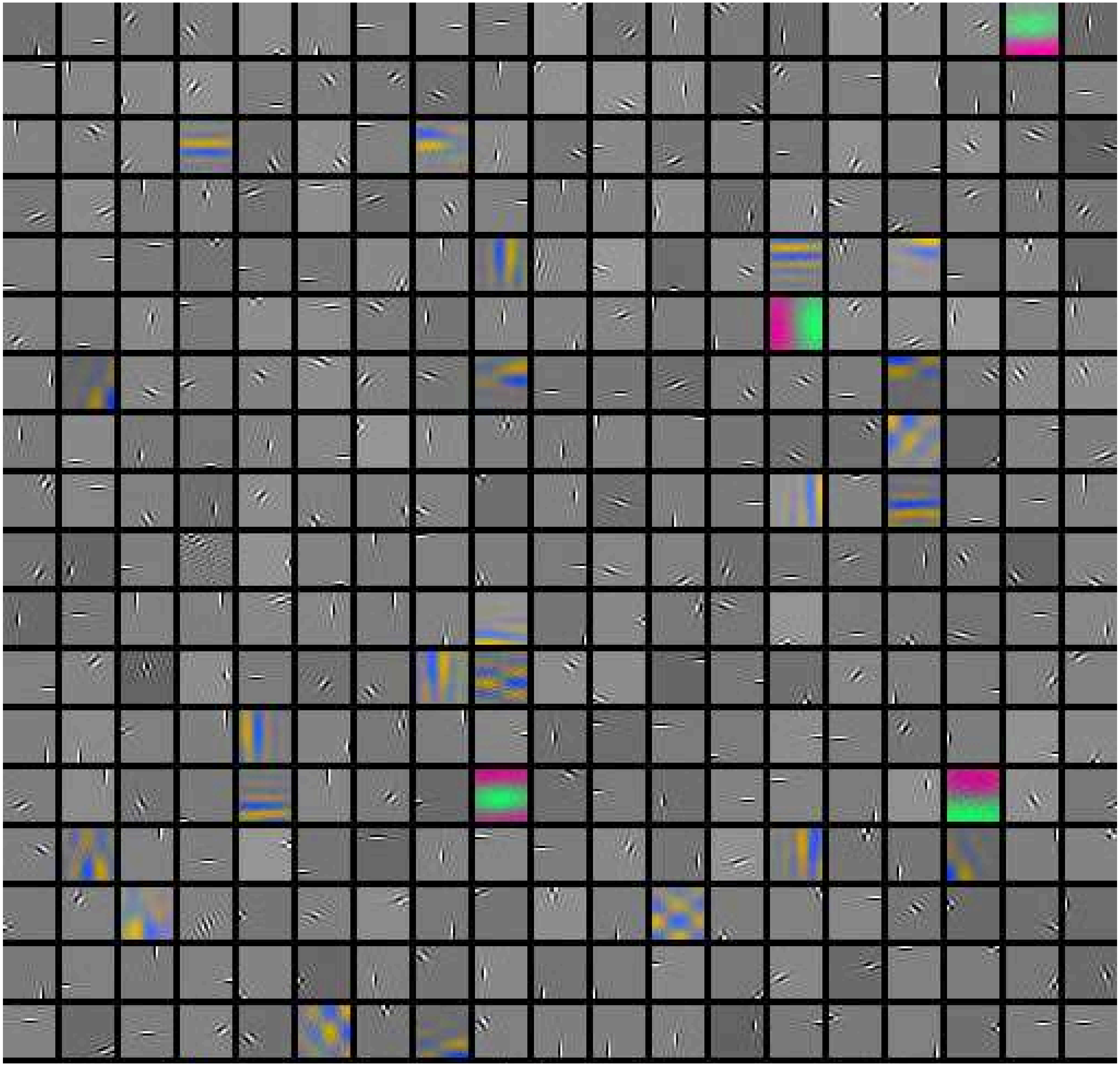
Can We Use EGS To Understand Learning At Longer Time Scales?

Initial effort with

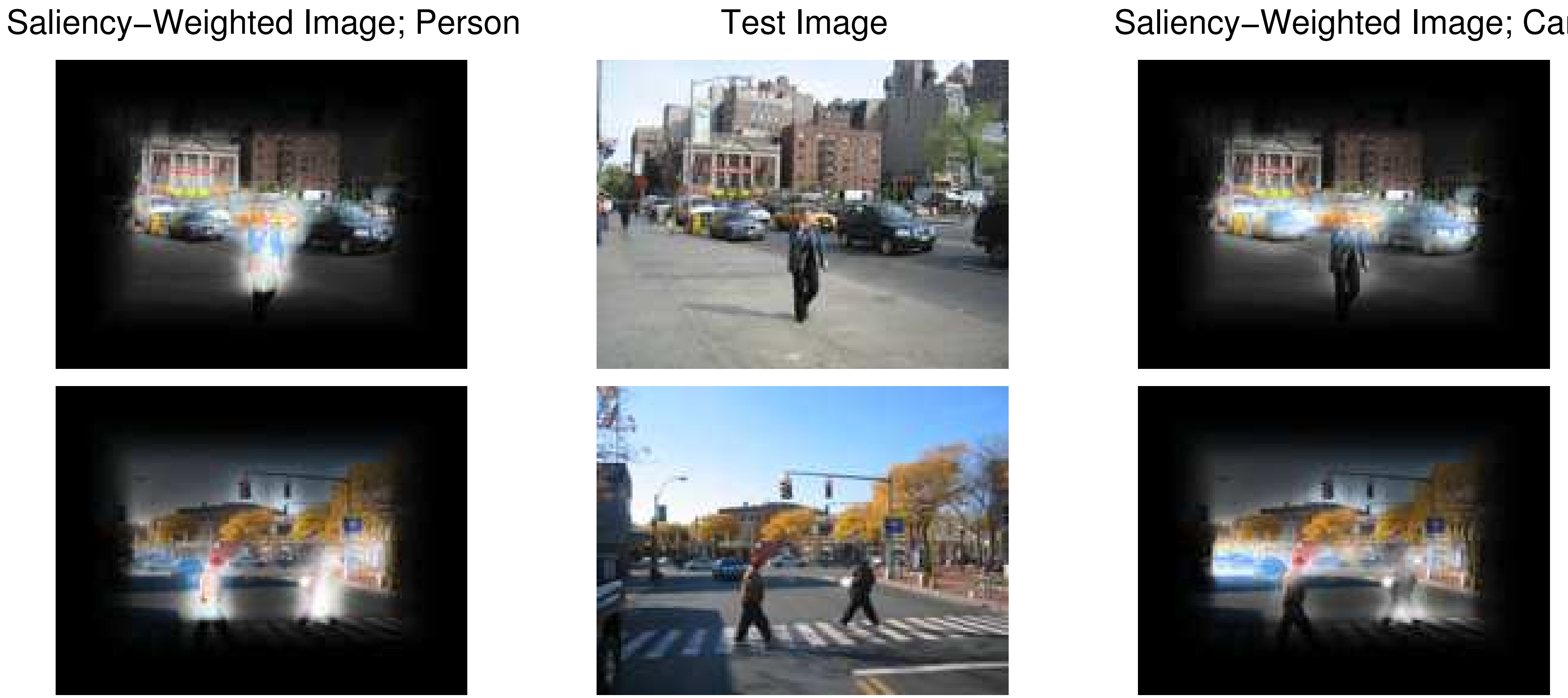
- real-world images (MIT LabelMe data base)
- long-term learning to find objects in images

Exactly same EGS model with

- much lower environment switch probability (~ memory decay)
- image based features (ICA/PCA)



Saliency−Weighted Image: Person | Test Image | Saliency−Weighted Image; Car



Wolfe’s theme

“Beyond feature guidance”

Mozer’s theme

How far can feature guidance take us—can we avoid “syntax”, “semantics”, “scene gist”, etc.? (Similar philosophy as Zelinsky)

In progress: still need to compare to human fixation data, other models

What Does EGS Offer?

Formalization of the intuition we share about attentional gains

What Does EGS Offer?

Formalization of the intuition we share about attentional gains

A computational-level theory of visual search

Goal of attention: target detection $\Longrightarrow$ saliency: target probability

Control emerges as a consequence of statistical inference on the task environment in which an individual operates.

Adaptation (at every time scale) results from updating beliefs about task environment.

What Does EGS Offer?

Formalization of the intuition we share about attentional gains

A computational-level theory of visual search

A perspective on attentional control

Intertrial priming effects have been viewed as either

- a passive, bottom up influence on behavior (e.g., Maljkovic & Nakayama, 1994; Pinto, Olivers, & Theeuwes, 2005)
- driven by or modulated by implicit top-down control processes (e.g., Guyer & Müller, 2009; Wolfe, Butcher, Lee, & Hyle, 2003)

What Does EGS Offer?

Formalization of the intuition we share about attentional gains

A computational-level theory of visual search

A perspective on attentional control

Intertrial priming effects have been viewed as either

- a passive, bottom up influence on behavior (e.g., Maljkovic & Nakayama, 1994; Pinto, Olivers, & Theeuwes, 2005)
- driven by or modulated by implicit top-down control processes (e.g., Guyer & Müller, 2009; Wolfe, Butcher, Lee, & Hyle, 2003)

What Does EGS Offer?

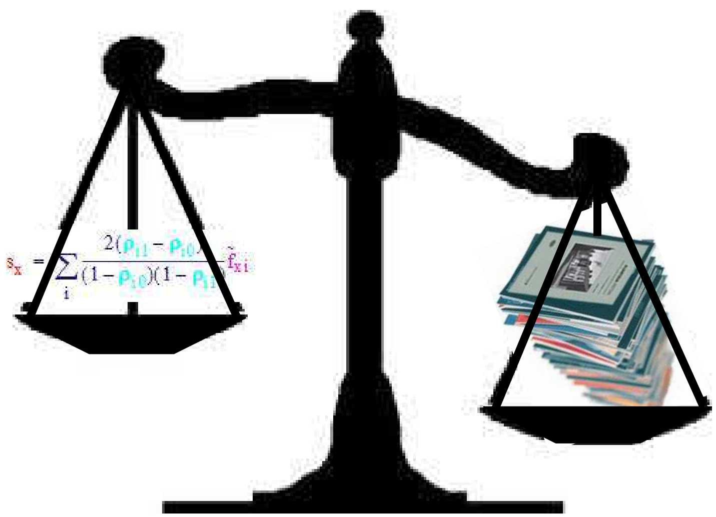
Formalization of the intuition we share about attentional gains

A computational-level theory of visual search

A perspective on attentional control

Learning is the key to a unified, elegant theory of search.

$\Rightarrow$ parsimonious explanations of data from many different paradigms and tasks with the same principles.



STOP HERE MIKE!

Search For Die Bohne

Even when you recognize the object name...

there may be variability in the object’s appearance.

And even when you have a good visual representation of the target...

distractor statistics matter.

Control Of Attention Depends On Learning Statistical Regularities On Multiple Time Scales

Time Scale	What Is Learned	Example
Coarse	Visual invariants	word ⇒ visual features
Intermediate	Structure of environments	reward-associated features (Anderson); object categories (Shiffrin)
Fine	Properties of the immediate environment	white coffee cups; intertrial priming (Müller, Lamy, Becker)

My hunch

If we successfully understand adaptation on the finest time scale, we'll be able to handle coarser time scales as well.

(Same learning mechanisms, different decay constants)

Assumptions Of Environment Model

1. **Feature detectors are conditionally independent of one another.**

$$P(\mathbf{F}_x | T_x, \rho) = \prod_i P(F_{xi} | T_x, \rho)$$

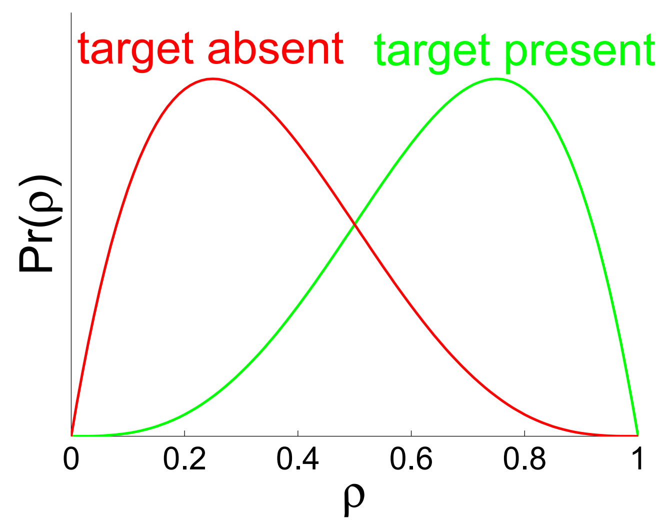
Assumptions Of Environment Model

1. **Feature detectors are conditionally independent of one another.**

$$P(\mathbf{F}_x | T_x, \rho) = \prod_i P(F_{xi} | T_x, \rho)$$

2. **Feature detectors are rate-coded neurons with firing rate ρ_{it}.**

ρ_{it} is unknown but learned through experience.



Assumptions Of Environment Model

1. **Feature detectors are conditionally independent of one another.**

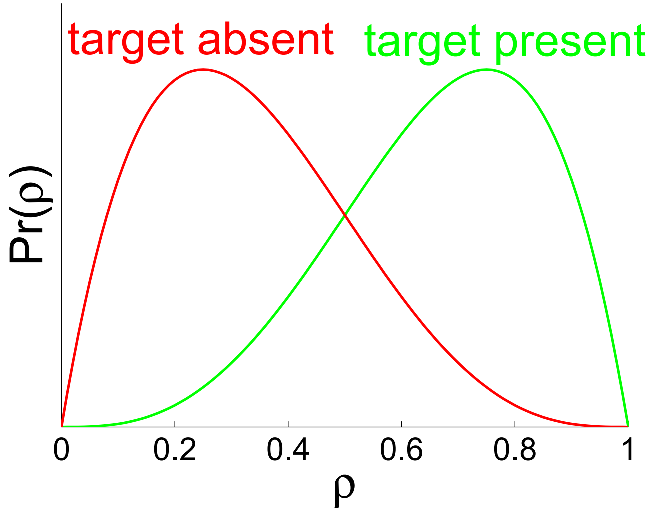
$$P(\mathbf{F}_x | T_x, \rho) = \prod_i P(F_{xi} | T_x, \rho)$$

2. **Feature detectors are rate-coded neurons with firing rate ρ_{it}.**

ρ_{it} is unknown but learned through experience.

3. **Prior to experience, all features are considered relevant.**

ρ_{i1} initialized to be greater than ρ_{i0}



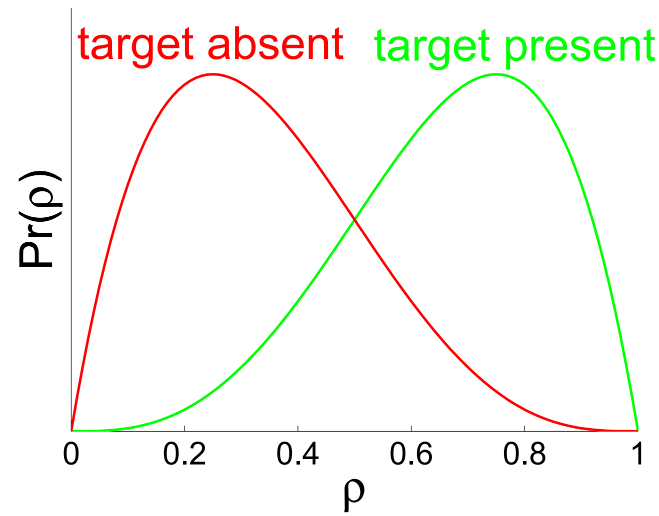
Assumptions Of Environment Model

1. **Feature detectors are conditionally independent of one another.**

$$P(\mathbf{F}_x | T_x, \rho) = \prod_i P(F_{xi} | T_x, \rho)$$

2. **Feature detectors are rate-coded neurons with firing rate ρ_{it}.**

ρ_{it} is unknown but learned through experience.



3. **Prior to experience, all features are considered relevant.**

ρ_{i1} initialized to be greater than ρ_{i0}

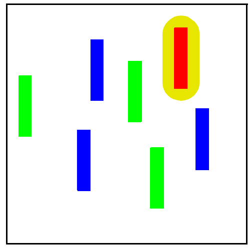
4. **Environment is nonstationary.**

With probability λ, environment and/or task can change.

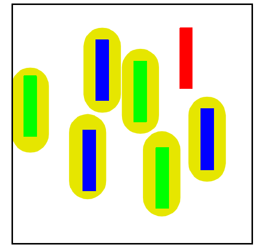
How Are Environment Statistics (ρ) Estimated?

As locations are inspected during a trial, a supervisory process labels each element as target or nontarget.

target locations



nontarget locations



Given these observations, update the ρ_{it} via Bayesian inference.

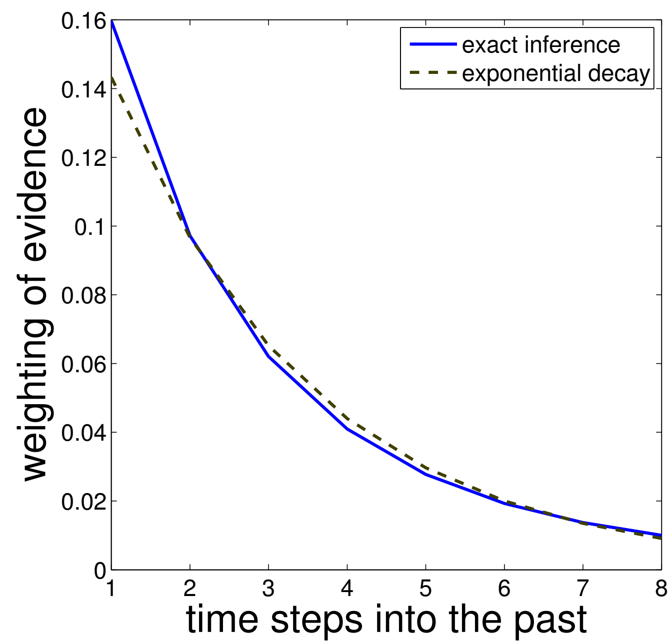
Exact inference

ρ_{it} is a mixture-of-Betas RV with mixture length linear in number time steps

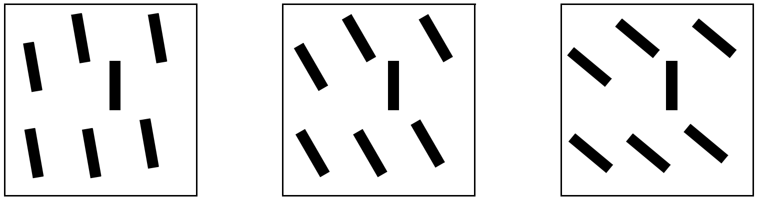
easy to simulate, hard to conceive of as biologically plausible

Approximately...

update a decaying average of each ρ_{it} based on the new observations

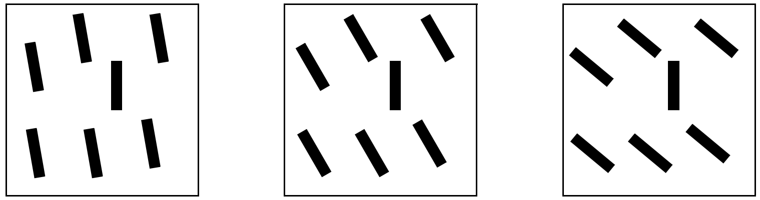


Target-Distractor Similarity Effects 1

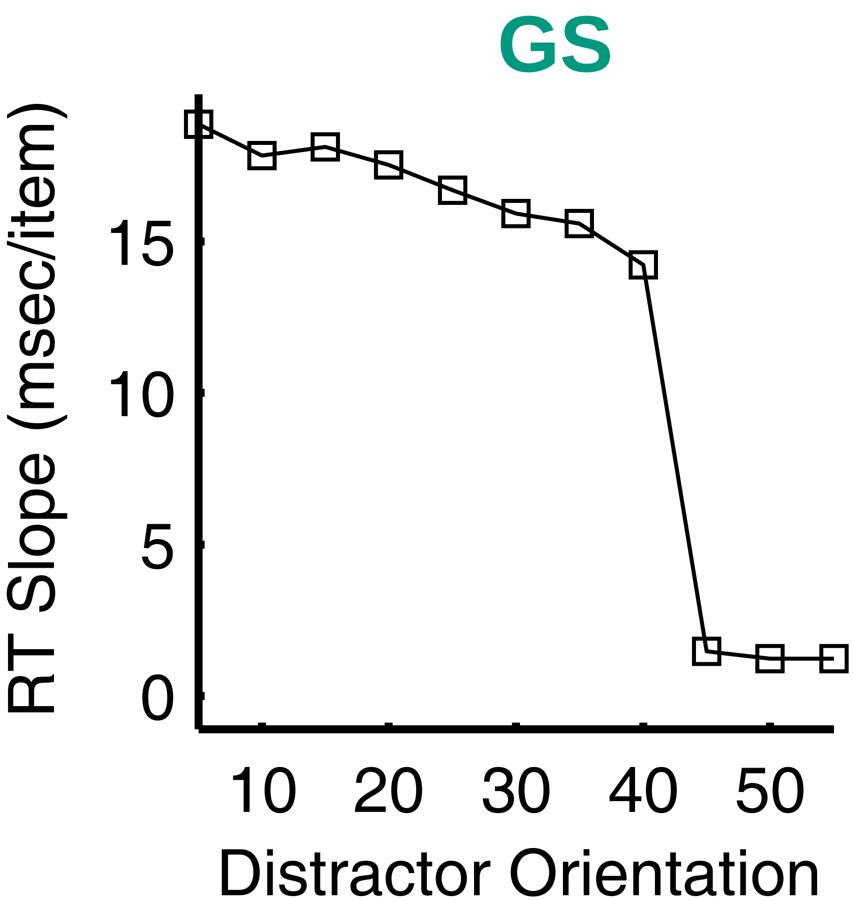


hard -------------------> easy

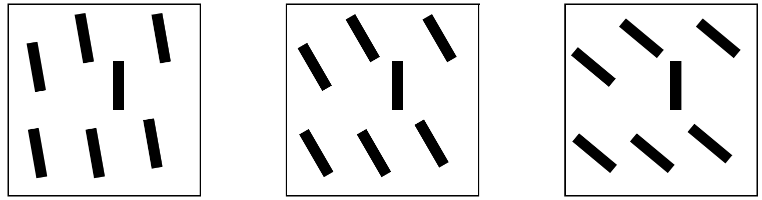
Target-Distractor Similarity Effects 1



hard -------------------> easy

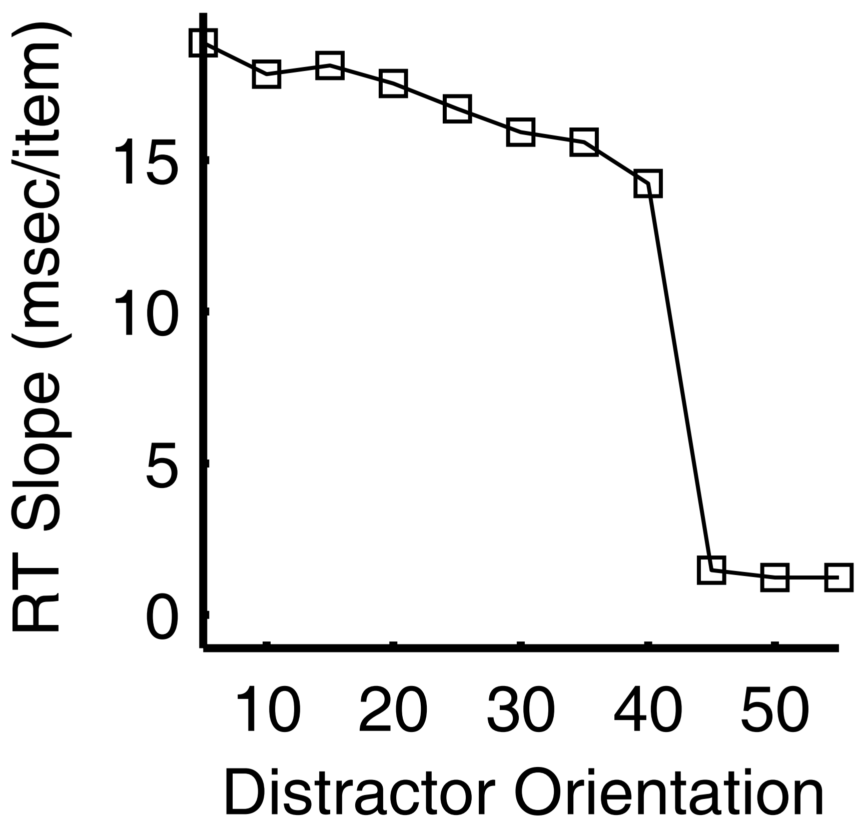


Target-Distractor Similarity Effects 1

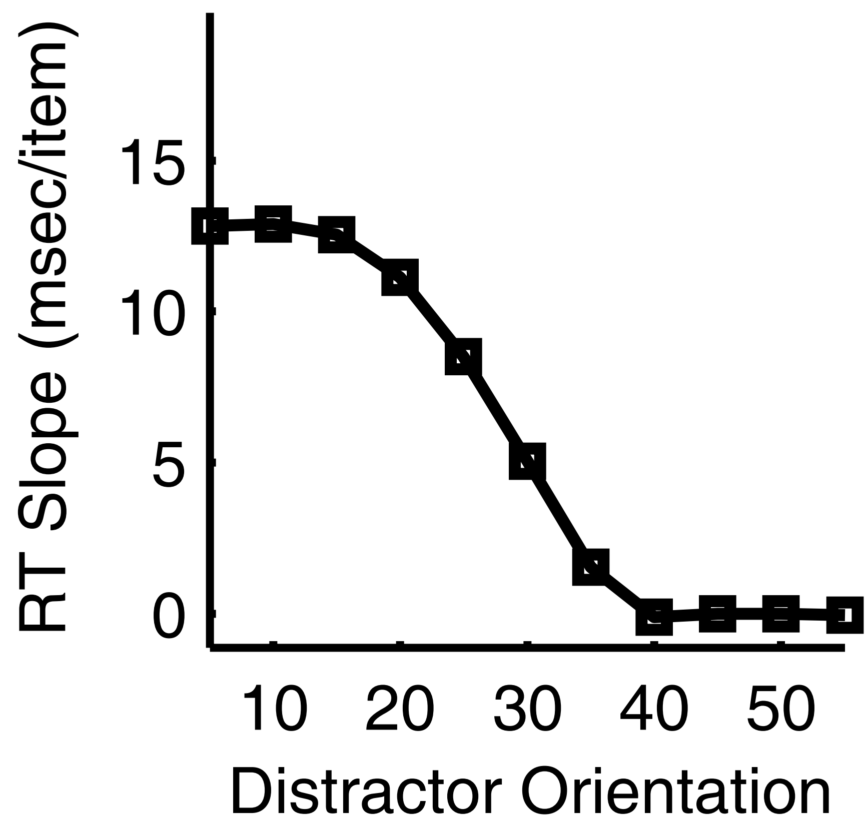


hard -------------------> easy

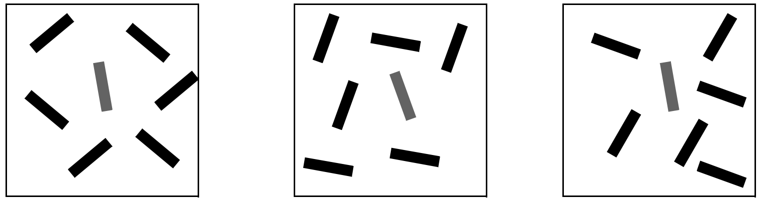
GS



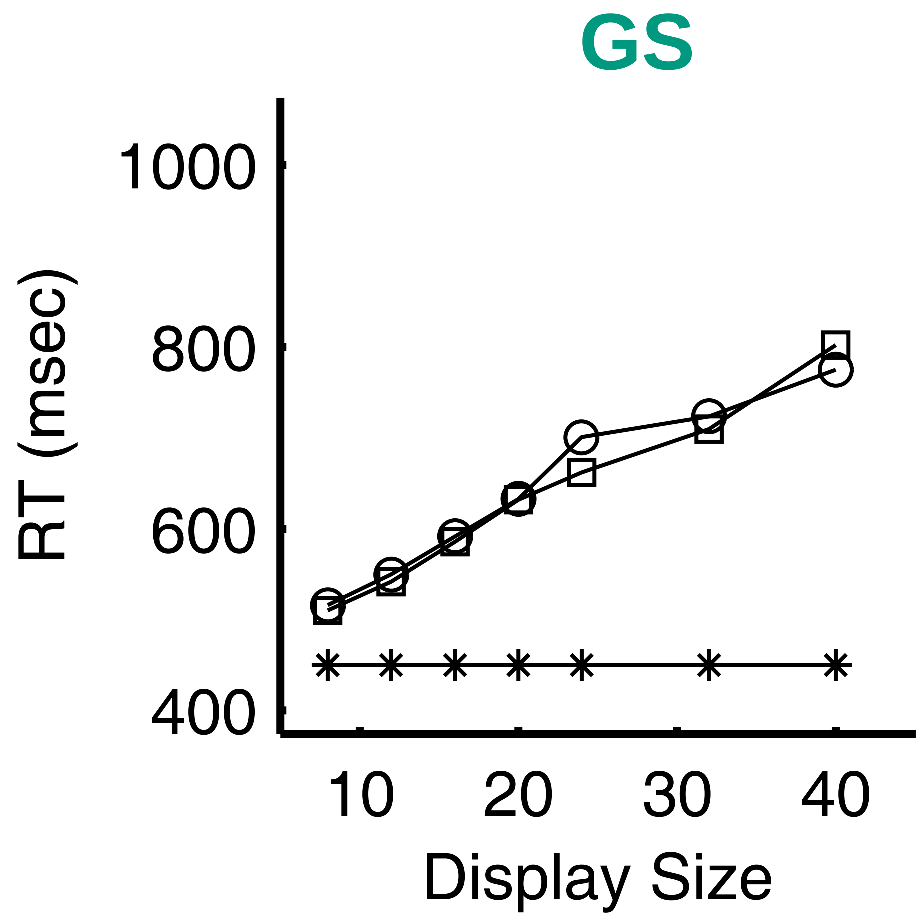
EGS

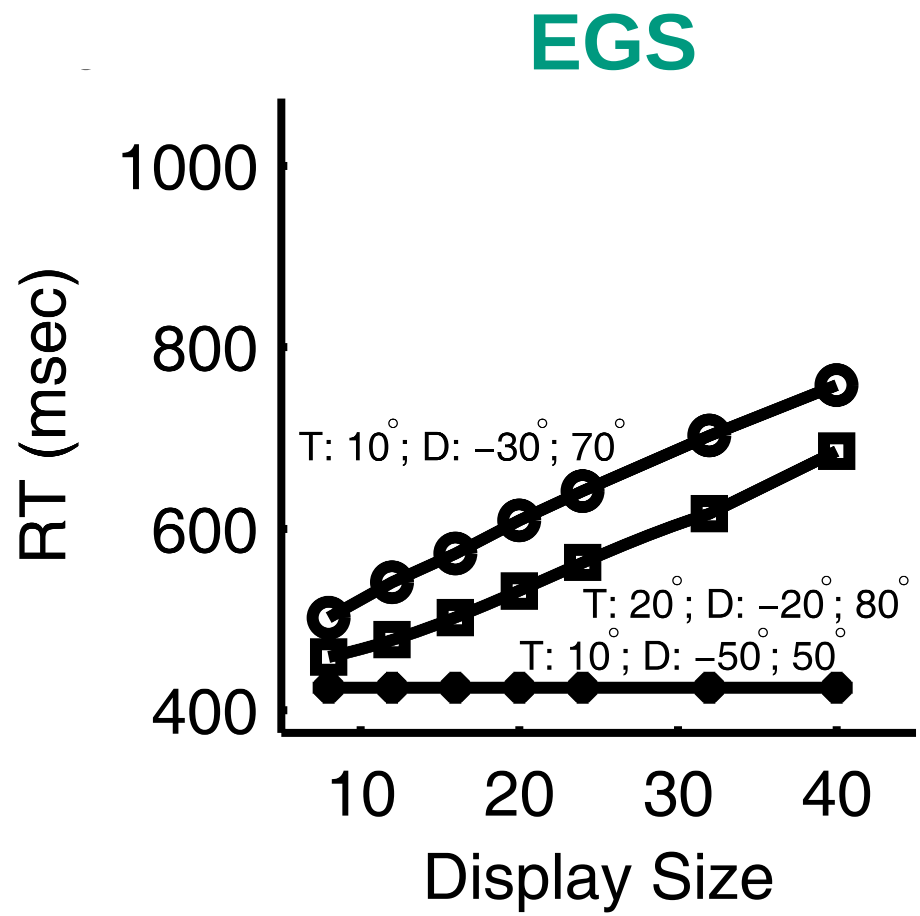


Target-Distractor Similarity Effects 2

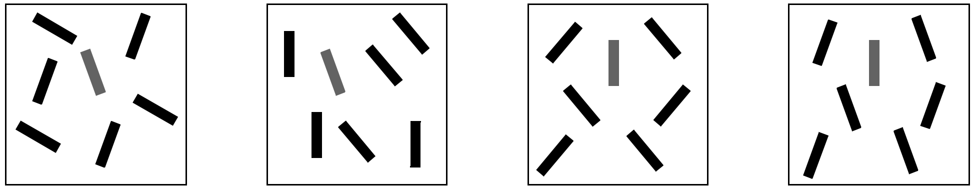


easy -------------------> hard

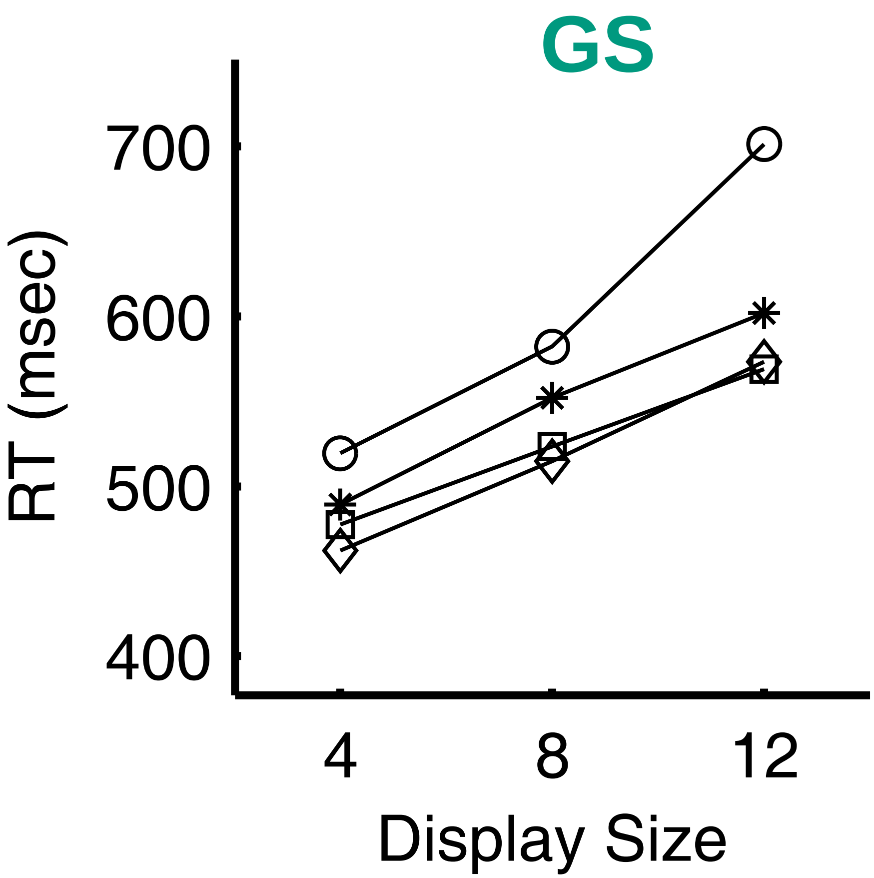


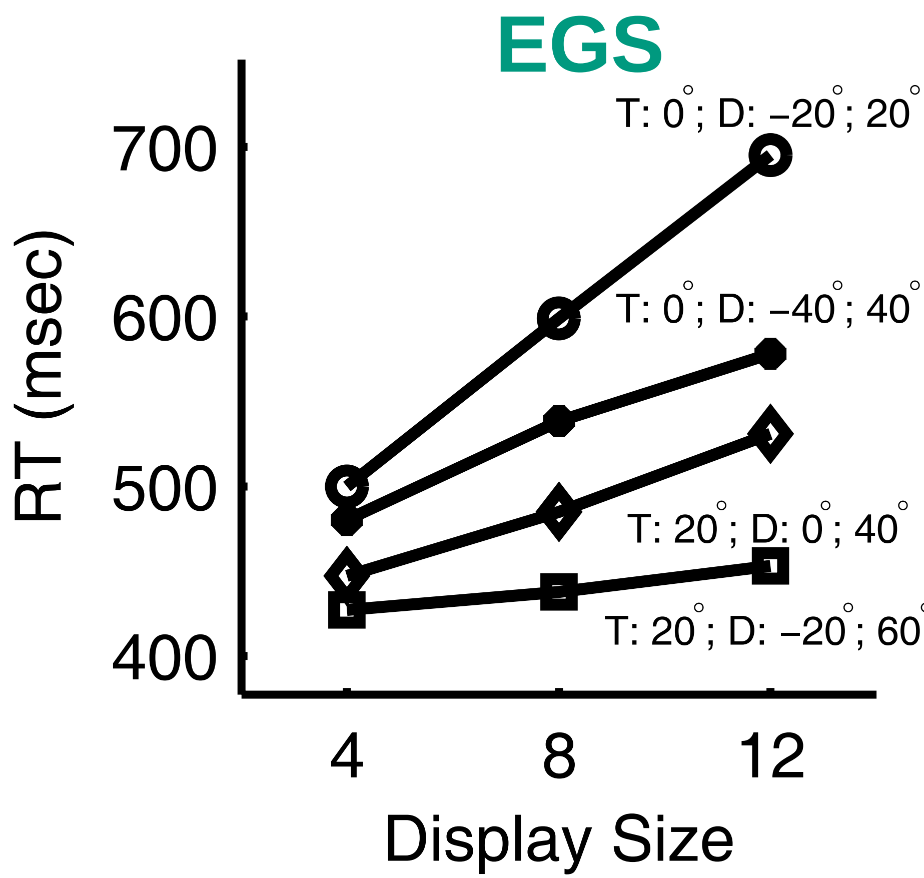


Target-Distractor Similarity Effects 3

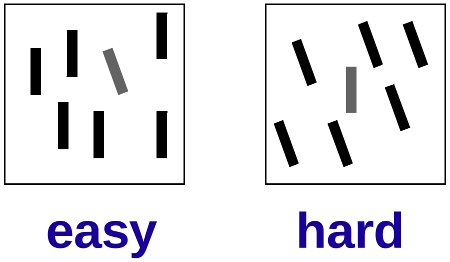


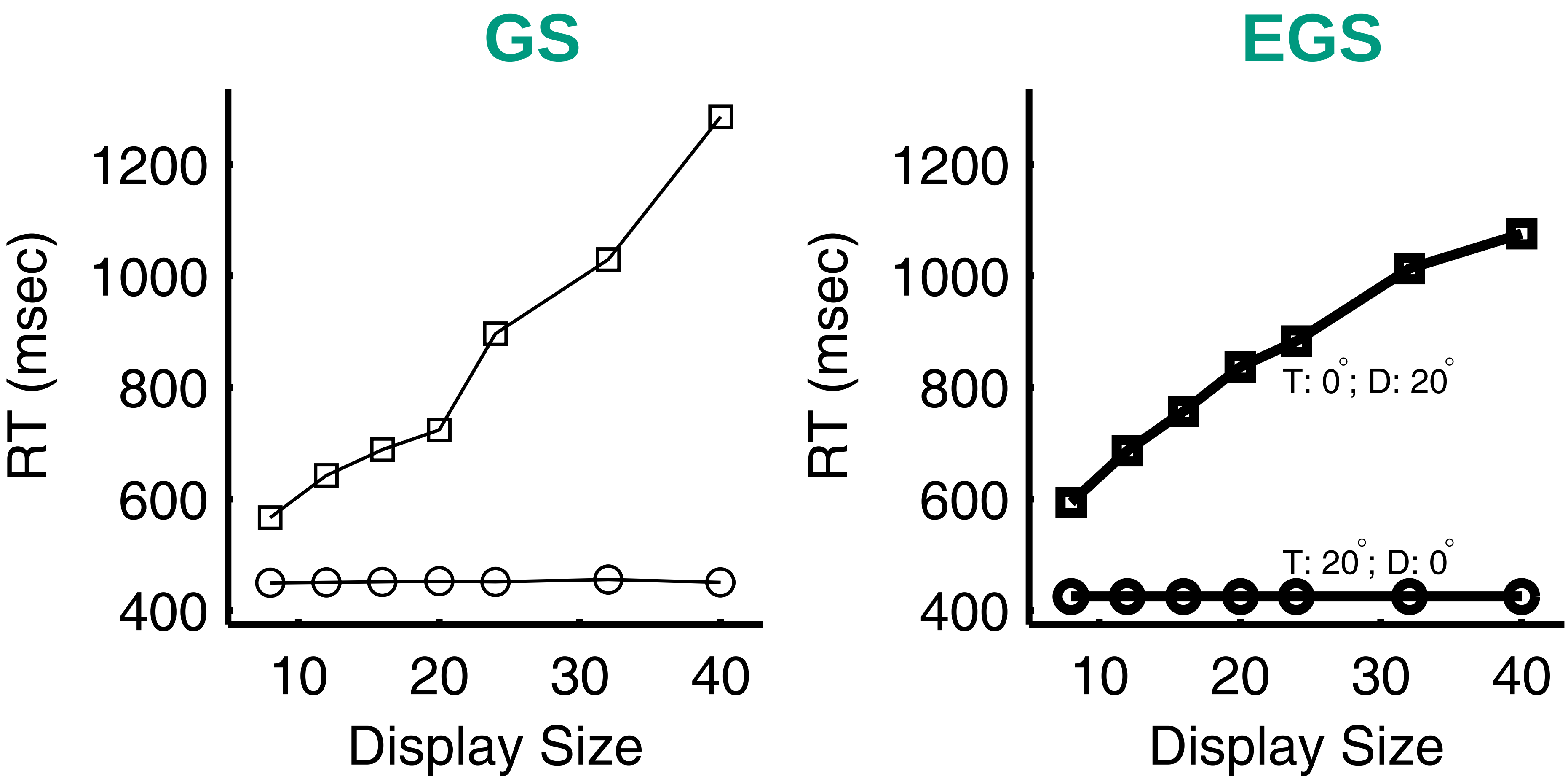
easy -----------------------------------> hard





Target-Distractor Similarity Effects 4





What It Boils Down To

What It Boils Down To

- Generate stimulus sequence corresponding to experiment.

What It Boils Down To

- **Generate stimulus sequence corresponding to experiment.**
- **Initialize task statistics**

$\rho_{i0} \sim \theta$, $\rho_{i1} \sim \varphi$, such that $\theta < \varphi$

What It Boils Down To

- Generate stimulus sequence corresponding to experiment.
- Initialize task statistics
- On each trial, perform feature extraction on display.

What It Boils Down To

- **Generate stimulus sequence corresponding to experiment.**
- **Initialize task statistics**
- **On each trial, perform feature extraction on display.**
- **Compute saliency at each location x**

$$s_x = \sum_i \sum_{t=0}^{1} \frac{1-2t}{\rho_{it}(1-\rho_{it})}(f_{xi} - \rho_{it})^2$$

What It Boils Down To

- **Generate stimulus sequence corresponding to experiment.**
- **Initialize task statistics**
- **On each trial, perform feature extraction on display.**
- **Compute saliency at each location x**
- **Determine response time based on ranking**

  Response_time = $\mu_0 + \mu_1$ saliency_ranking_of_target

What It Boils Down To

- **Generate stimulus sequence corresponding to experiment.**
- **Initialize task statistics**
- **On each trial, perform feature extraction on display.**
- **Compute saliency at each location x**
- **Determine response time based on ranking**
- **Update task statistics based on current trial feature activity**

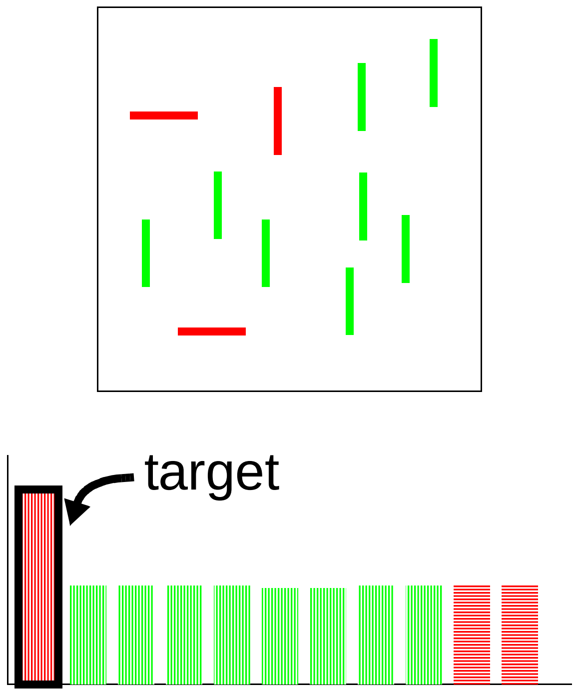
$$\alpha_{it} \leftarrow \lambda \alpha_{it}^{0} + (1-\lambda)\left(\alpha_{it} + \sum_{x \in \chi_t} f_{xi}\right)$$

$$\beta_{it} \leftarrow \lambda \beta_{it}^{0} + (1-\lambda)\left(\beta_{it} + \sum_{x \in \chi_t} 1 - f_{xi}\right)$$

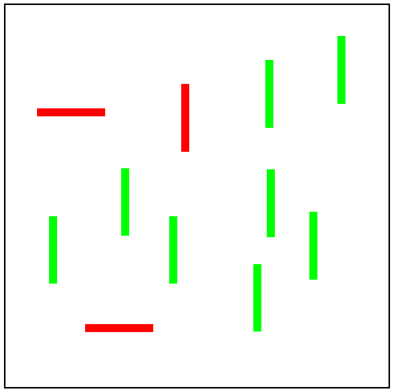
where $\rho_{it} = \dfrac{\alpha_{it}}{(\alpha_{it} + \beta_{it})}$

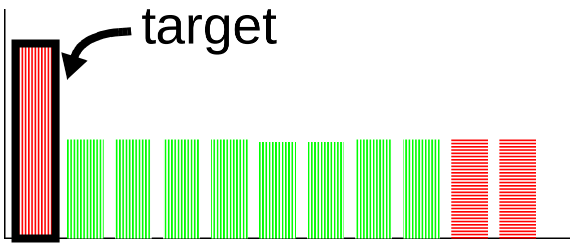
Basically, compute mean activity of feature when target is present and when target is absent.

TRIAL 1

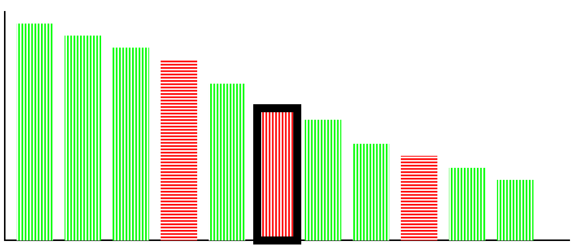


TRIAL 1

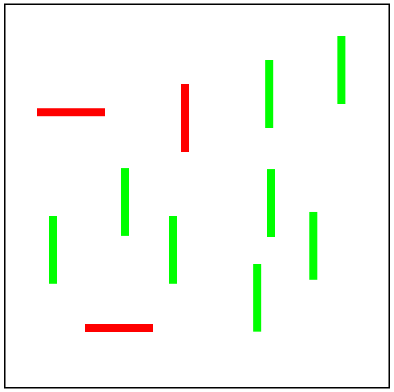




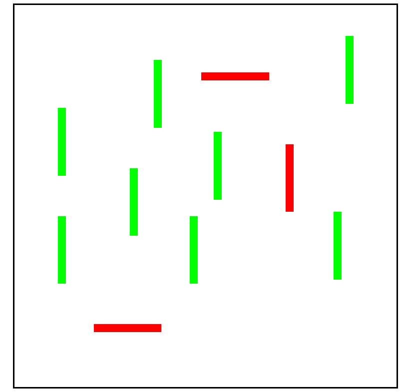
Noisy rank-based prioritization of display elements



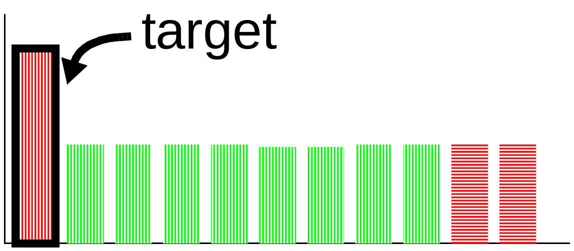
TRIAL 1



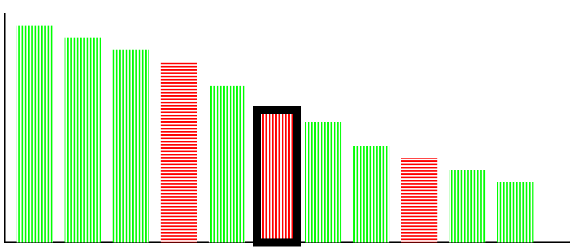
TRIAL 2

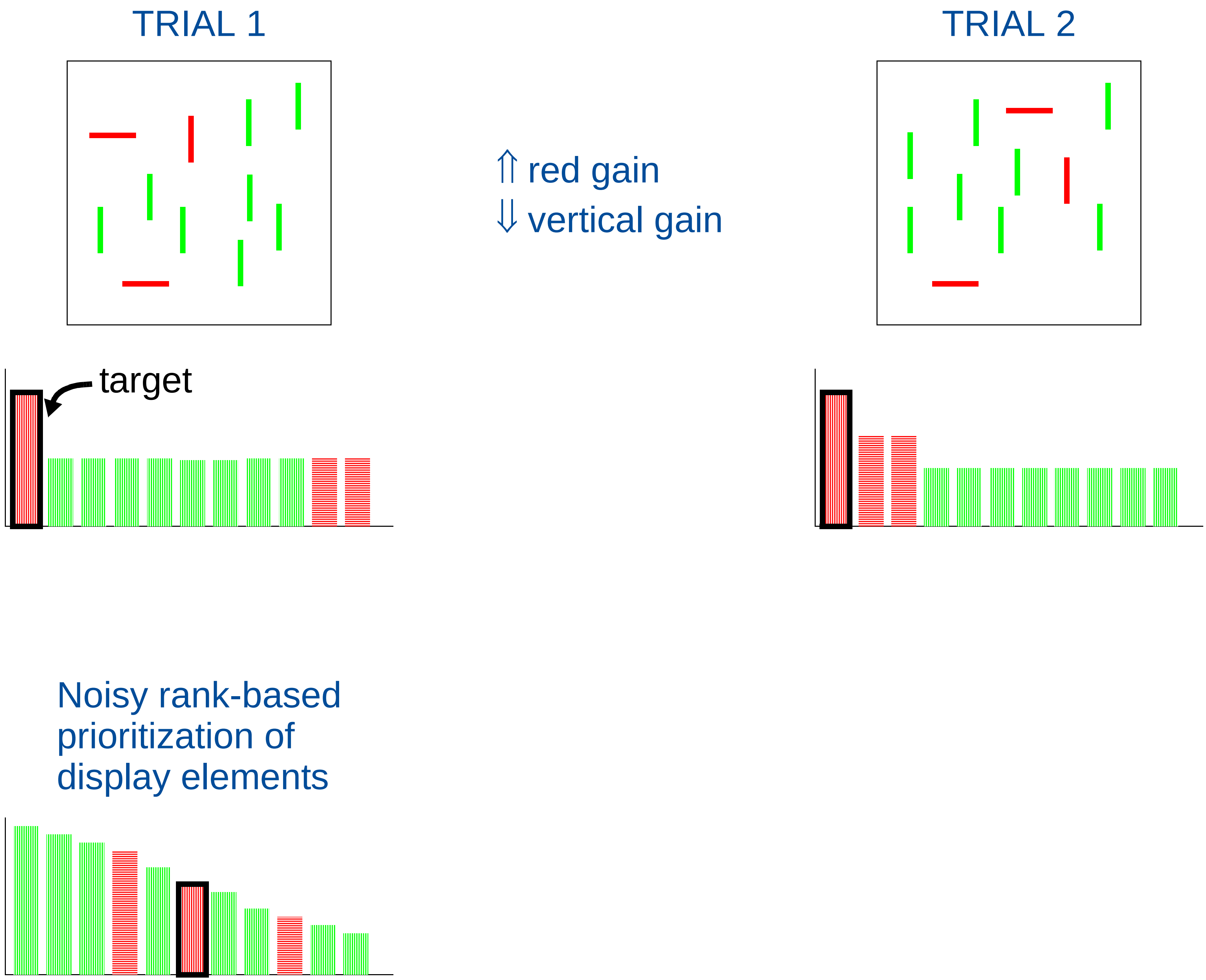


⇑ red gain
⇓ vertical gain

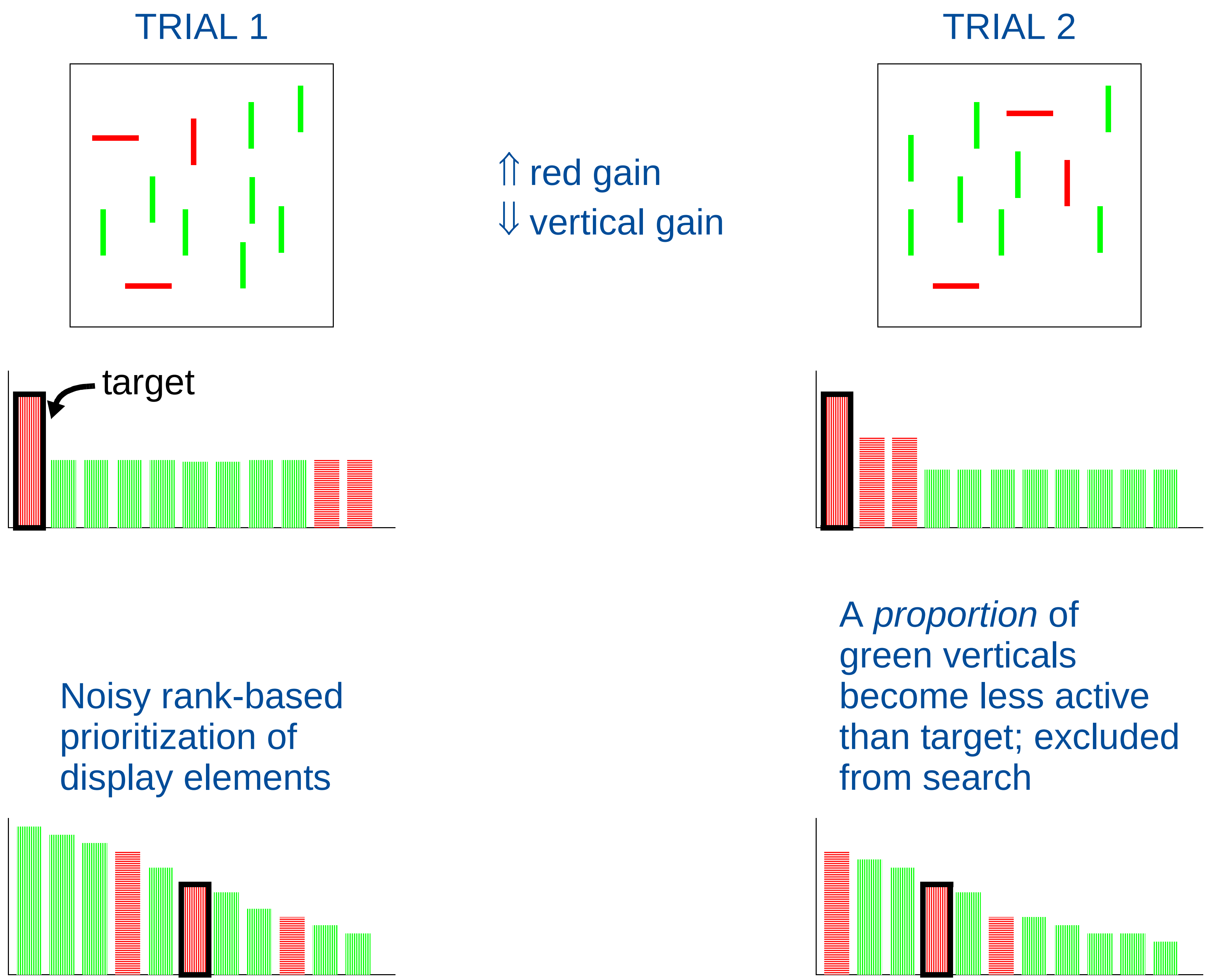


Noisy rank-based
prioritization of
display elements





TRIAL 1
TRIAL 2
⇑ red gain
⇓ vertical gain
target
Noisy rank-based prioritization of display elements



TRIAL 1
TRIAL 2
⇑ red gain
⇓ vertical gain
target
Noisy rank-based prioritization of display elements
A *proportion* of green verticals become less active than target; excluded from search

Can we blame oddball task?

Task involves not just search, but comparing display elements to determine which is target.

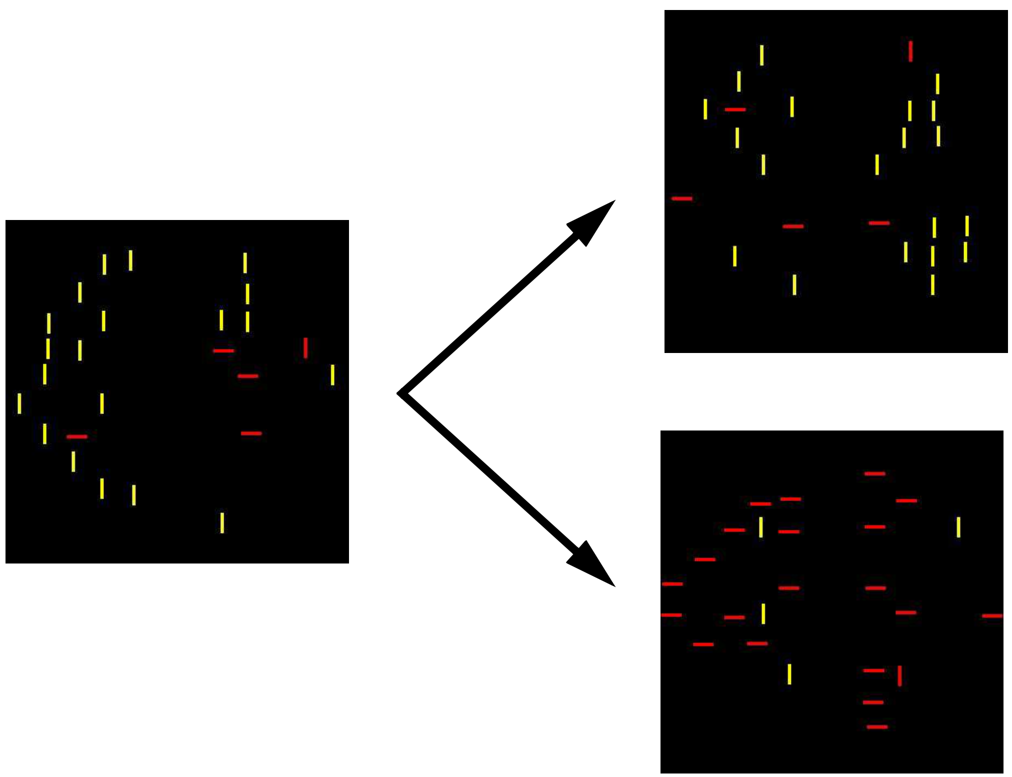
Can we blame oddball task?

Task involves not just search, but comparing display elements to determine which is target.

Alternative

Fixed target search

Trial-to-trial variation in distractor statistics



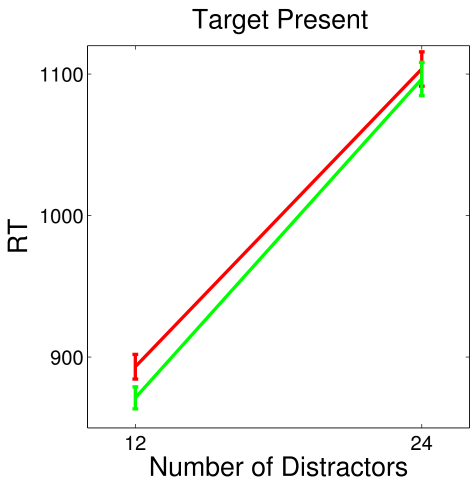
repeat

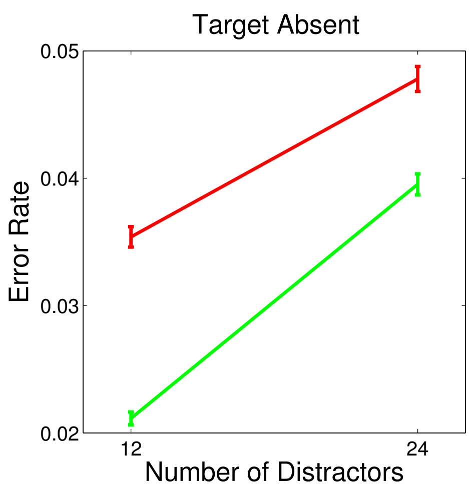
switch

A Desperate And Failed Effort...

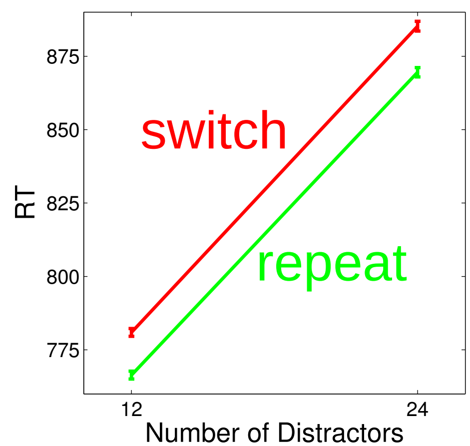
A Desperate And Failed Effort...

present or absent?

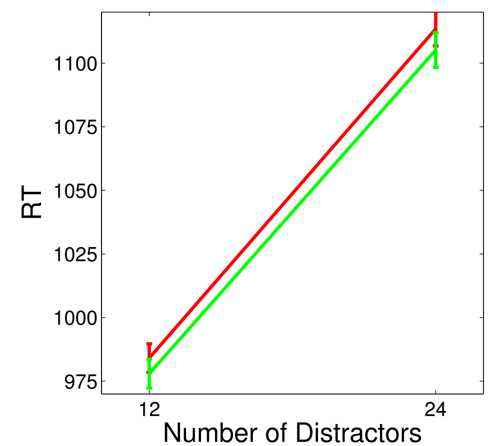




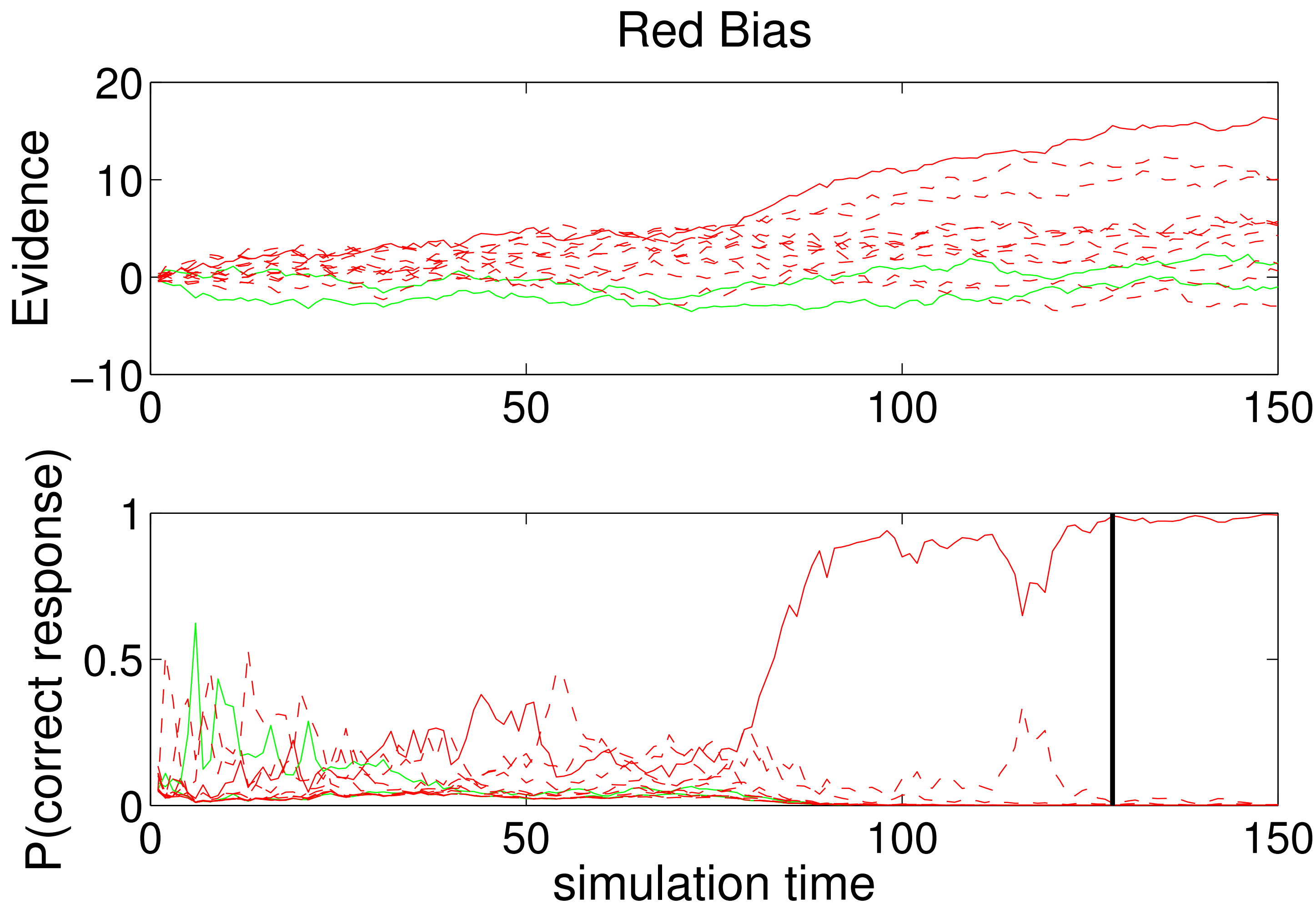
left or right side of display?



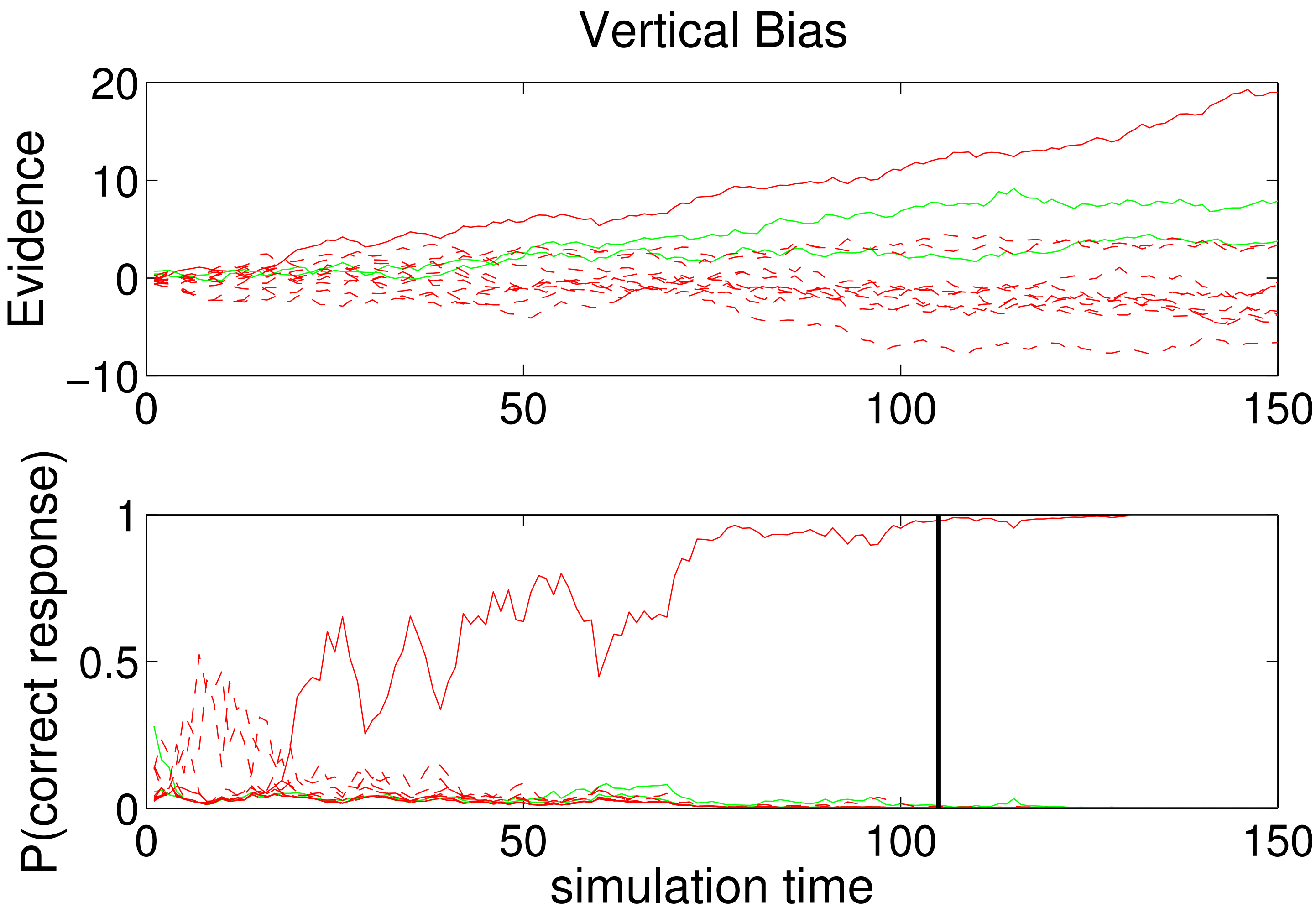
target contains gap?

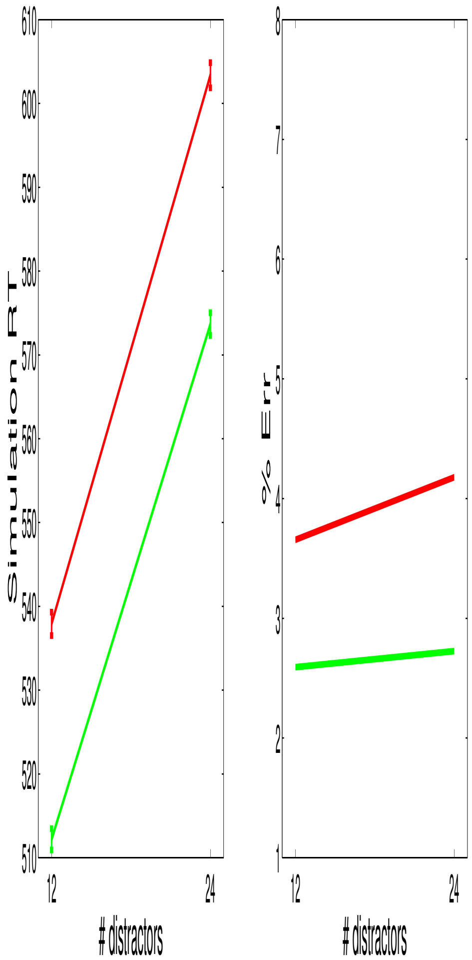
Accumulator Model



Accumulator Model



Accumulator Model



Lesson for experimentalists

No interaction ≠ priming effects are post- or pre-attentional

* BEGIN DIGRESSION *

Estimating ρ

ρ depends on task environment.

- ρ is estimated based on experience performing task.

E.g., what's the response of a red feature detector for a target ($\rho_{red,1}$)?

- Collect responses of red feature detector at locations containing a target.
- Suppose we observe: .47, .62, .91, .55, .80

Could compute maximum likelihood estimate, i.e.,

- $\rho_{red,1} = (.47 + .62 + .91 + .55 + .80) / 5$

Instead, model uses Bayesian parameter estimation.

- Consider all possible values of ρ and determine their plausibility based on how well they fit the data.

Intuitive Example

Coin with unknown bias, ρ = probability of heads

Sequence of observations: H T T H T T T H

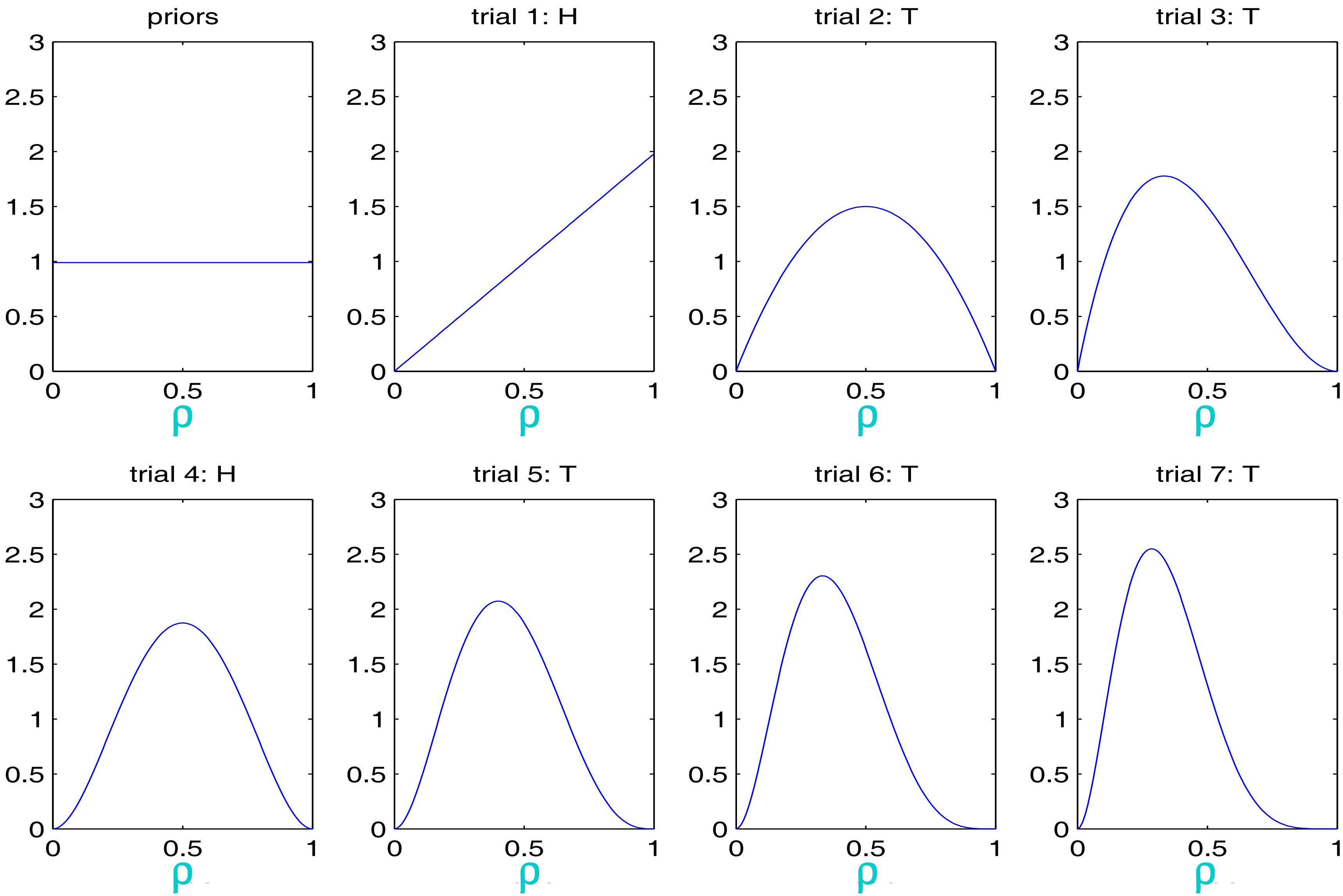
Maximum likelihood approach

$\rho = 3 / 8$

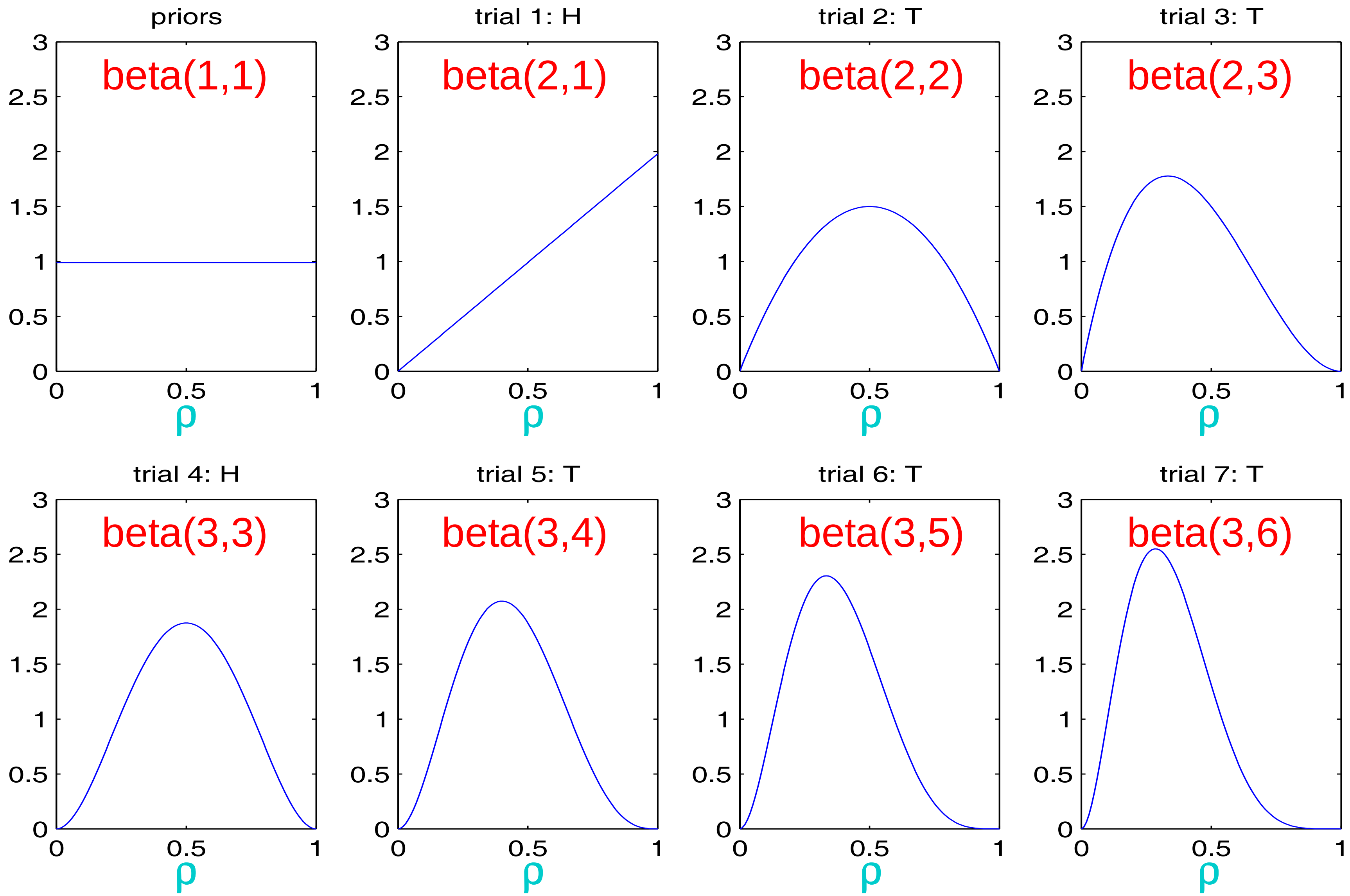
Bayesian approach

set of hypotheses, each associated with a different value of ρ

Coin Flip Sequence: H T T H T T T

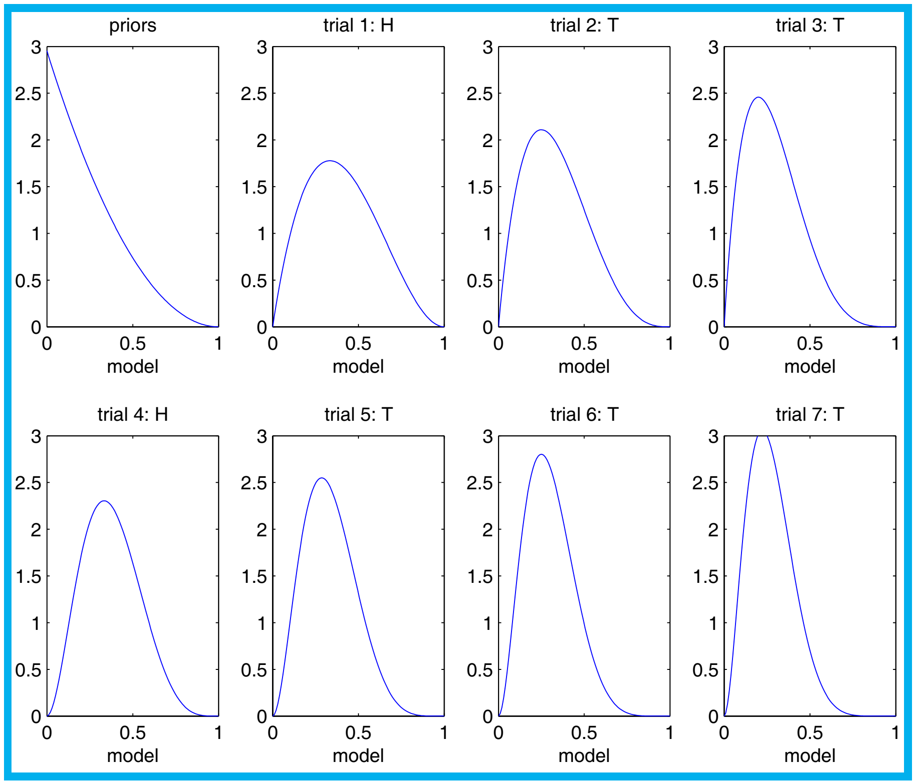


Coin Flip Sequence: H T T H T T T

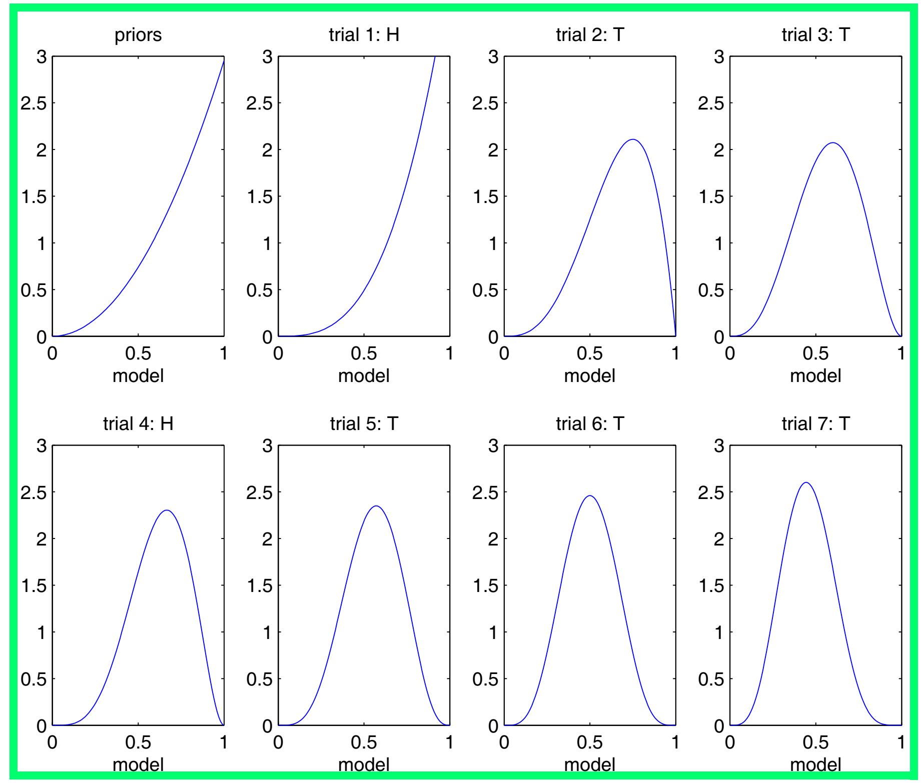


Effect of Prior Knowledge

low head-probability bias

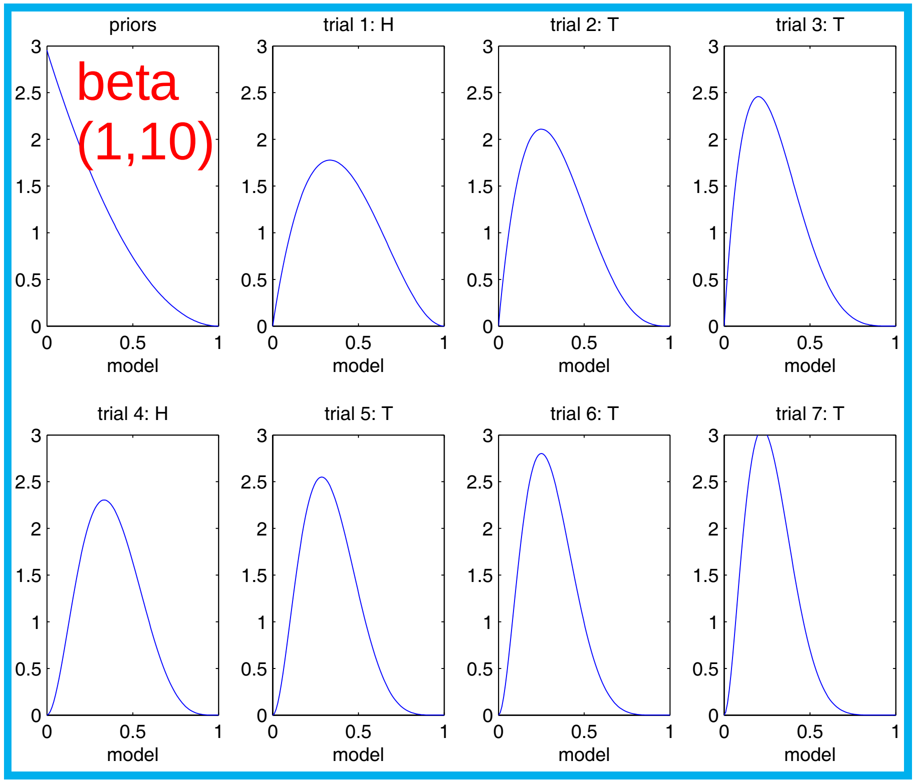


high head-probability bias

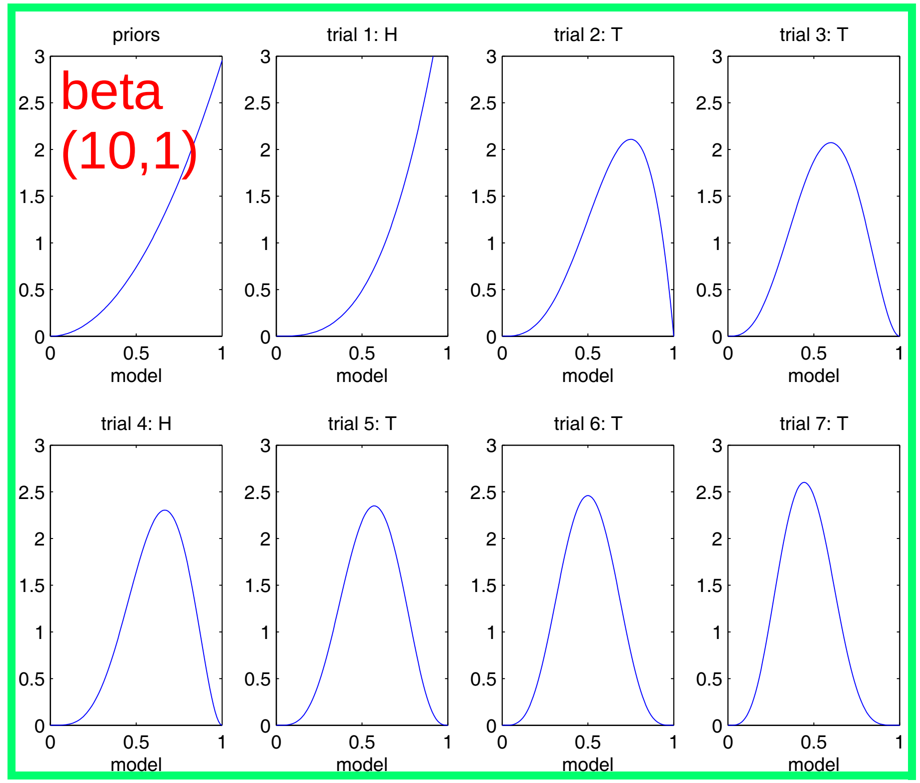


Effect of Prior Knowledge

low head-probability bias



high head-probability bias



* END OF DIGRESSION *

$$\textbf{Saliency} \equiv P(T_x \mid \mathbf{F}_x, \rho)$$

task statistics – learned thru experience

feature activity (vector) at location x

target at location x? (1=true, 0=false)

Saliency $\equiv P(T_x \mid \mathbf{F}_x, \rho)$

ρ: task statistics – learned thru experience

$\mathbf{F}_x$: feature activity (vector) at location x

T_x: target at location x? (1=true, 0=false)

Bayes' Rule

$$P(T_x \mid \mathbf{F}_x, \rho) = \frac{P(T_x)P(\mathbf{F}_x \mid T_x, \rho)}{\sum_{t=0}^{1} P(T_x = t)P(\mathbf{F}_x \mid T_x = t, \rho)}$$

Saliency $\equiv P(T_x | \mathbf{F}_x, \rho)$

task statistics – learned thru experience

feature activity (vector) at location x

target at location x? (1=true, 0=false)

Bayes' Rule

$$P(T_x | \mathbf{F}_x, \rho) = \frac{P(T_x)P(\mathbf{F}_x | T_x, \rho)}{\sum_{t=0}^{1} P(T_x = t)P(\mathbf{F}_x | T_x = t, \rho)}$$

$P(\mathbf{F}_x | T_x, \rho)$ **is a task-specific model of the environment.**

Indicates visual system response ($\mathbf{F}_x$) for targets (T_x=1) vs. nontargets (T_x=0)

We make specific claims about the form of this model that resides in our head.

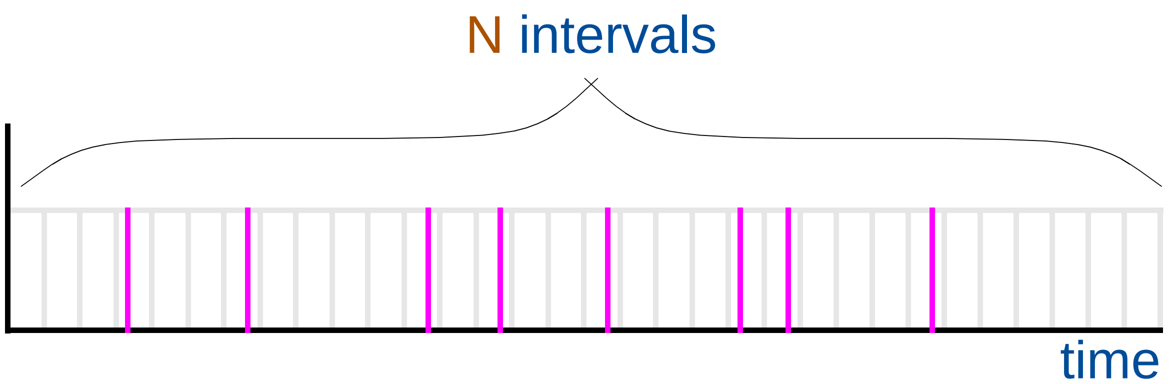
This model does not have to match the actual environment.

Generally, models are simplified to be mathematically tractable.

1. **Assume feature responses are conditionally independent of one another, i.e.,**

$$P(\mathbf{F}_x | T_x, \rho) = \prod_i P(F_{xi} | T_x, \rho)$$

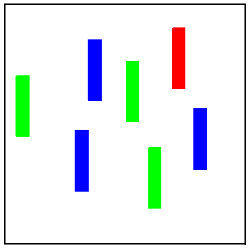
2. **Assume feature detection is carried out by a rate-coded spiking neuron**



F_{xi}: count of the number of spikes observed for feature i at location x

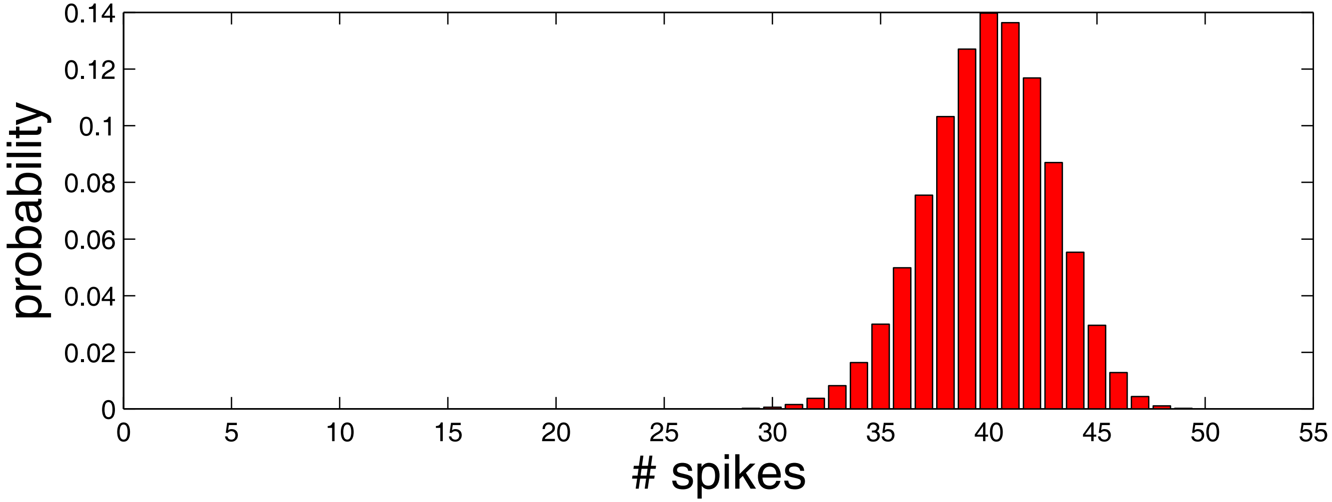
ρ_{it}: spike rate for feature *i* if the target is of type *t*

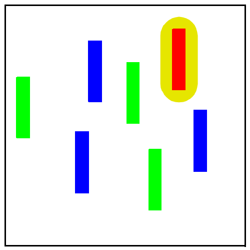
Consider search for a red object among non-red.
What will the response of the red feature detector be?



Consider search for a red object among non-red.
What will the response of the red feature detector be?

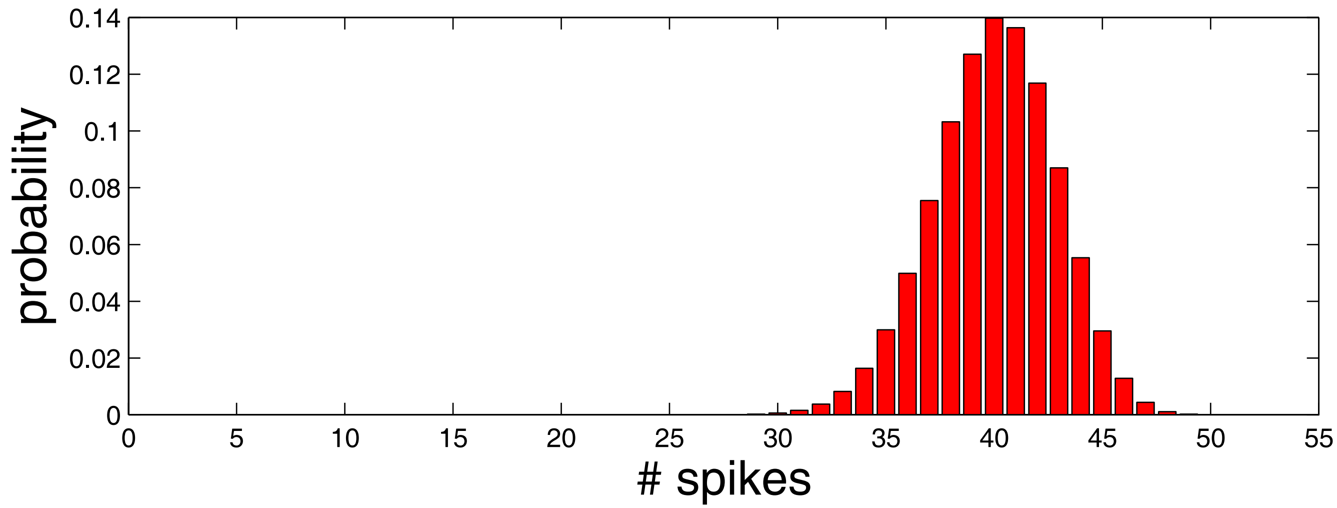
$P(F_{red} | T = 1, \rho_{red,1} = 0.8)$ **for N=50**

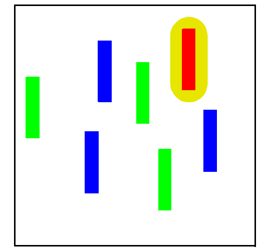




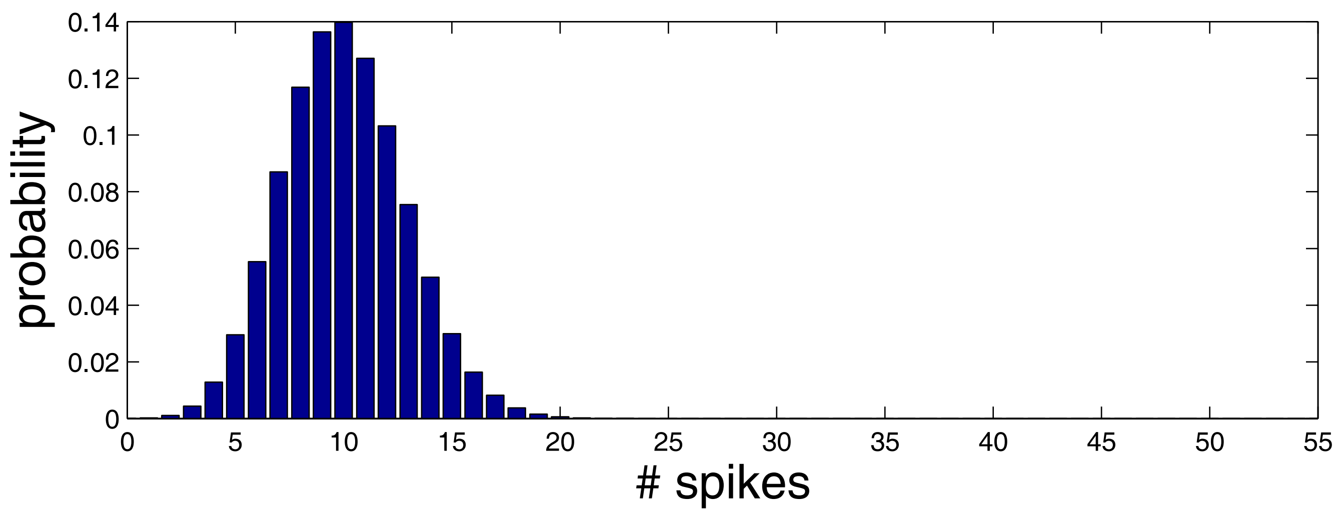
Consider search for a red object among non-red.
What will the response of the red feature detector be?

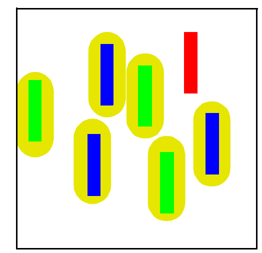
$P(F_{red} | T = 1, \rho_{red,1} = 0.8)$ **for N=50**





$P(F_{red} | T = 0, \rho_{red,0} = 0.2)$ **for N=50**





$$P(F_{xi} | T_x, \rho) \sim \text{Binomial}(\rho_{it}, N)$$

number of time intervals

spiking rate of feature i
for target (t=1) or distractor (t=0)

$$P(F_{xi} \mid T_x, \rho) \sim \cancel{\text{Binomial}(\rho_{it}, N)}$$

number of time intervals

spiking rate of feature i
for target (t=1) or distractor (t=0)

$$\sim \text{Gaussian}(N\rho_{it}, N\rho_{it}(1 - \rho_{it}))$$

To Recap

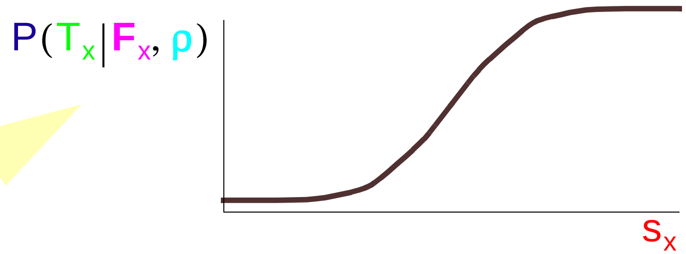
Saliency is defined to be the probability that a location contains a target, $P(T_x \mid \mathbf{F}_x, \rho)$

Given the following assumptions:

1. Feature responses are conditionally independent of one another.
2. Feature detection is carried out by neurons that spike at rate ρ_{it}.
3. The neuron is samples for a time window that would allow >30 spikes.

we get...

$$P(T_x \mid \mathbf{F}_x, \rho) = \frac{1}{1 + e^{-(c_1 + c_2 s_x)}}$$



$$P(T_x|\mathbf{F}_x, \rho) = \frac{1}{1 + e^{-(c_1 + c_2 s_x)}}$$

Because attentional priority depends on relative saliency, we can substitute s_x **for** $P(T_x|\mathbf{F}_x, \rho)$**.**

$$P(T_x \mid \mathbf{F}_x, \rho) = \frac{1}{1 + e^{-(c_1 + c_2 s_x)}}$$

response (spike rate)
of feature i in location x

$$s_x = \sum_i c_{i1}\tilde{f}_{xi} + c_{i2}\tilde{f}_{xi}^2$$

$$s_x = \sum_i c_{i1}\tilde{f}_{xi} + c_{i2}\tilde{f}_{xi}^2$$

$$c_{i1} = \frac{2(\rho_{i1} - \rho_{i0})}{(1-\rho_{i0})(1-\rho_{i1})}$$

$$c_{i2} = \frac{1}{\rho_{i0}(1-\rho_{i0})} - \frac{1}{\rho_{i1}(1-\rho_{i1})}$$

Experience-Guided Search

$$s_x = \sum_i c_{i1}\tilde{f}_{xi} + c_{i2}\tilde{f}_{xi}^2$$

Guided Search

$$s_x = \sum_i c_{i1}\tilde{f}_{xi}$$

Two Further Assumptions of Model

Two Further Assumptions of Model

1. **Bias that all features are considered relevant in the absence of experience**

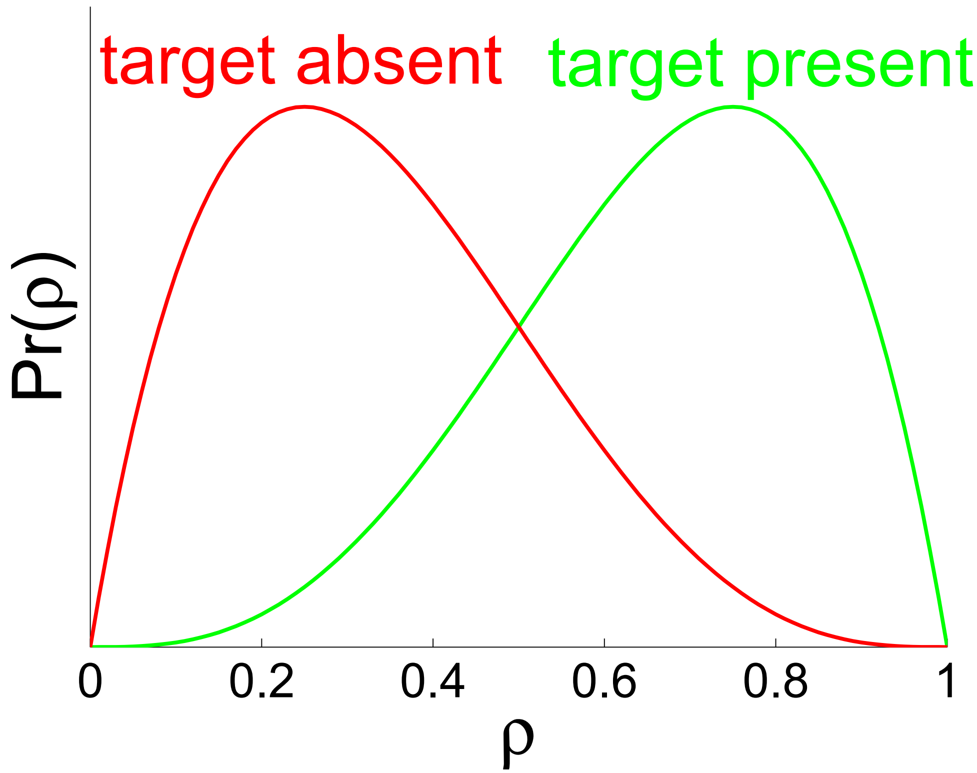
   Achieved by treating ρ as a Beta random variable with imaginary-count prior

Two Further Assumptions of Model

1. Bias that all features are considered relevant in the absence of experience

~~Achieved by treating ρ as a Beta random variable with imaginary-count prior~~

Encode not just the most likely value of ρ, but uncertainty distribution.



Two Further Assumptions of Model

1. **Bias that all features are considered relevant in the absence of experience**

   ~~Achieved by treating ρ as a Beta random variable with imaginary-count prior~~

   Encode not just the most likely value of ρ, but uncertainty distribution.

2. **Environment is nonstationary**

   With probability λ, environment and/or task can change.

Two Further Assumptions of Model

1. Bias that all features are considered relevant in the absence of experience

~~Achieved by treating ρ as a Beta random variable with imaginary-count prior~~

Encode not just the most likely value of ρ, but uncertainty distribution.

2. Environment is nonstationary

With probability λ, environment and/or task can change.

From these two claims, we have *three* free parameters total.

Qualitative performance does not depend on parameters as long as $\lambda>0$ and $E[\rho_{i0}] < E[\rho_{i1}]$