A Processing Ghost in a Tank Machine

Mario Fifić

Grand Valley State University, Michigan

Midwestern Cognitive Science Conference
May, 2014

Working memory (WM) under consideration

- Processing order
- Capacity status
- Subdivision by modalities
- Dual or one system
- Status of mental representations & resource allocation

Resource allocation models of WM

- The discrete-slot model proposes that WM operates on the **ALL-OR-NONE** principle: holding only high-resolution item representations stored in a limited number of memory slots.
  - The slots+averaging model is variant of the discrete-slot model assuming that more than one slot could be allocated to a single item representation

- the variable-resources model WM operates on the **ALL-GET-SOME** principle: a pool of limited resources is dynamically allocated across a set of memorized items representations.

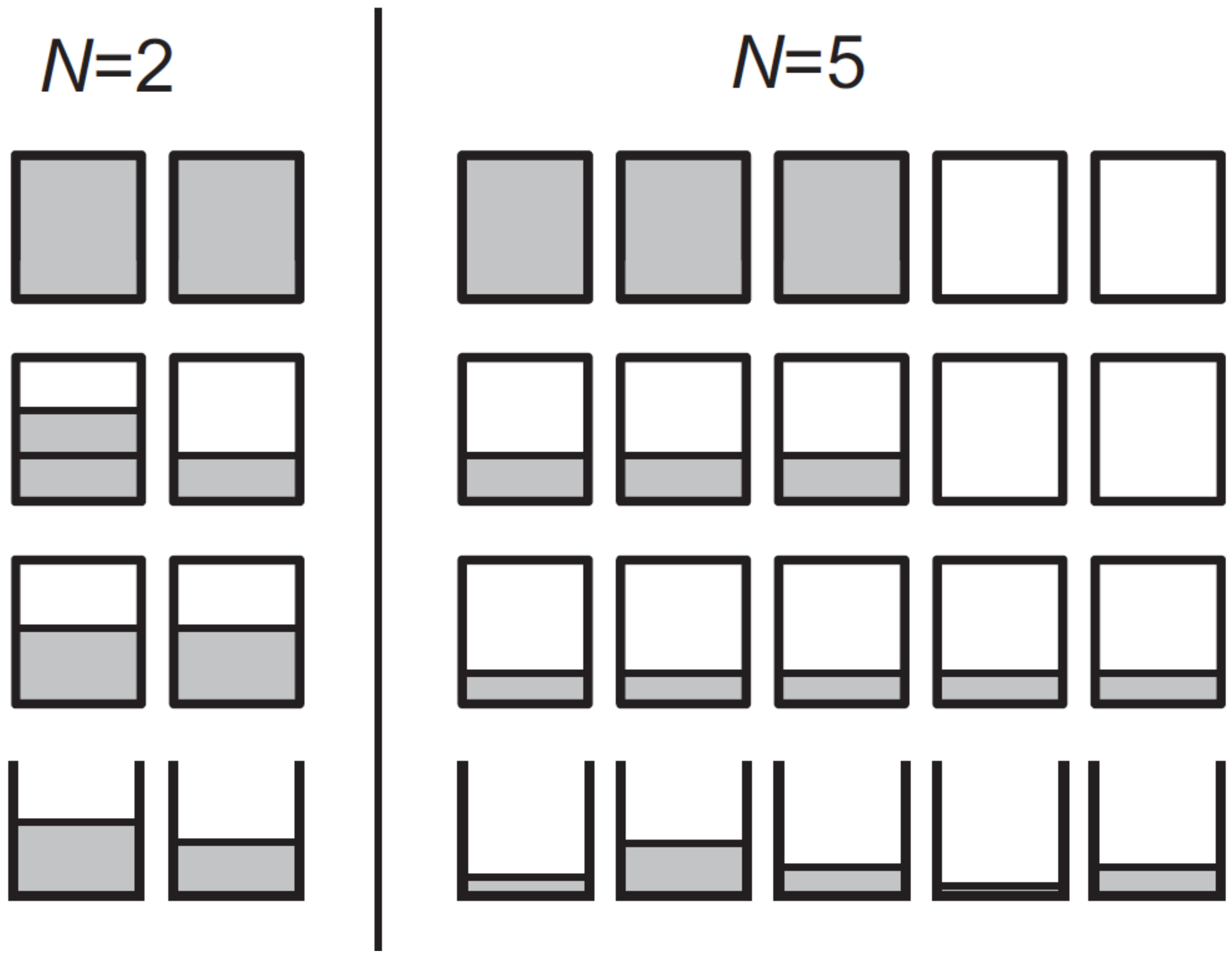
Resource allocation in model of WM

ALL-OR-NONE
Discrete slots model

Slots+Averiging

Equal distribution

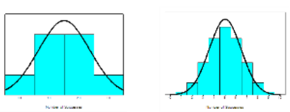
ALL-GET-SOME
A continuous resource model



discrete-slot model

slots+averaging model

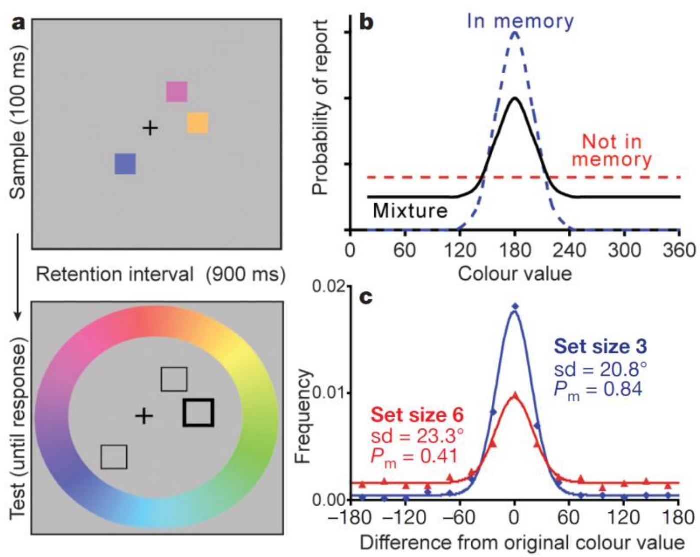
variable-resources model

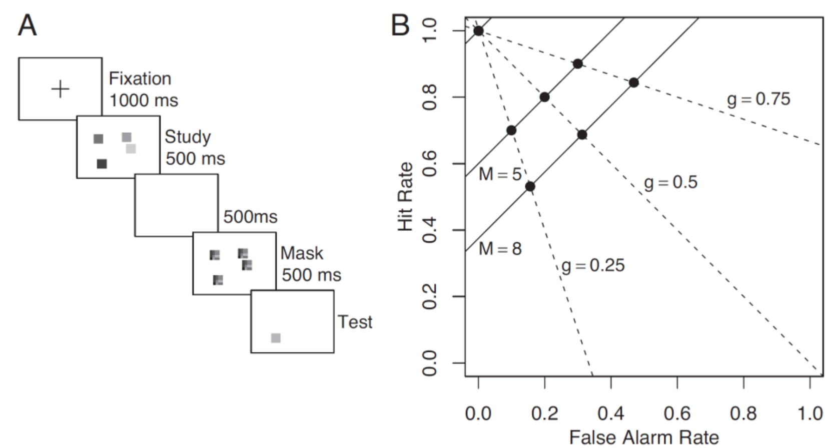


Lim f(x)
n->Infinity

Evidence supporting Discrete Slots Model

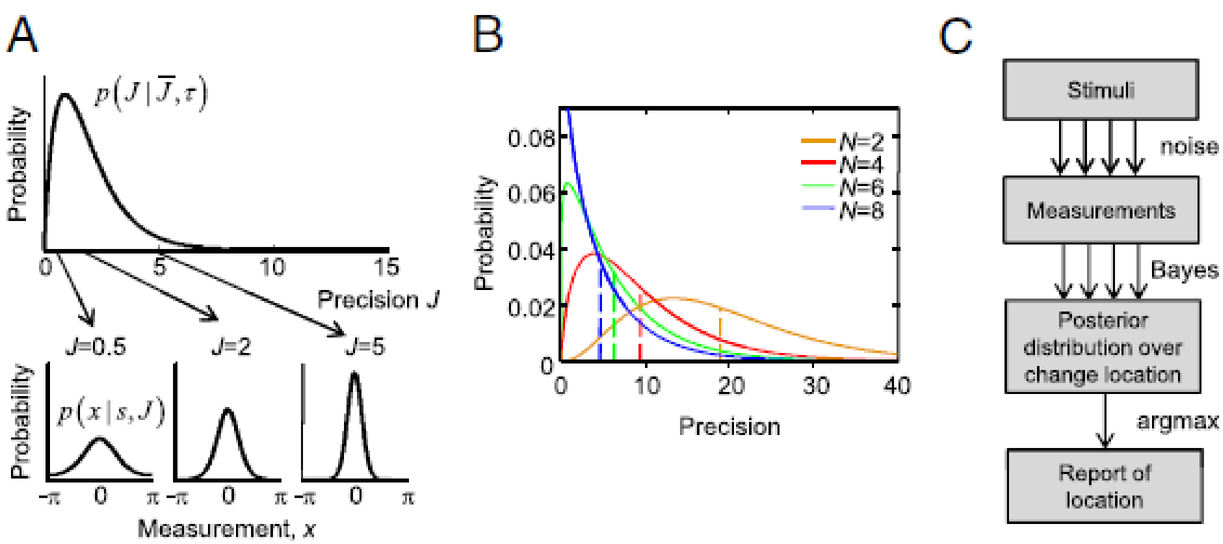
- Zhang & Luck 2008
- Cowan (2001) The magical number 4 in short-term memory
- Rouder, Morey, Cowan, Zwilling, Morey, & Pratte (2008).
- Donkin, Nosofsky, Gold, & Shiffrin, (2013).





Evidence supporting the Variable-resource model

- Van den Berg, Shin, Chou, George, & Ma, (2012)
- Bays & Husain (2008)



van den Berg, et al. (2012). Appendix..

PNAS PNAS PNAS

ments $\mathbf{x} = (x_1, \ldots, x_N)$ and $\mathbf{y} = (y_1, \ldots, y_N)$, we use a Bayesian-observer model. The Bayesian observer computes a probability distribution over the location of the change, $p(L \mid \mathbf{x}, \mathbf{y})$, and then reports the location with the highest probability. The posterior distribution over L is proportional to the joint distribution, $p(\mathbf{x}, \mathbf{y}, L)$, which in turn is evaluated as an integral over the remaining variables, namely Δ, $\boldsymbol{\theta}$, and $\boldsymbol{\varphi}$,

$$\begin{aligned} p(\mathbf{x},\mathbf{y},L) &= \iiint p(\mathbf{x},\mathbf{y},\boldsymbol{\theta},\boldsymbol{\varphi},\Delta,L)\,d\Delta d\boldsymbol{\theta} d\boldsymbol{\varphi} \\ &= \iiint p(L)p(\Delta)p(\boldsymbol{\theta})p(\boldsymbol{\varphi}\mid L,\boldsymbol{\theta})\,p(\mathbf{x}\mid\boldsymbol{\theta})\,p(\mathbf{y}\mid\boldsymbol{\varphi})\,d\Delta d\boldsymbol{\theta} d\boldsymbol{\varphi}, \end{aligned}$$

where in going from the first to the second line we have used the structure of the generative model in Fig. S1*B*. Substituting distributions and evaluating the integral over $\boldsymbol{\varphi}$ gives

$$p(\mathbf{x},\mathbf{y},L) = \frac{1}{N}\left(\frac{1}{2\pi}\right)^{N+1} \int \prod_{i=1}^{N}\left(\int p(x_i \mid \theta_i)p(y_i \mid \varphi_i = \theta_i + \Delta\delta_{L,i})\right)d\Delta, \qquad \textbf{[S16]}$$

where $\delta_{L,i} = 1$ when $L = i$ and 0 otherwise. Because we are interested only in the dependence on L, we can freely divide by the L-independent product $\prod_{i=1}^{N}\left(\int p(x_i \mid \theta_i)p(y_i \mid \varphi_i = \theta_i)\right)$, leaving only integrals pertaining to the Lth location:

$$p(\mathbf{x},\mathbf{y},L) \propto \frac{\iint p(x_L \mid \theta_L)\,p(y_i \mid \varphi_L = \theta_L + \Delta)\,d\theta_L d\Delta}{\int p(x_L \mid \theta_L)\,p(y_L \mid \varphi_L = \theta_L)}. \qquad \textbf{[S17]}$$

that is among the encoded ones. In analogy to Eq. **S16**, this probability is

$$\begin{aligned} (L\text{ encoded})\; p(\mathbf{x},\mathbf{y},L) &= \frac{1}{N}\left(\frac{1}{2\pi}\right)^{K+1} \\ &\times \int \prod_{i=1}^{K}\left(\int p(x_i \mid \theta_i)p(y_i \mid \varphi_i = \theta_i + \Delta\delta_{L,i})\right)d\Delta. \end{aligned} \qquad \textbf{[S19]}$$

Now we evaluate the joint probability of $\mathbf{x}$, $\mathbf{y}$ and that the change occurred at a location L that is not among the encoded ones. This probability is equal to

$$\begin{aligned} (L\text{ not encoded})\; p(\mathbf{x},\mathbf{y},L) &= \iint p(\mathbf{x},\mathbf{y},\boldsymbol{\theta},\boldsymbol{\varphi},L)\,d\boldsymbol{\theta} d\boldsymbol{\varphi} \\ &= \iint p(L)\,p(\boldsymbol{\theta})\,p(\boldsymbol{\varphi}\mid L,\boldsymbol{\theta})\,p(\mathbf{x}\mid\boldsymbol{\theta})\,p(\mathbf{y}\mid\boldsymbol{\varphi})\,d\boldsymbol{\theta} d\boldsymbol{\varphi} \\ &= \frac{1}{N}\left(\frac{1}{2\pi}\right)^{K}\prod_{i=1}^{K}\left(\int p(x_i \mid \theta_i)\,p(y_i \mid \varphi_i = \theta_i)\right). \end{aligned} \qquad \textbf{[S20]}$$

As one would expect, this probability does not depend on L. Because we are interested only in the location L for which $p(\mathbf{x}, \mathbf{y}, L)$ is largest (i.e., the argmax), we divide both Eqs. **S19** and **S20** by Eq. **S20**. Then, in analogy to Eq. **S17**, we have to take the argmax of

$$\begin{cases} (L\text{ encoded}) & \dfrac{1}{2\pi}\dfrac{\iint p(x_L \mid \theta_L)\,p(y_i \mid \varphi_L = \theta_L + \Delta)\,d\theta_L d\Delta}{\int p(x_L \mid \theta_L)\,p(y_L \mid \varphi_L = \theta_L)} = \dfrac{1}{2\pi\int p(x_L \mid \theta_L)\,p(y_L \mid \varphi_L = \theta_L)} \\ (L\text{ not encoded}) & 1. \end{cases}$$

Unresolved question(s)
(what’s under the All-or-none carpet)
or
“Why I am not an enthusiast “

- (1) We argue that the above research advances have been downplaying the experimental approaches to directly manipulate the allocation of resources across item representations held by WM.
  - Our study showed that, when instructed, subjects adaptively allocated a limited amount of resources and shared them across memorized item representations.

Unresolved question(s) (what’s under the carpet)

- (2) The exact mechanism of resource allocation has not been specified.

Specific Research Questions

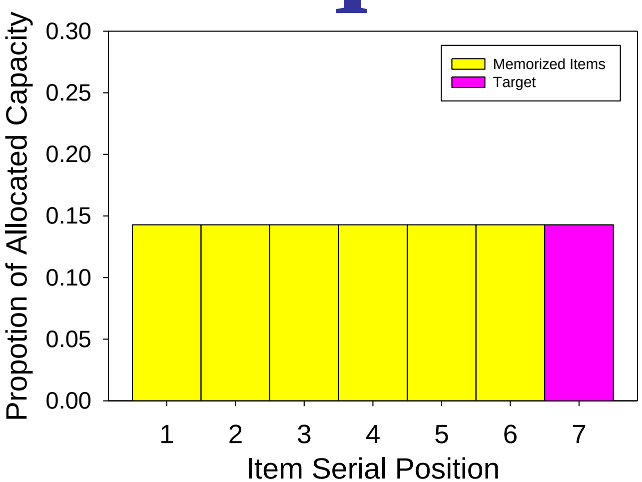
- How are the resources allocated in WM? (Half-Half rule)

- What is the status of mental representations in WM?

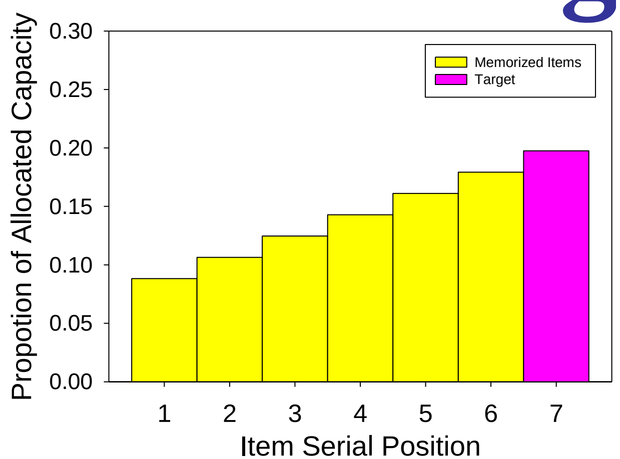
How are the capacity resources allocated?

Attentional gating function in WM

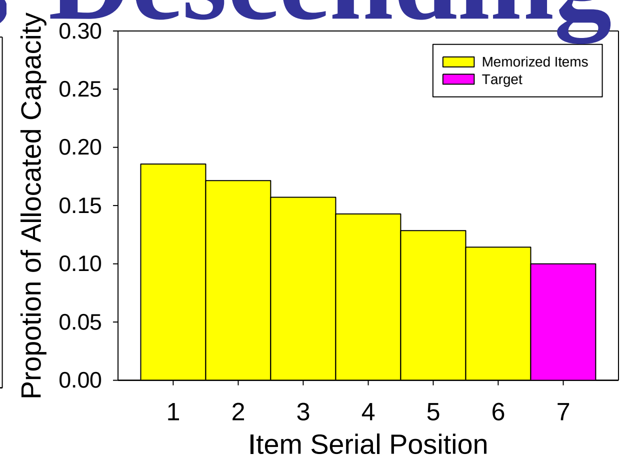
Equal



Ascending



Descending



How are the capacity resources distributed?

The Half-Half Optimal Rule

- The optimal solution for allocation of a limited amount of resources: one Half of resources should be allocated to memorized items and another Half to a target.

$$\arg\max_{\mathrm{Target}}\left[\sum_{i}^{N-1} \mathrm{Target} \cdot \mathrm{Item}_i\right]$$

$$\sum_{i}^{N-1} \mathrm{Target} \cdot \mathrm{Item}_i = \mathrm{Target} \sum_{i}^{N-1} \mathrm{Item}_i = \mathrm{Target}\,(\mathrm{TotalCapacity} - \mathrm{Target}) =$$

$$\mathrm{Target} \cdot \mathrm{TotalCapacity} - \mathrm{Target}^2$$

$$\frac{d}{d\mathrm{Target}}\left[\mathrm{Target} \cdot \mathrm{TotalCapacity} - \mathrm{Target}^2\right] = \mathrm{TotalCapacity} - 2 \cdot \mathrm{Target}$$

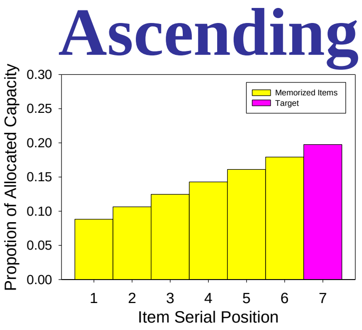
$$\mathrm{TotalCapacity} - 2 \cdot \mathrm{Target} = 0$$

$$\mathrm{Target} = \frac{1}{2}\mathrm{TotalCapacity}$$

The Target Locking Hypothesis

- Implication for non-optimal strategies, after the Half-Half rule→

Attentional gating should aim to allocate more capacity resources to the target than to memorized items.



The model

The Exemplar-Based STM Retrieval Model EBRW and the Item-Target Product Rule

Capacity Distributed via Attentional Gating

N = number of items stored in STM + Target

$$\text{Item Capacity (Target Capacity)} = f(\mathrm{N}) = -\frac{2\,(\text{Intercept}\cdot \mathrm{N} - \text{TotalCap})}{\mathrm{N}\,(1+\mathrm{N})}$$

$$\text{Assuming limited capacity} = \int_N f(N, Intercept) = \text{TotalCap}$$

EBRW

Distance

$$d_{ij} = \left[\sum_{k=1}^{K} w_k \left|x_{ik} - x_{jk}\right|^r\right]^{\frac{1}{r}},$$

d_{ij} = fixed value distance (free paramter)

$d_{ij} = 0$, on positive match

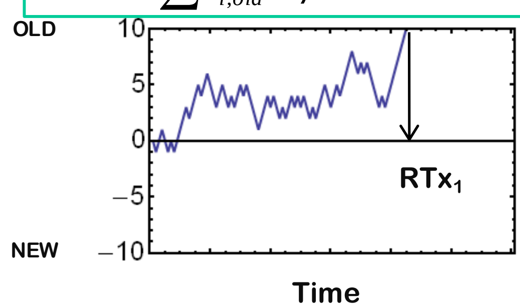
Activation*

$$s_{ij} = \underbrace{\exp\left(-c\cdot d_{ij}\right)}_{\text{Similarity}} \cdot \underbrace{\text{Item Capacity}_i \times \text{Target Capacity}_j}_{\text{Capacity via Attentional Gating}}$$

Stepping Probability

$$p_i = \frac{\sum s_{i,old}}{\sum s_{i,old} + \beta}$$

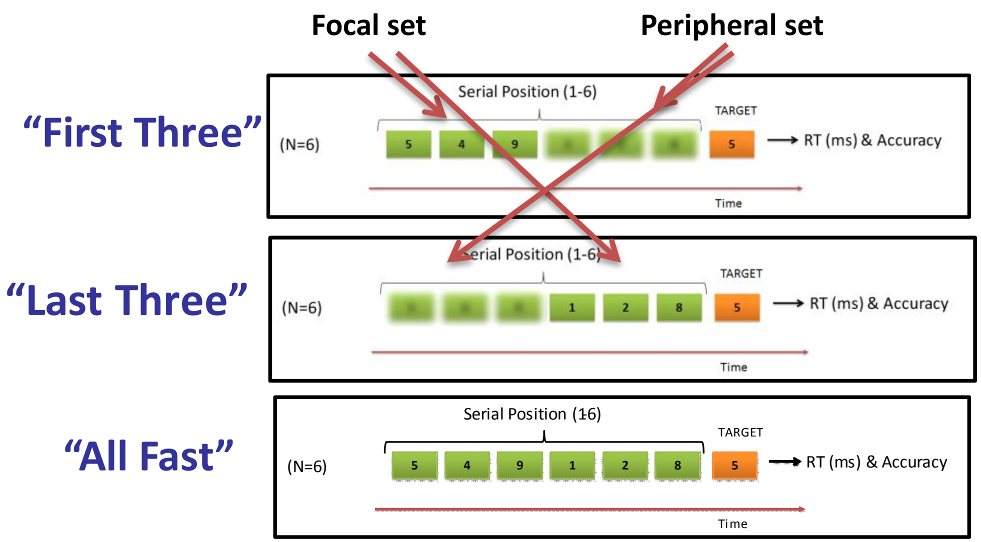
Random Walk



e.g. Nosofsky, Little ,Donkin, & Fific, 2011

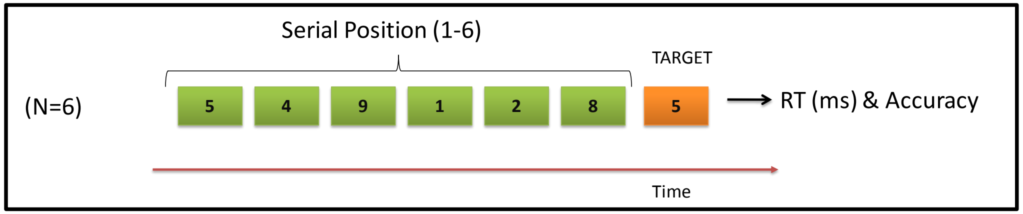
New Method
The attention-to-position paradigm

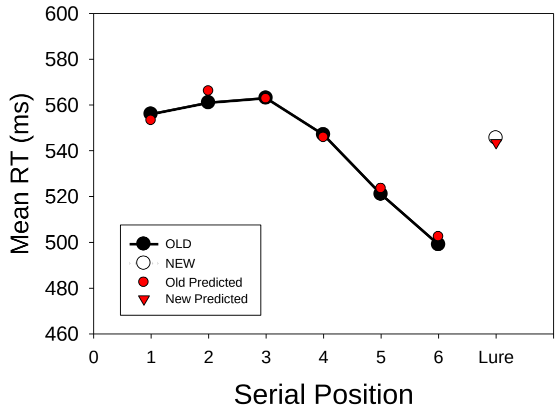
- Rapid short-term memory paradigm
- **Focal set** : To pay special attention to certain item positions in the memorized list, called a “focal set”. This means that if a target item was a member of a focal set, a response decision had to be extremely fast, and accurate
- **Peripheral set**: The rest items not contained in a focal set.

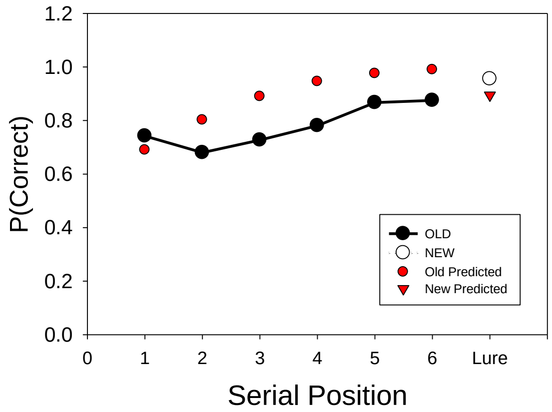


The data

"All Fast"







- This is a typical RT pattern observed in the STM research, the primacy and recency

~~(1) Equal-precision~~

~~(2) ALL-OR-NONE~~

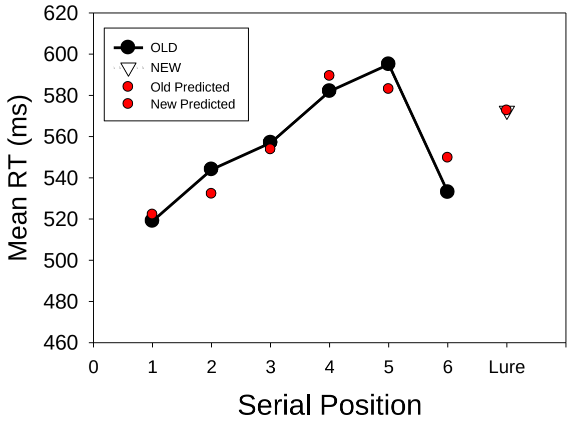
(3) The decay-representation WM model

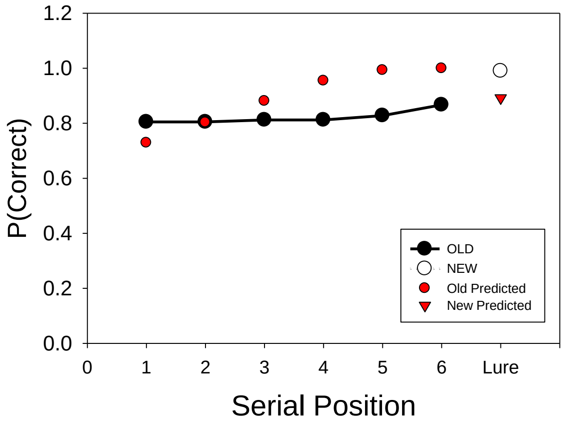
(4) Fluid-resource model

(5) Slots+averaging model

The data

"First Three"





- Not typical strong primacy RT effect

~~(1) Equal-precision~~

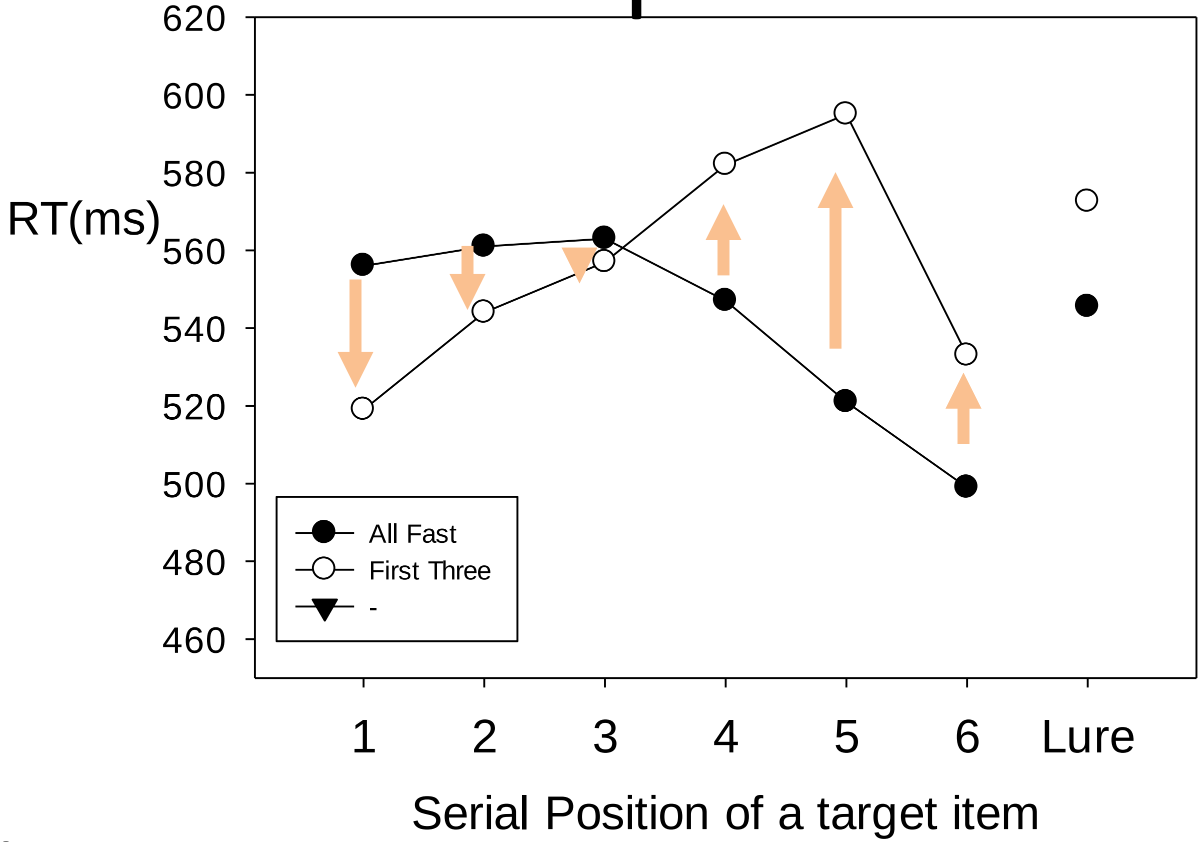
~~(2) ALL-OR-NONE~~

~~(3) Decay-representation WM model~~

(4) Fluid-resource model

(5) Slots+averaging model

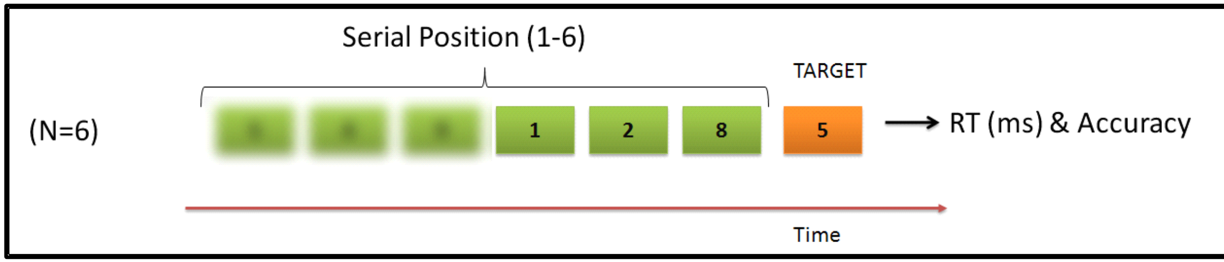
Comparison

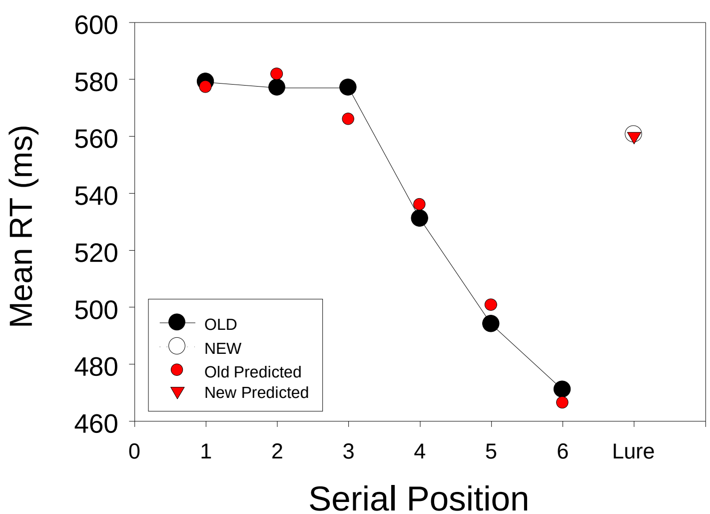


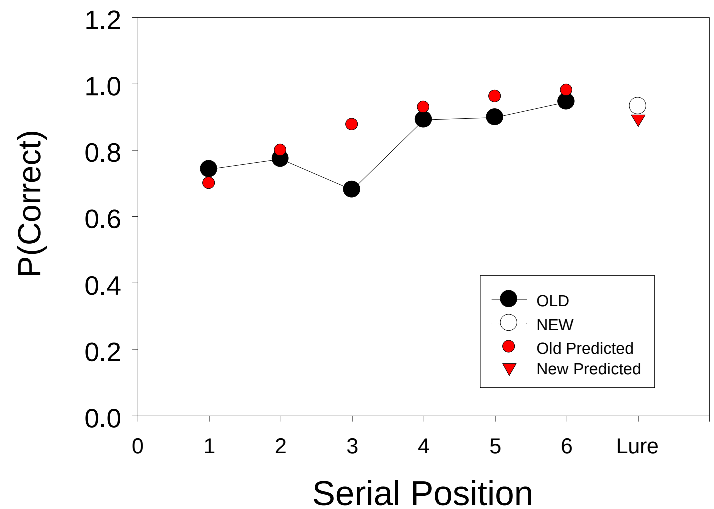
- Implications
  - A discontinuous serial position effect -> Dual WM systems
  - The principle of resource conservation-> Strictly fixed capacity

The data – further validation of resource allocation

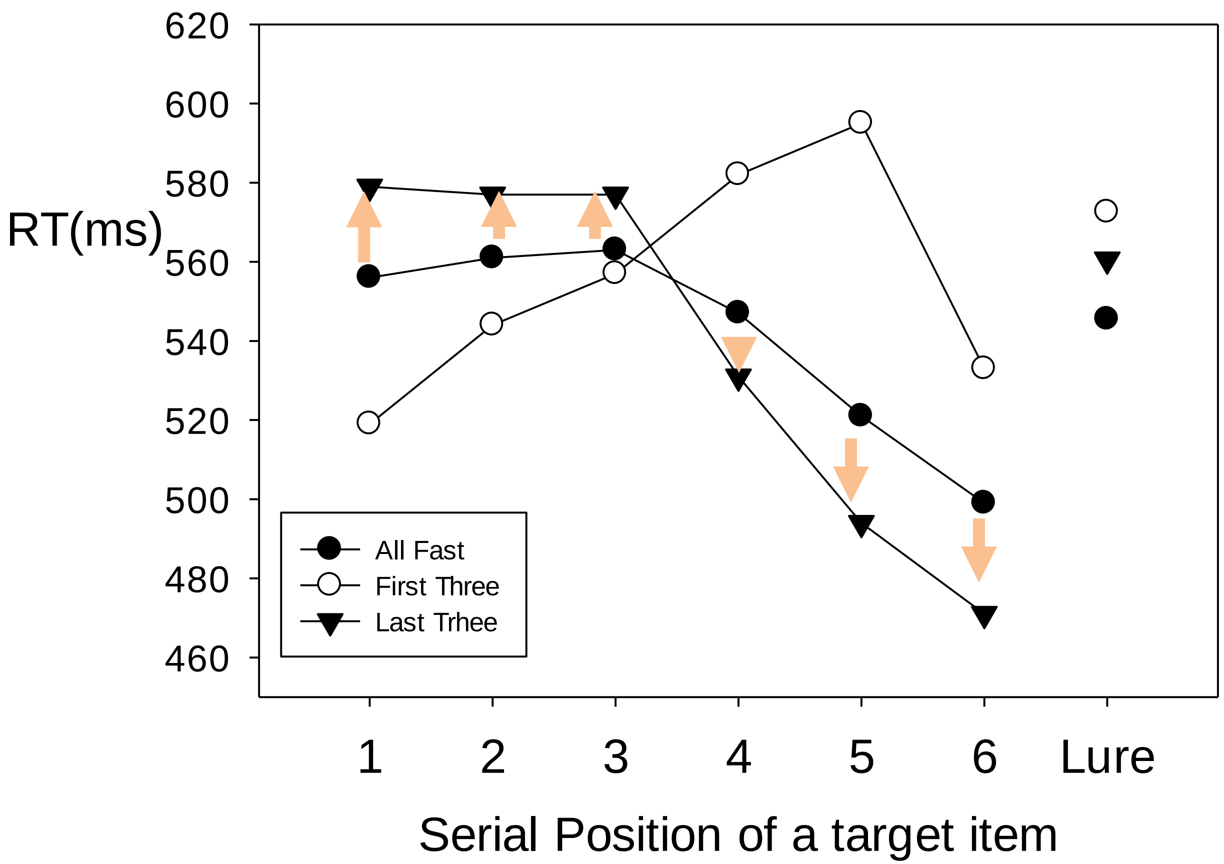
“Last Three”





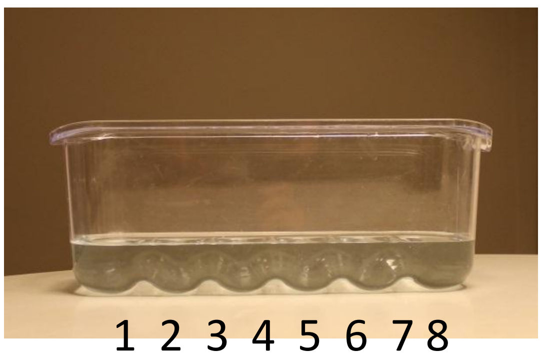


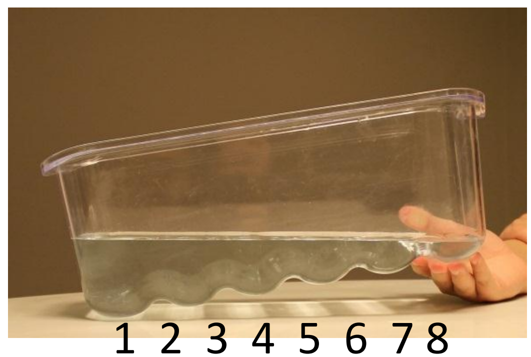
Comparison

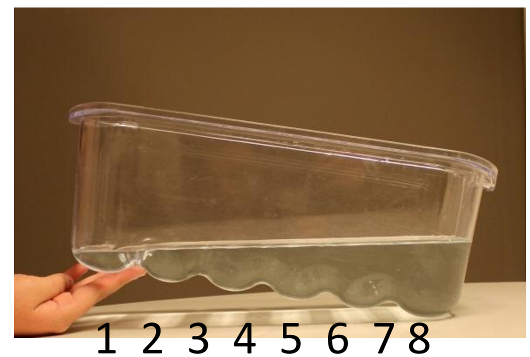


620
600
580
560
540
520
500
480
460
RT(ms)
All Fast
First Three
Last Trhee
1
2
3
4
5
6
Lure
Serial Position of a target item

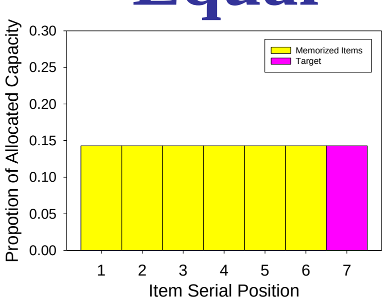
The proposed resource allocation model :The Tilted Water Tank



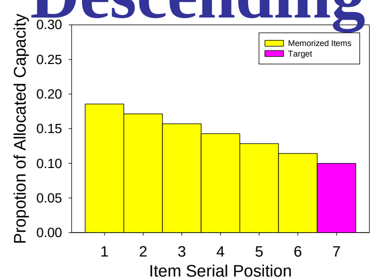




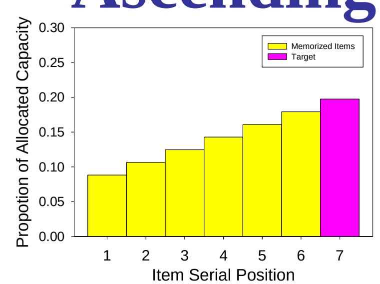
Equal



Descending

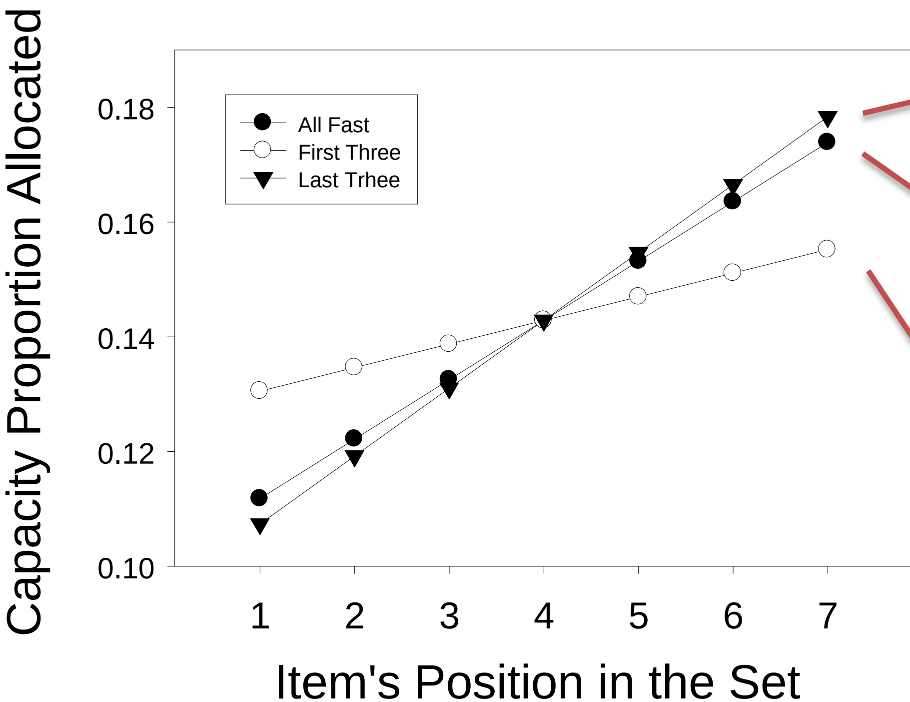


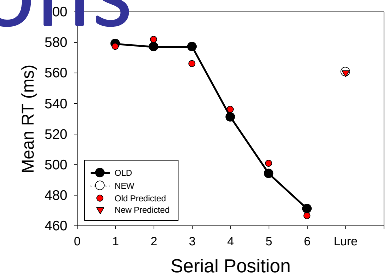
Ascending

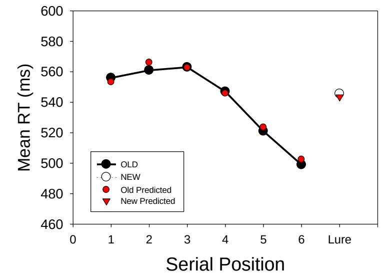


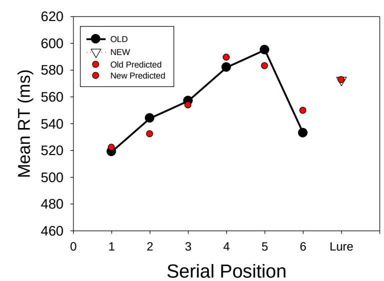
Model fitting: the linear distribution function of resource allocation

Estimated Capacity Allocations







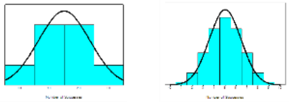


discrete-slot model

slots+averaging model

?

variable-resources model



Lim f(x)
n->Infinity

How many boxes?

- Conduct data fitting of the EBRW model that can freely allocate fixed amount of resources across memorized items, including the parameter which defines a number of possible memory slots (boxes).
- In other words: find <u>the number</u> of possible resource allocation units (slots, boxes) that maximizes the goodness of fit of the model for resource allocation.

How many boxes?

	Free resource parameter-EBRW mode				
Params	All Fast		First Three		Last Three
c	0.959		2.383		1.6
acrit	2.52		3.556		11.378
bcrit	3.936		43.592		17.164
scale	44.706		4.576		1.901
mu	175.816		137.859		256.035
listbase	0.175		0.261		0.42
dscale	2.705		0.979		0.65
m1	0.145		0.161		0.128
m2	0.131		0.166		0.131
m3	0.132	$\sum mi = 1$	0.167	$\sum mi = 1$	0.128
m4	0.161		0.170		0.176
m5	0.201		0.172		0.208
m6	0.231		0.164		0.229
boxes	**801**		**670**		**711**

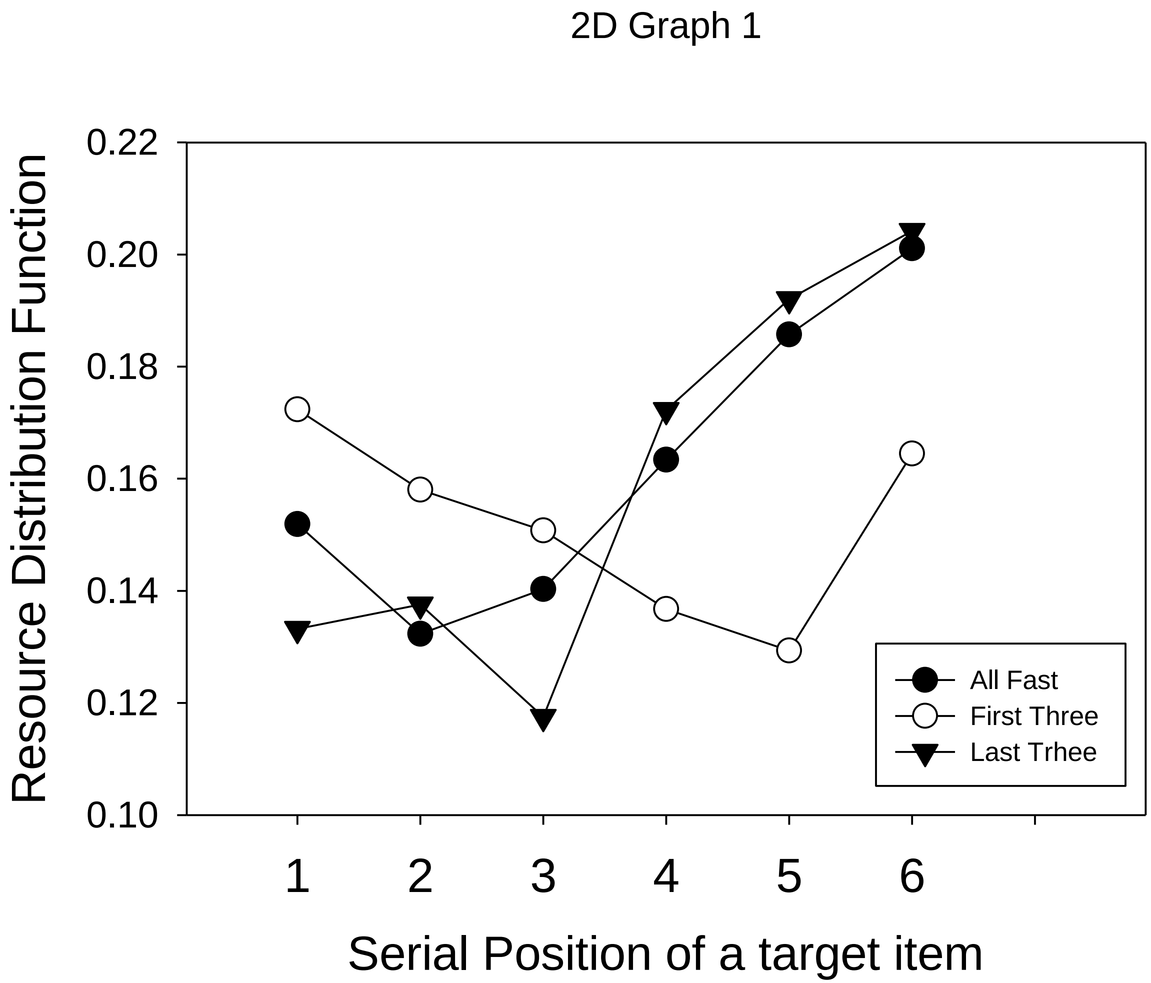
Conclusions

- New method for testing WM, attention by instruction
- Support for the variable-resources model WM, all-get-some
- Falsification of all-or-none approaches, discrete representations
- We specified a likely mechanism of resource allocation (Target locking) and provided rationale
- The ghost is likely to reside in a tilted tank!

Further Implications:

- Linear distribution function of resources could serve as a proxy to the Attentional Gating mechanism.
- Falsification of Dual system WM view: the last item position advantage
- A joint fit of mean RT and choice probabilities.[EBRW]
- A STM capacity resources are strictly limited (the conservation of resources principle)

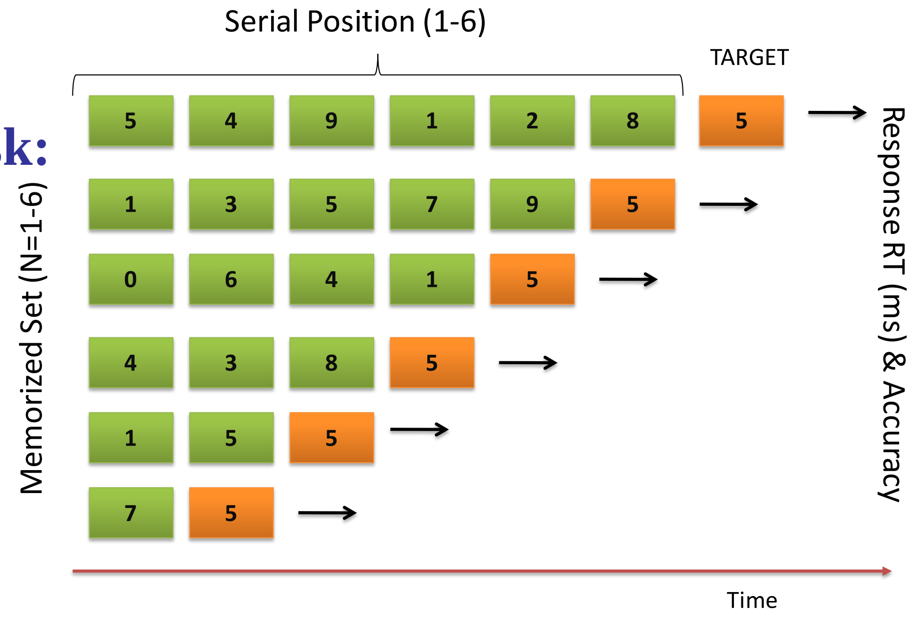
Free allocation of fixed capacity model



Tests & methods

A Short-Term Memory (STM) Retrieval Task

- The variant of the Sternberg task:
  Set size (1-6)
- Fast presentation rate (250msec)
- No target cuing



Tests & methods

Attention to Location Method

Manipulate subjects' temporal distribution of attention across items in the list

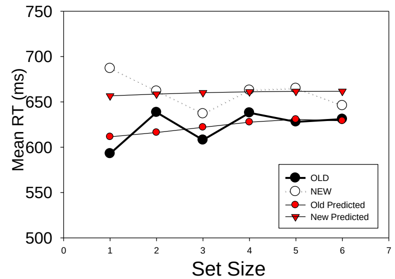
Focus on the first items in the list

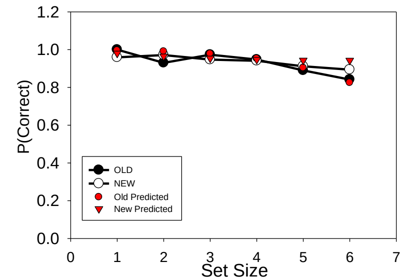
Focus on all items in the list

The data

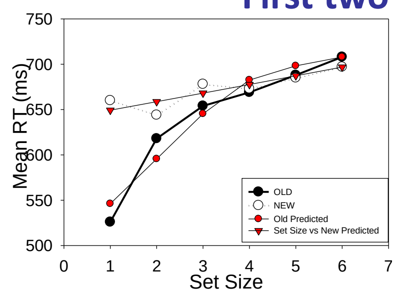
Attentional Gating and Set Size Effect

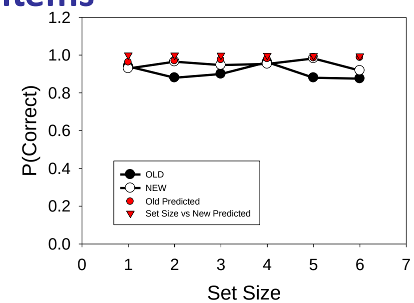
“No Target Cuing”



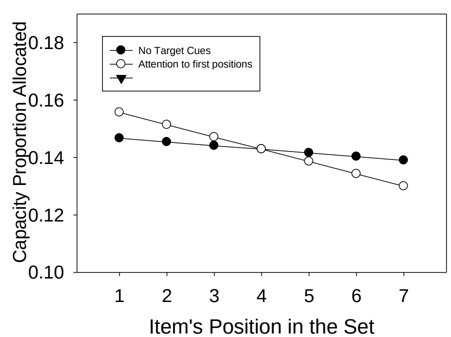


“First two items”





Estimated Capacity Allocations



Relationship between Short-term memory and Attention

Ahw & Jonides, 2001
Oberauer & Kliegl, 2006
Oberauer 2001; 2002; 2003
Garavan, 1998;
Reeves & Sperling, 1986

Relationship between Short-term memory and Attention

- **External Search**: when search items in a visual field attention uses short-term memory to mark the important spatial locations.

The idea:

- **Internal search**: when search items in memory (STM), the search mechanism uses attention to allocate processing resources to “mark”, or to lock, important memory locations.

Why status of mental representations in WM?

- Resource allocation.
- If representations are ALL-OR-NONE, and the system’s capacity is limited, then when there is information overload an operator must guess.
- Sophisticated guessing?
- Neural system’s implications.

New Method
The attention-to-position paradigm

- To prevent interference of extraneous variables with the process of resource allocation the subjects were instructed to pronounce each item in a set, without accentuation, and with a monotonic prosody

- Two measures: mean response time (RT) and accuracy.