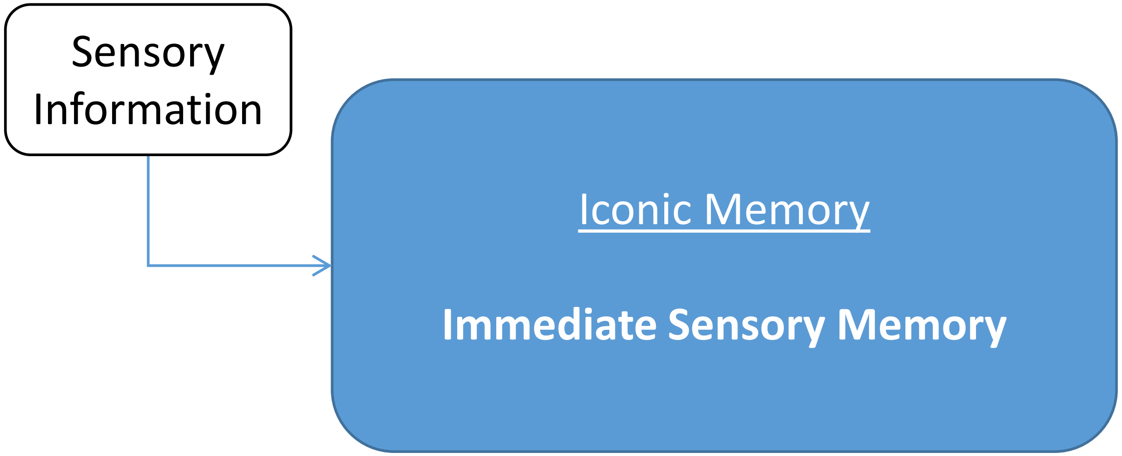
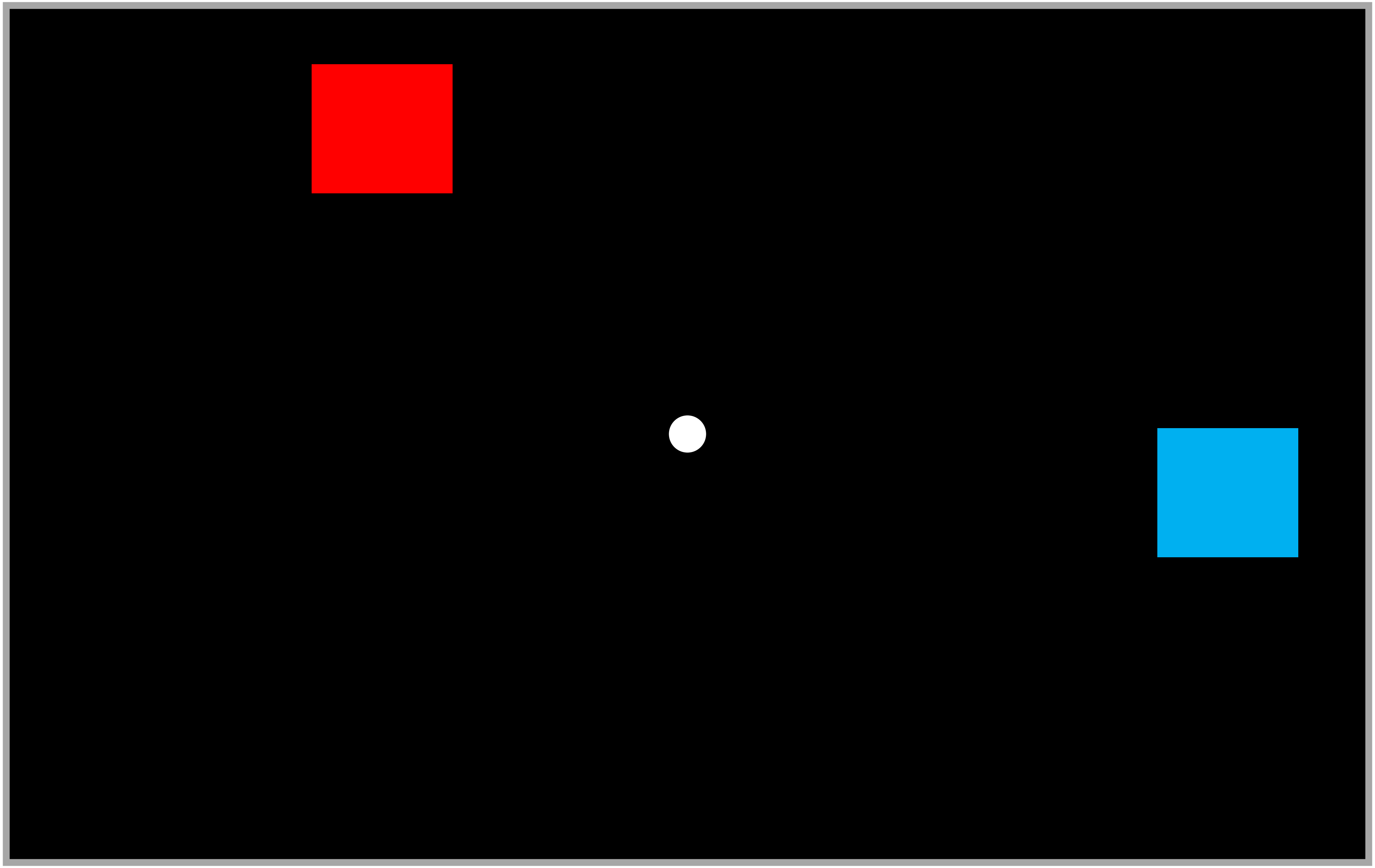
Working Memory

NEU/MOL 502A: **From Molecules to Systems to Behavior**

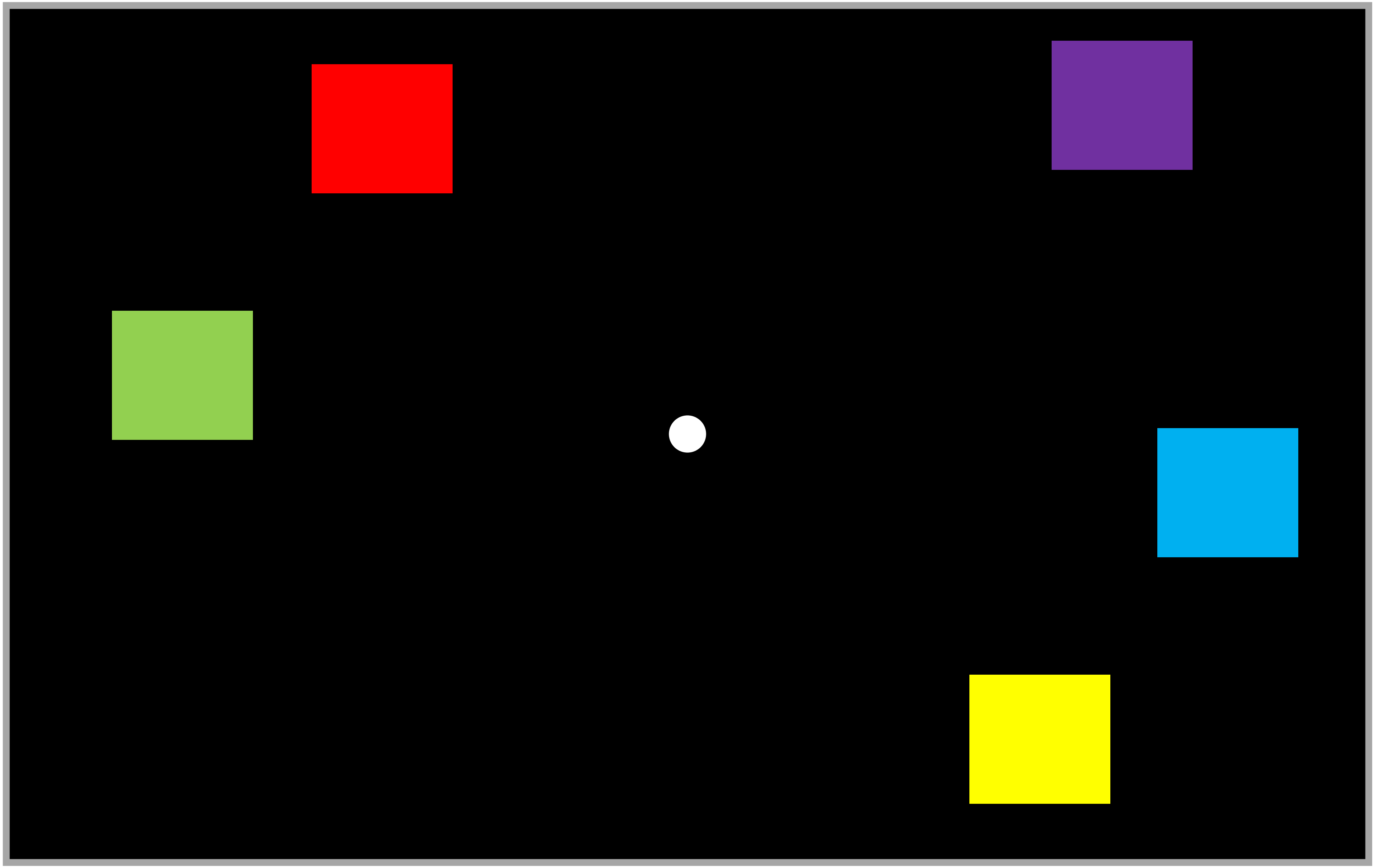
Different Scales of Memory



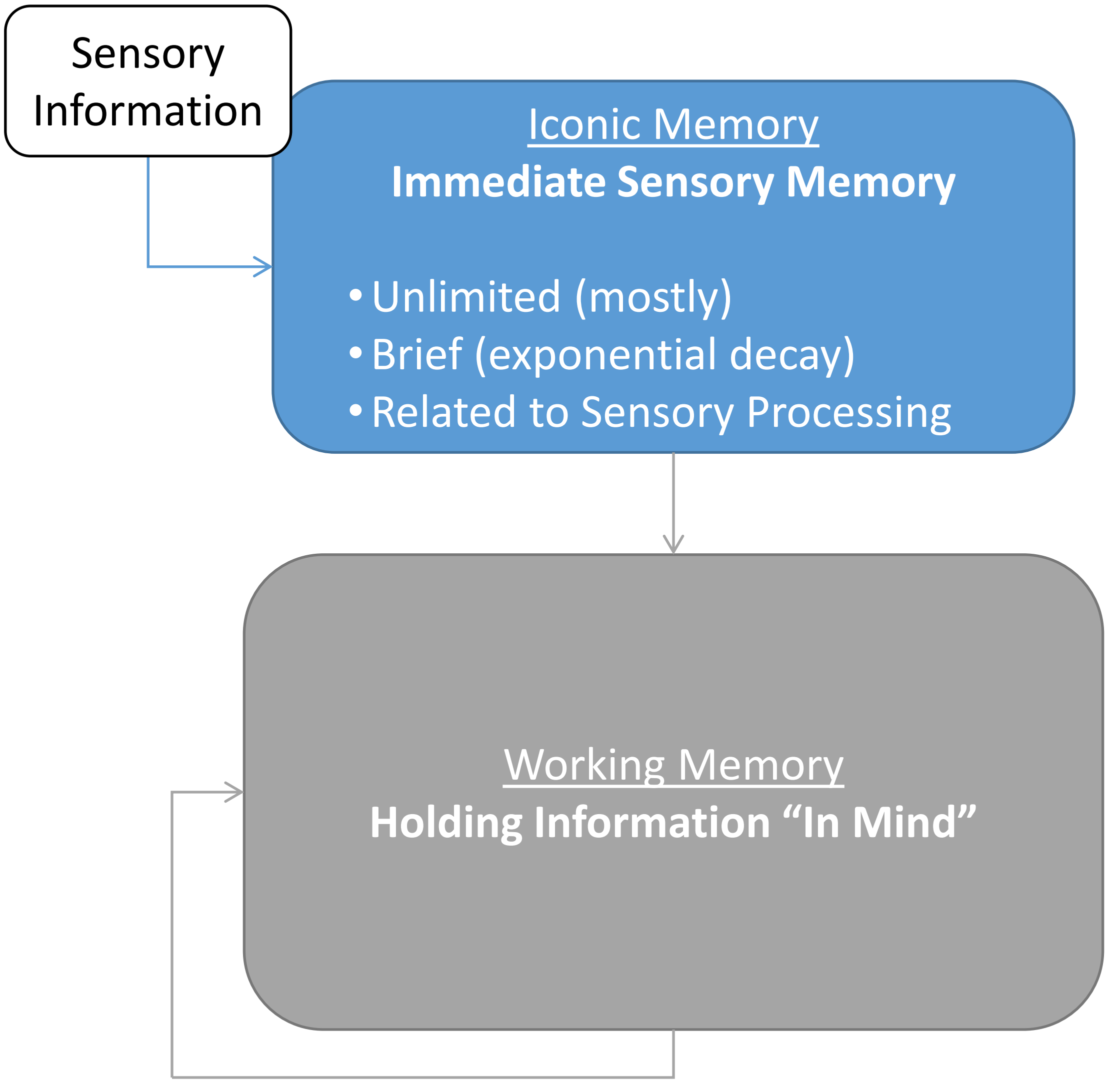
Example of Iconic Memory



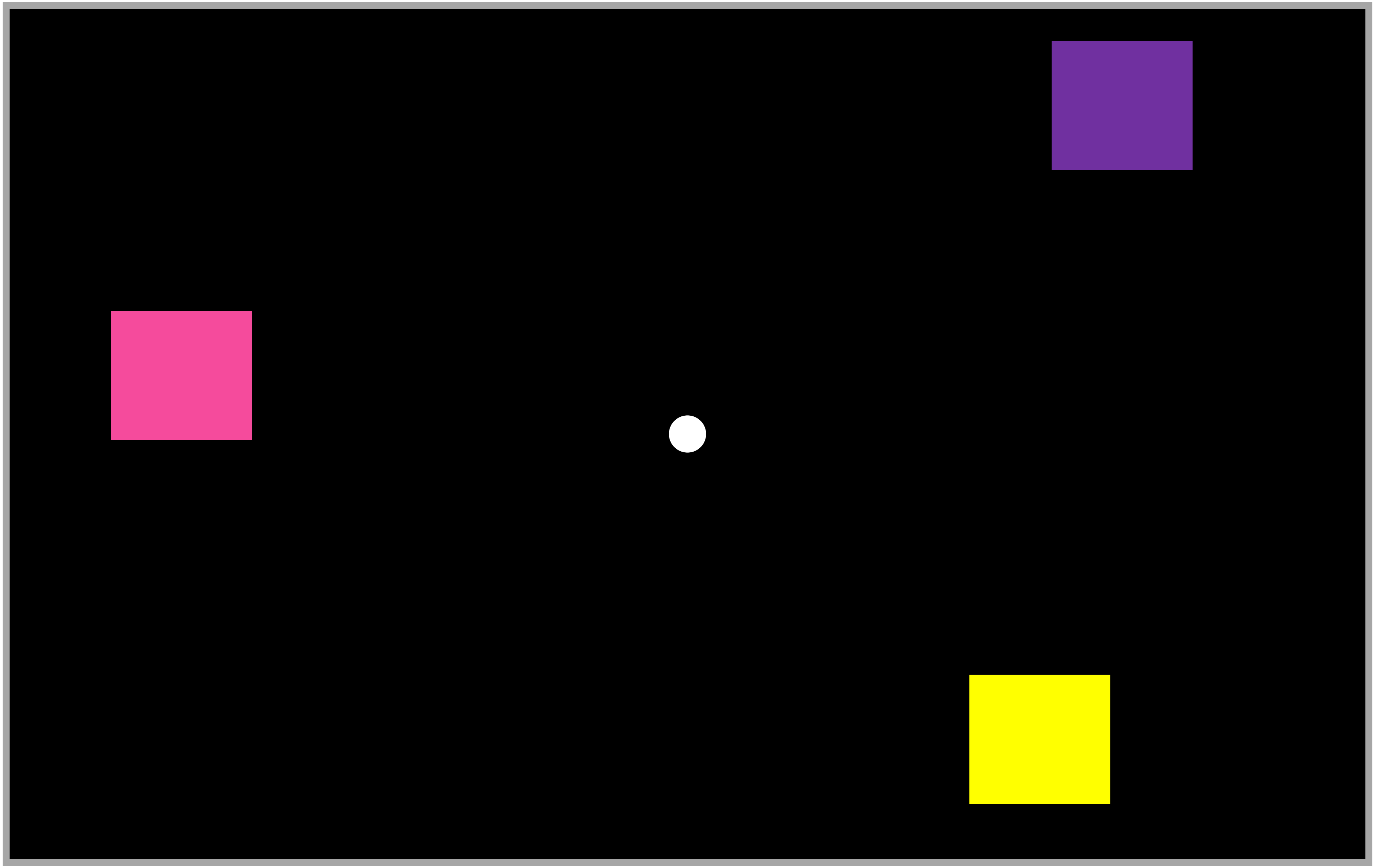
Example of Iconic Memory



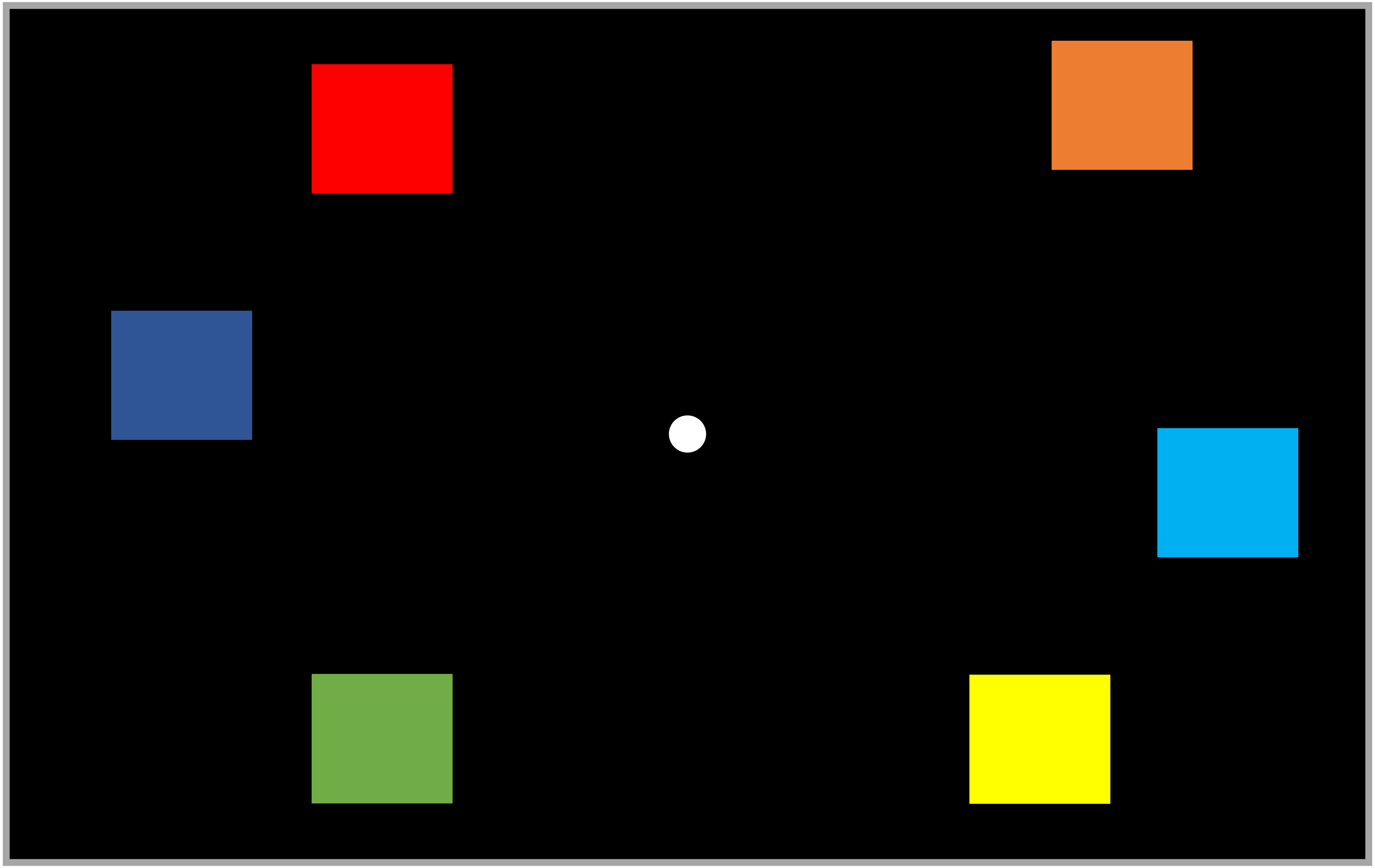
Multiple scales of memory in the brain



Working Memory



Working Memory



Multiple scales of memory in the brain

Sensory Information

Iconic Memory

Immediate Sensory Memory

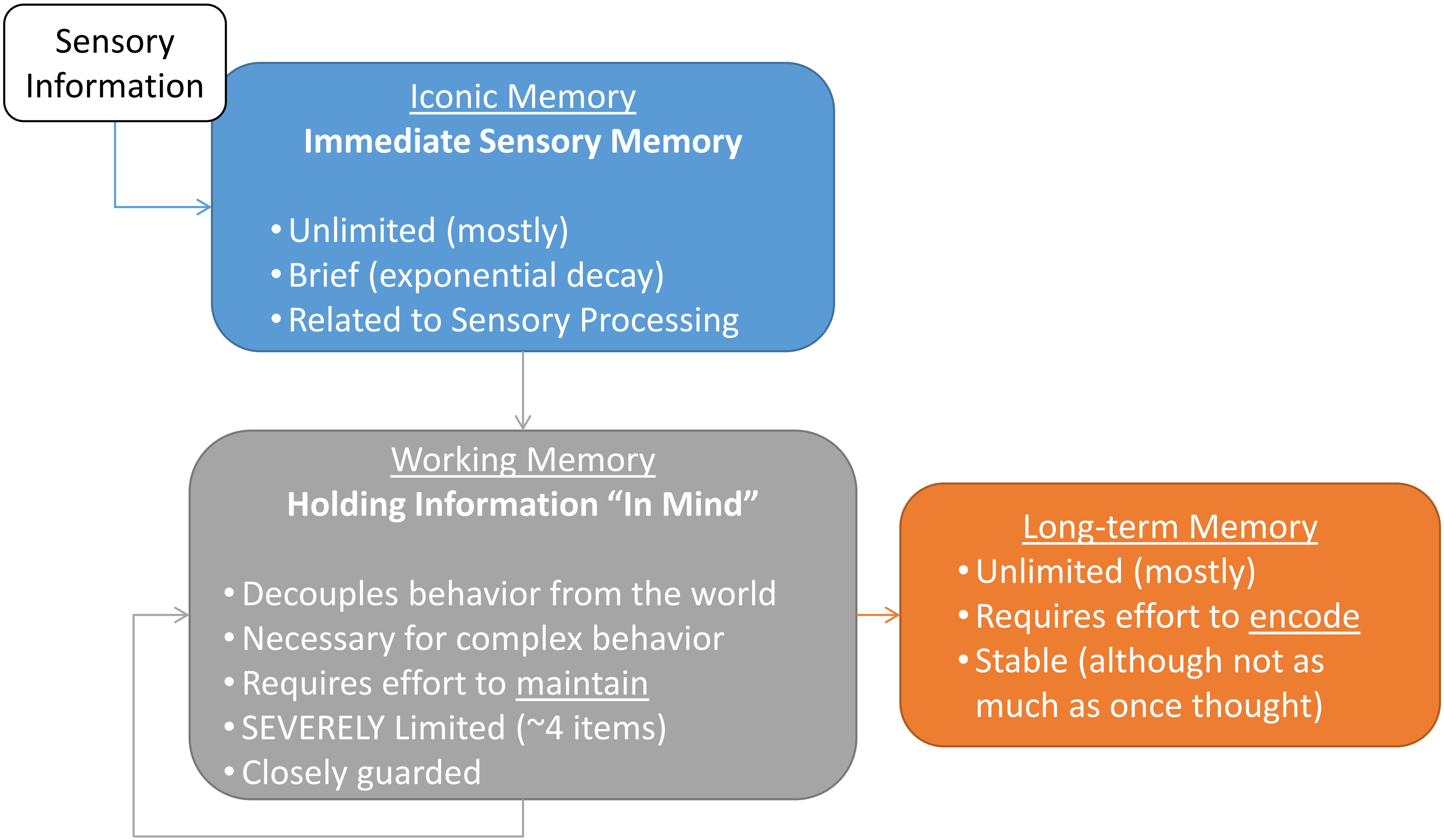
- Unlimited (mostly)
- Brief (exponential decay)
- Related to Sensory Processing

Working Memory

Holding Information "In Mind"

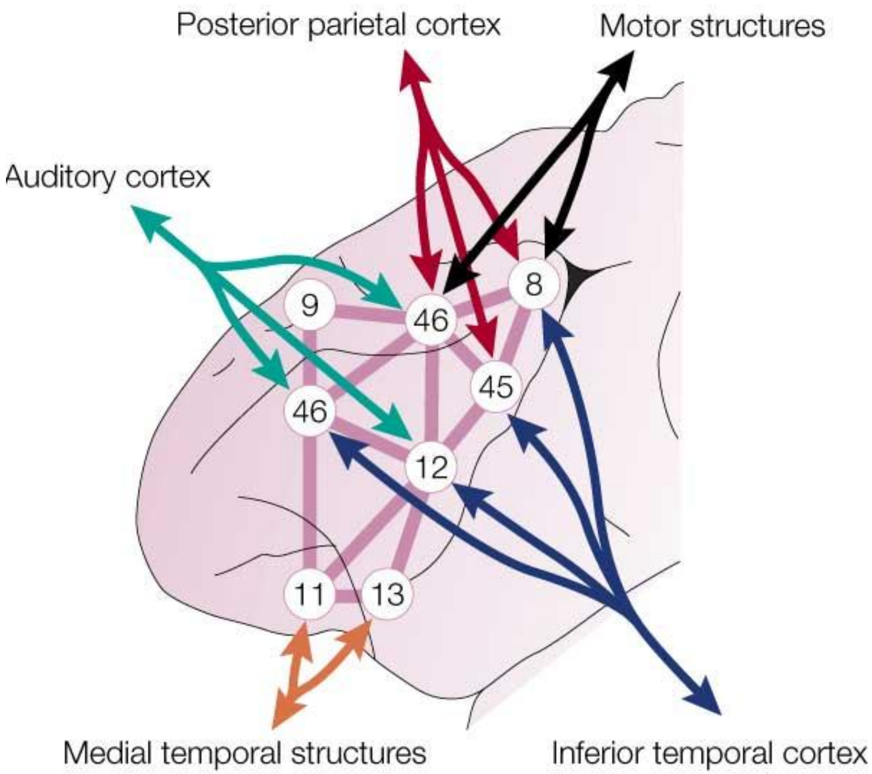
- Decouples behavior from the world
- Necessary for complex behavior
- Requires effort to maintain
- SEVERELY Limited (~4 items)
- Closely guarded

Multiple scales of memory in the brain



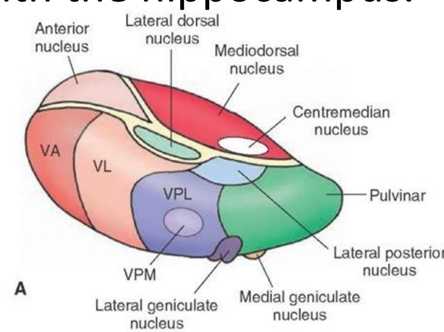
Neural Representations of Working Memory

Prefrontal Cortex is Densely Interconnected with Many Different Brain Regions

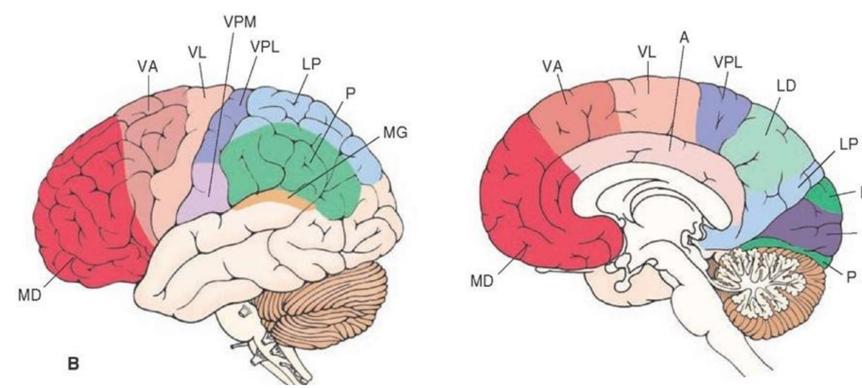


Prefrontal cortex is anatomically well situated to play a role in cognitive control.

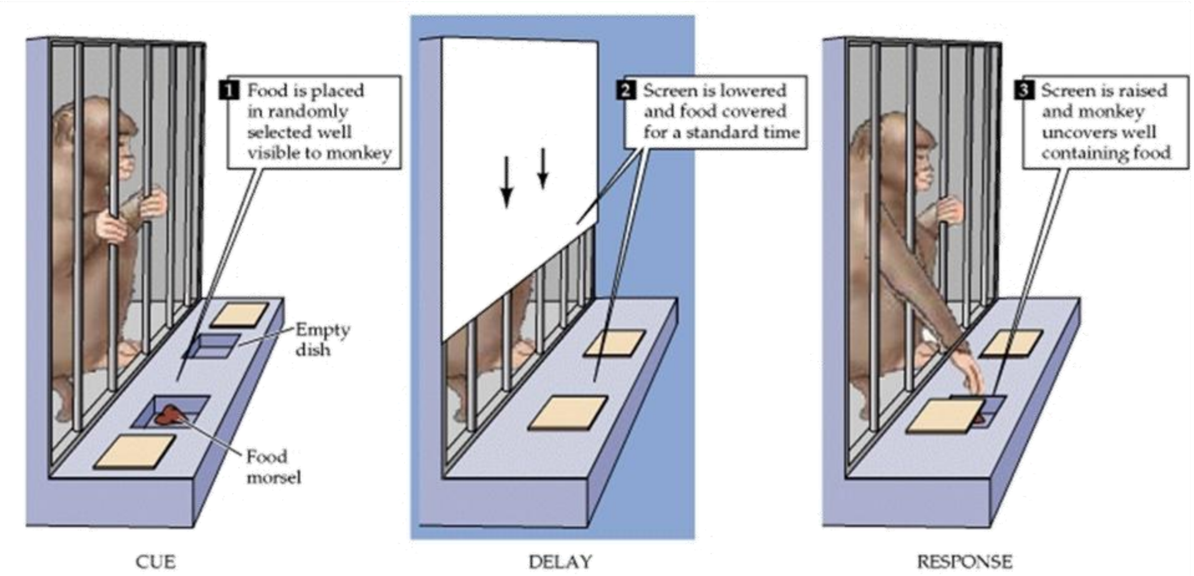
It is reciprocally connected with many different brain regions, including sensory, motor, and associative brain regions. It is also reciprocally connected with the hippocampus.

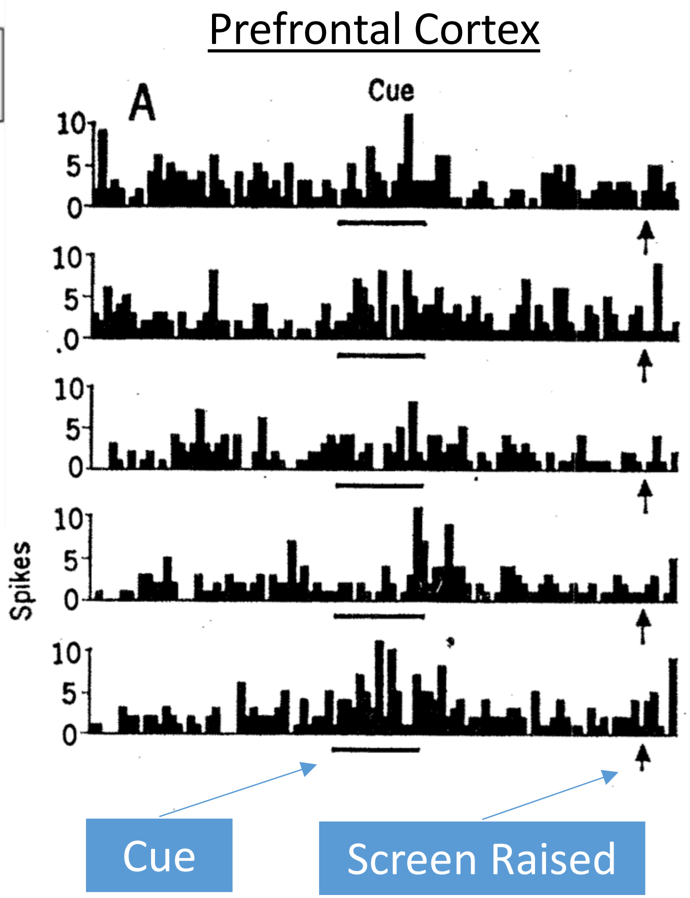


Prefrontal cortex is defined as the cortical region sending/receiving projections from the mediodorsal thalamus.



Delay Activity is the Hallmark of PFC Neurons





Monkeys were 'trained' to remember the location of a food reward over an extended delay.

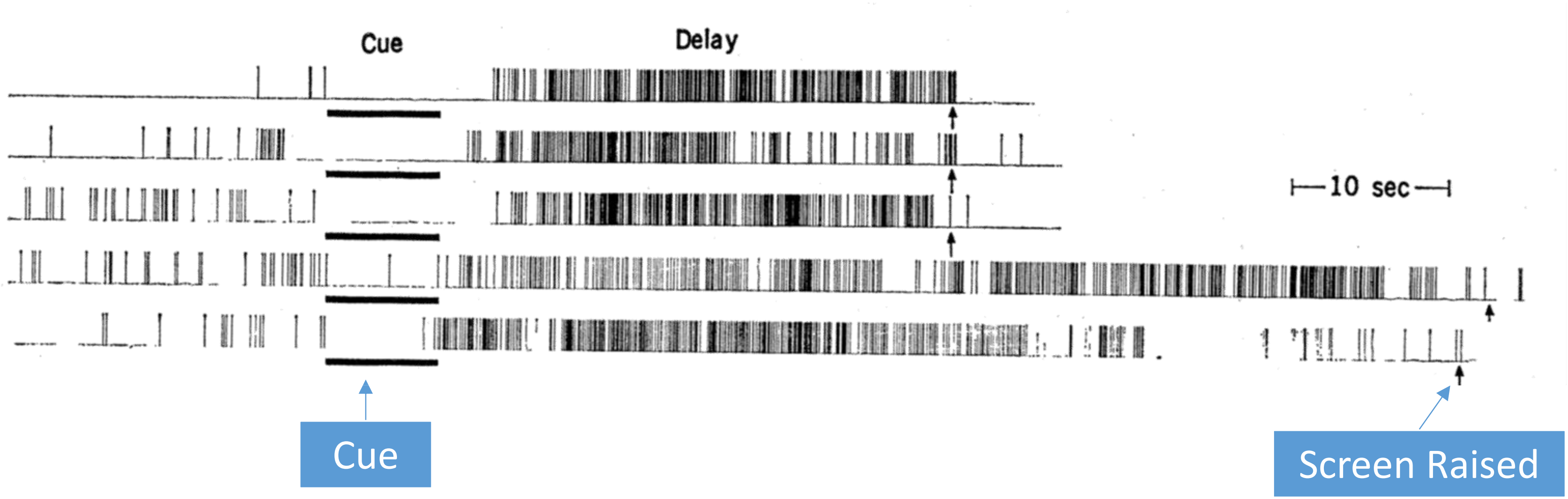
Delays ranged from 15 seconds to 60 seconds.

Fuster and Alexander, *Science* 1971

Delay Activity is the Hallmark of PFC Neurons

PFC neurons sustained activity over long delays, including up to several seconds. In this way, they could 'bridge the gap' between sensory stimulus and behavior.

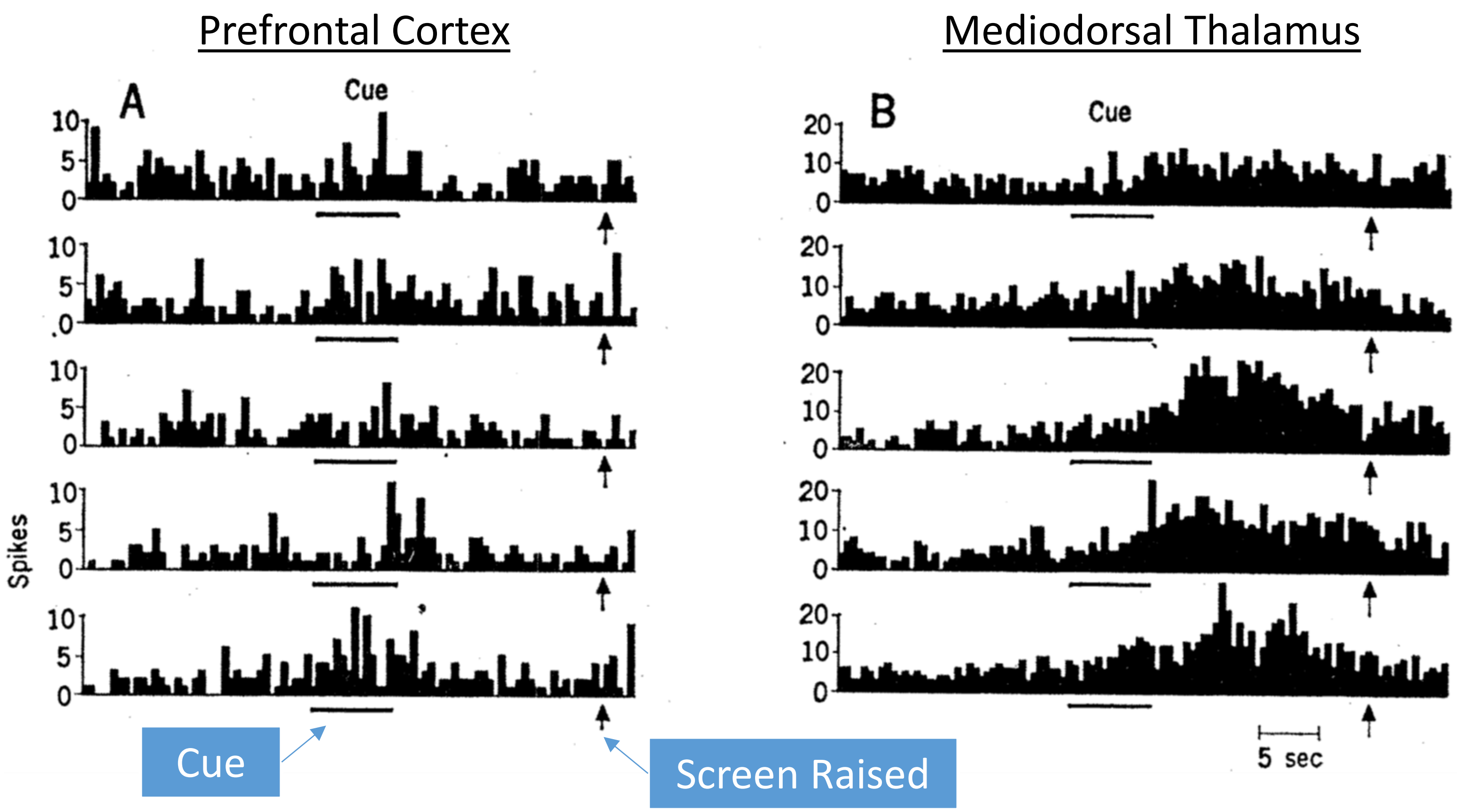
<u>Prefrontal Cortex</u>



Fuster and Alexander, *Science* 1971

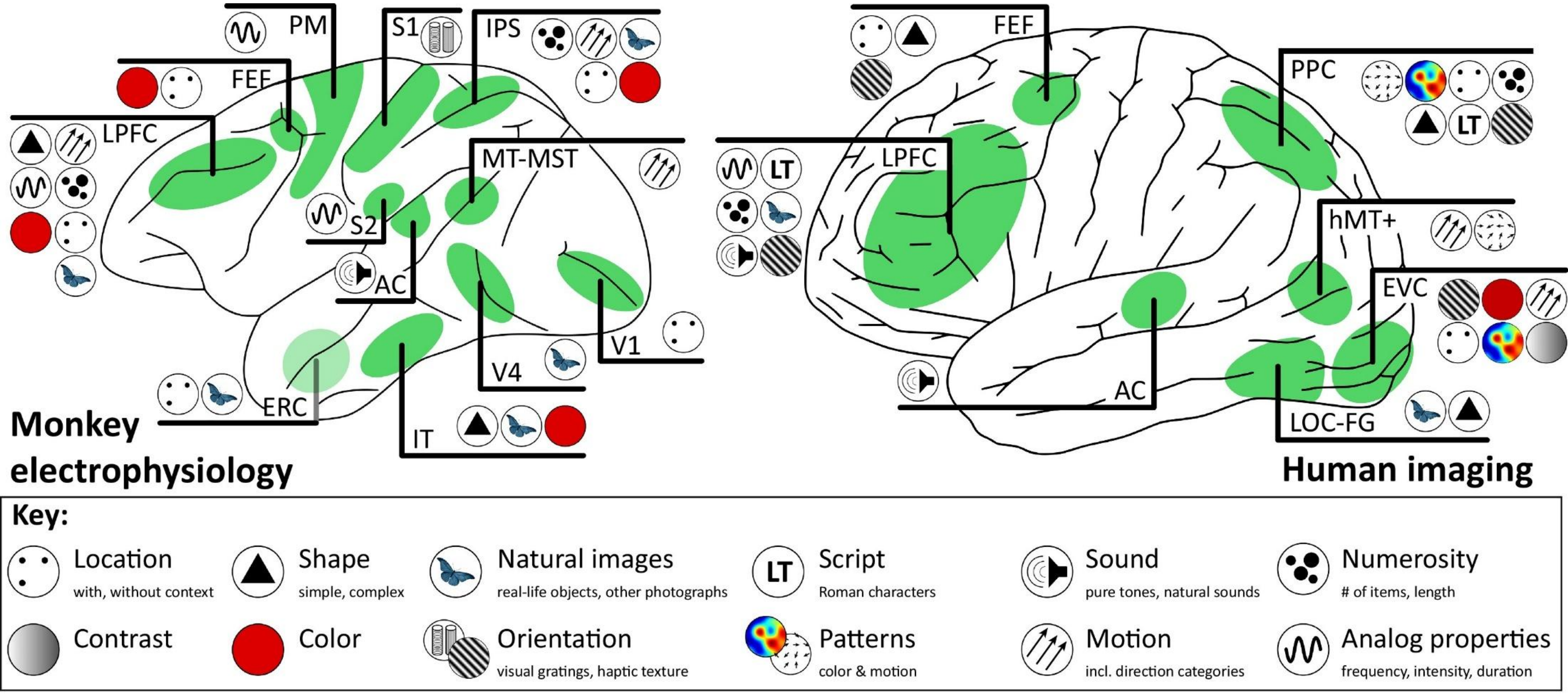
Delay activity is also seen in Mediodorsal Thalamus

MD neurons also show sustained activity over long delays. This is the basis for suggestions that recurrent loops between the prefrontal cortex and thalamus support sustained representations



Fuster and Alexander, *Science* 1971

Working memory representations are distributed across the brain



Trends in Cognitive Sciences

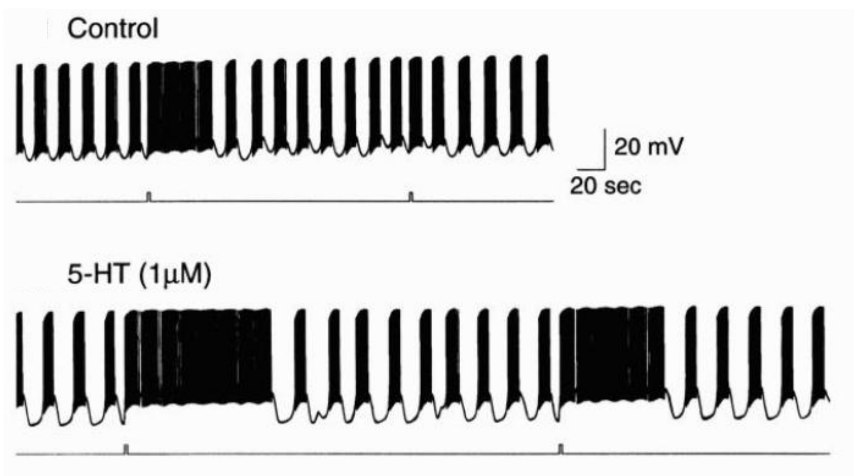
Note: It is controversial the degree of distribution in working memory representations. For example, Ranulfo Romo argues he sees no WM representations in S1.

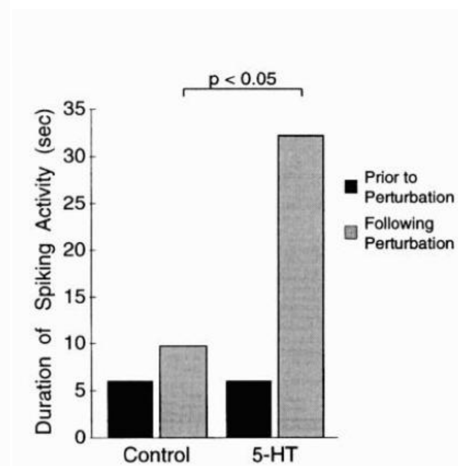
Brief Overview of Models of Working Memory

Models of Sustained, Mnemonic Activity

Neural mechanisms of persistent activity:

1) Changes in the biophysics of PFC neurons leading to tonic activity.

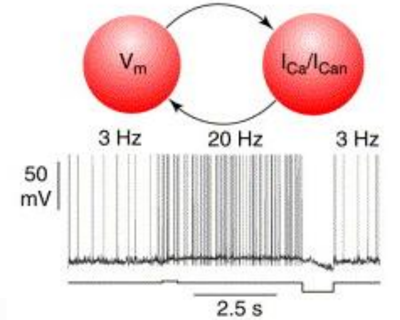


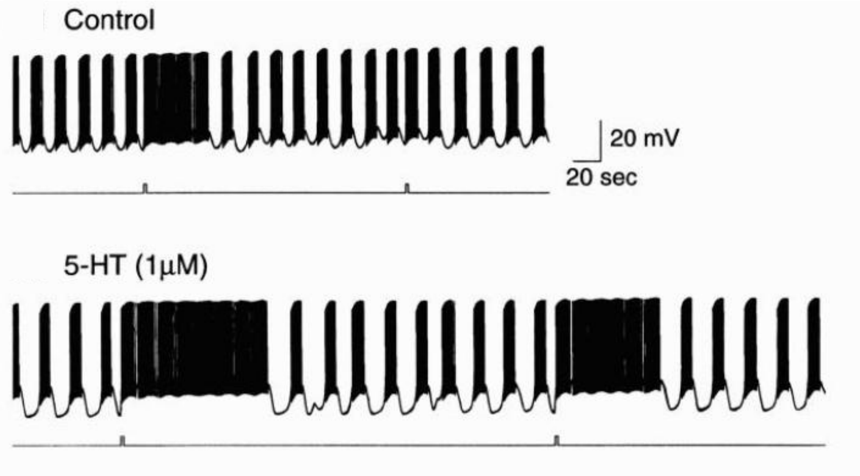


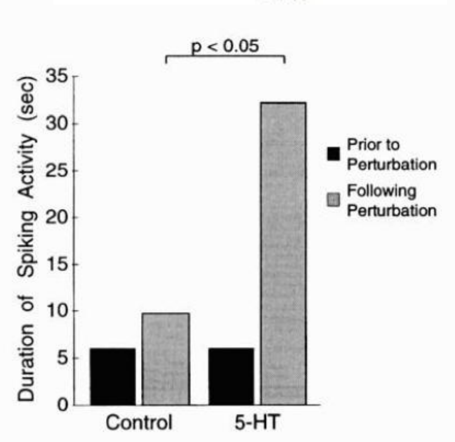
Models of Sustained, Mnemonic Activity

Neural mechanisms of persistent activity:

1) Changes in the biophysics of PFC neurons leading to tonic activity.





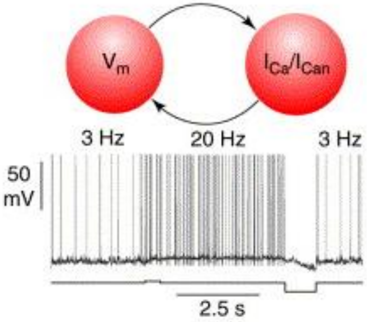


Models of Sustained, Mnemonic Activity

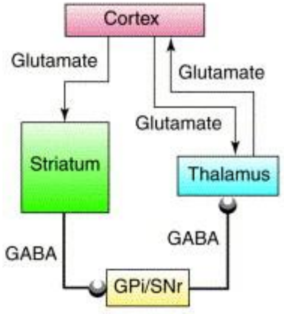
Neural mechanisms of persistent activity:

1) Changes in the biophysics of PFC neurons leading to tonic activity.

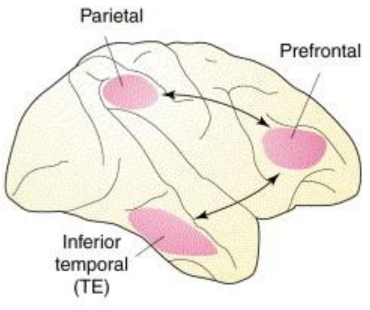
2) Reverberatory network mechanisms



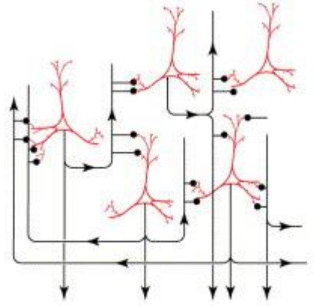
Cortical-Thalamic Loops



Cortico-Cortical Loops

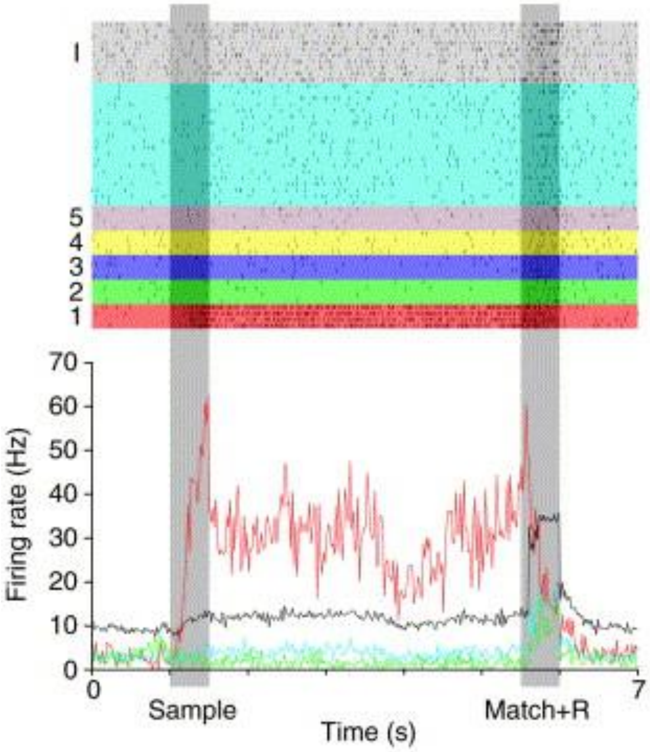


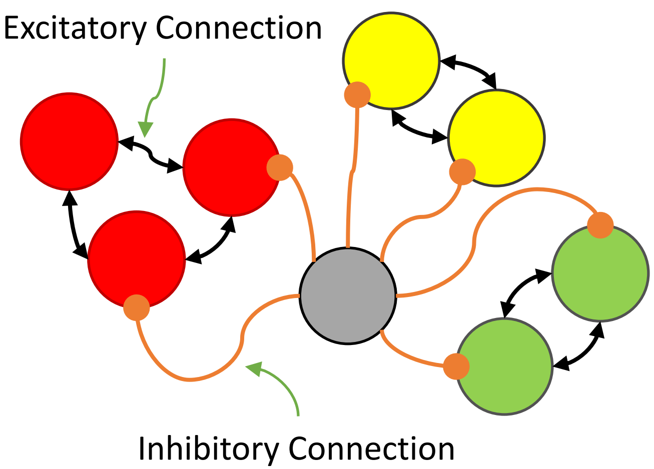
Local Cortical Circuits



Recurrent networks support mnemonic activity

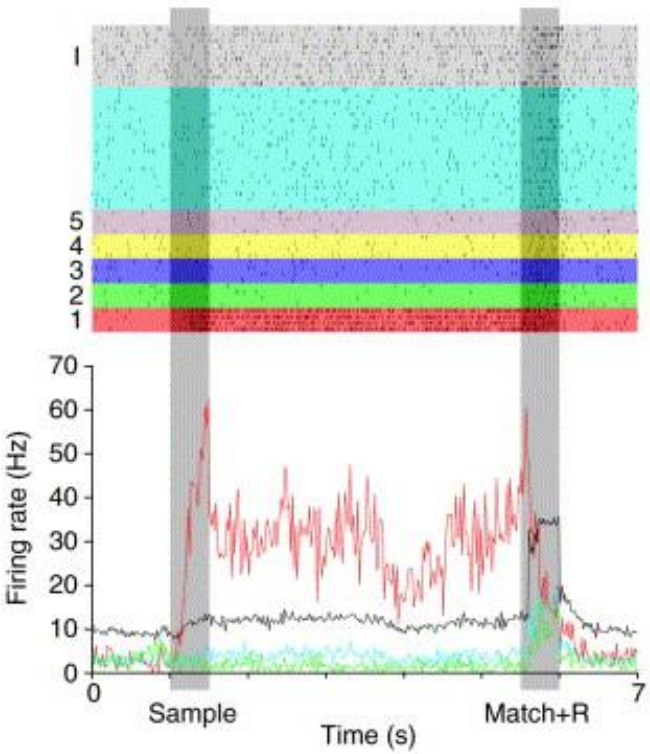
Discrete mnemonic representations:

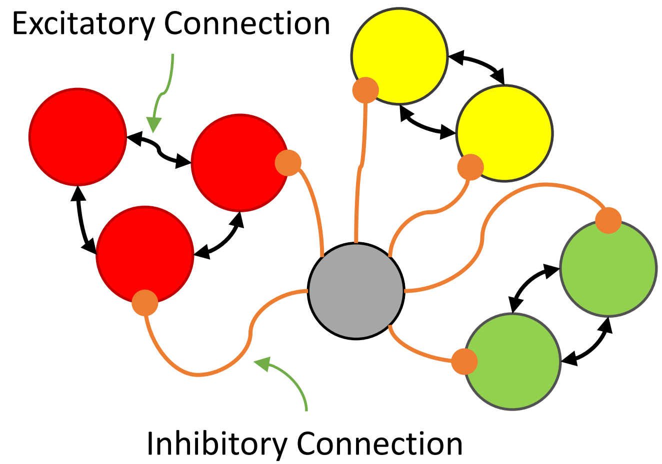


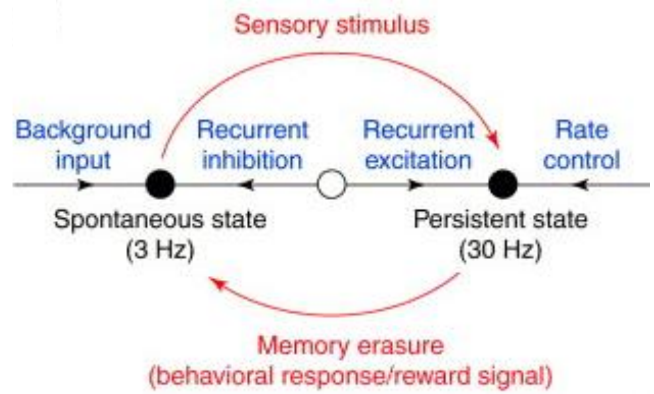


Recurrent networks support mnemonic activity

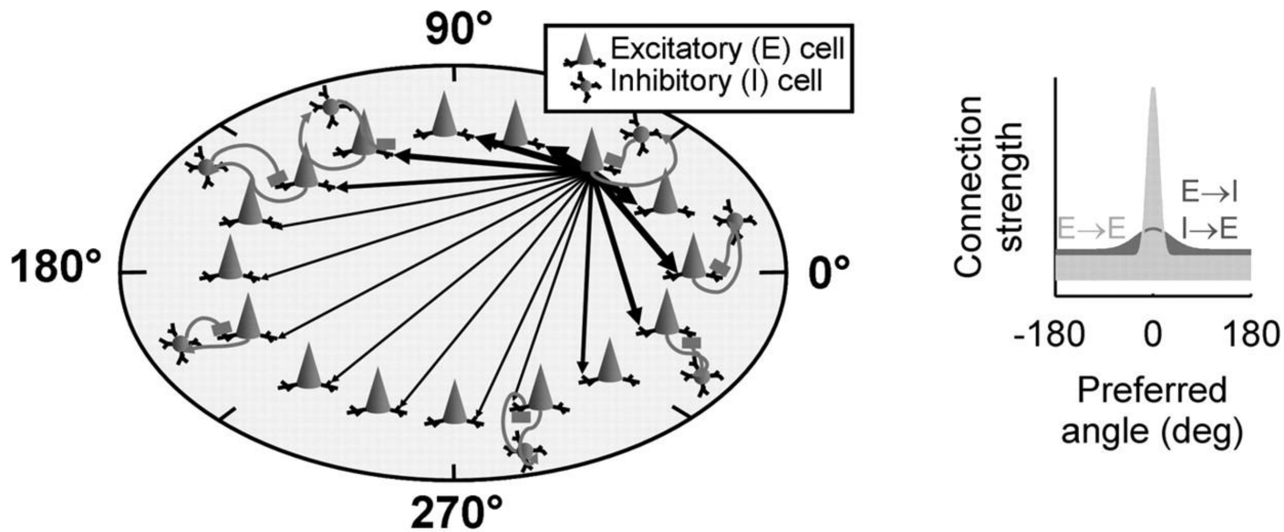
Discrete mnemonic representations:



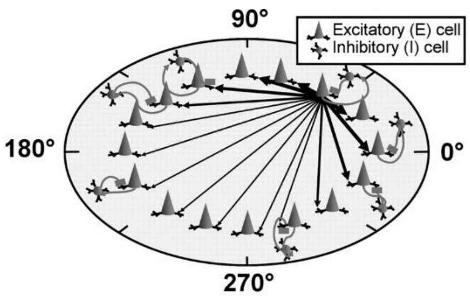


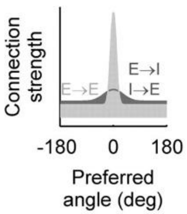


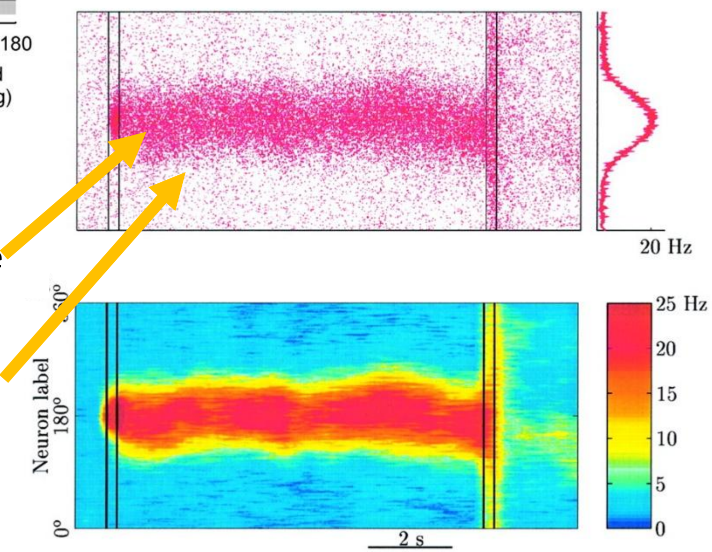
Construction of ‘bump’ attractor networks in PFC for maintaining representations



Construction of 'bump' attractor networks in PFC for maintaining representations





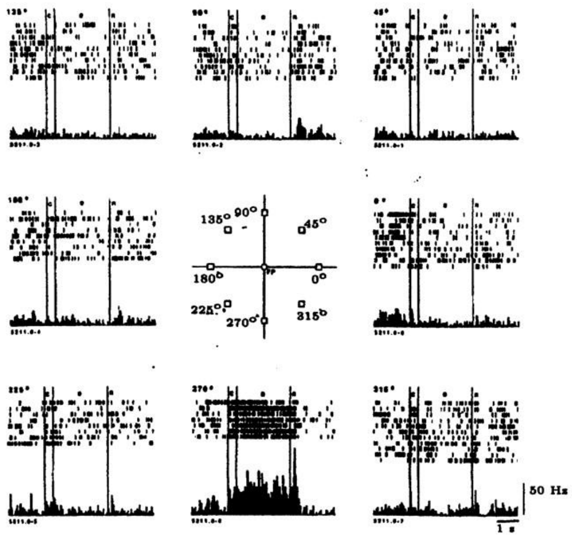


Re-current excitatory drive sustains initial activity

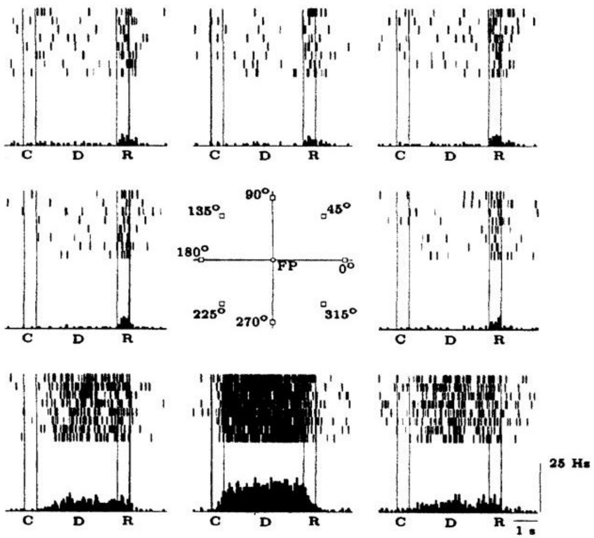
Surround inhibition keeps the bump from wandering

Bump attractors fit neurophysiological data

Delayed Saccade Task



Bump-Attractor Model

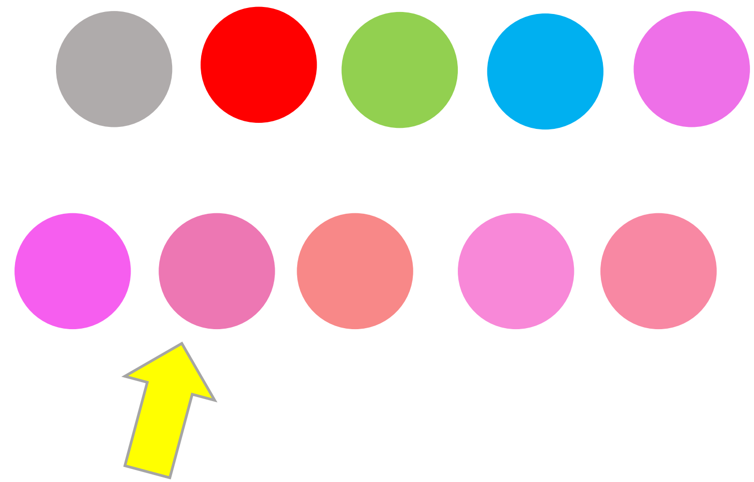


Flexibility of Working Memory

Working memory is incredibly flexible: you can hold anything in mind!

What animal was it?

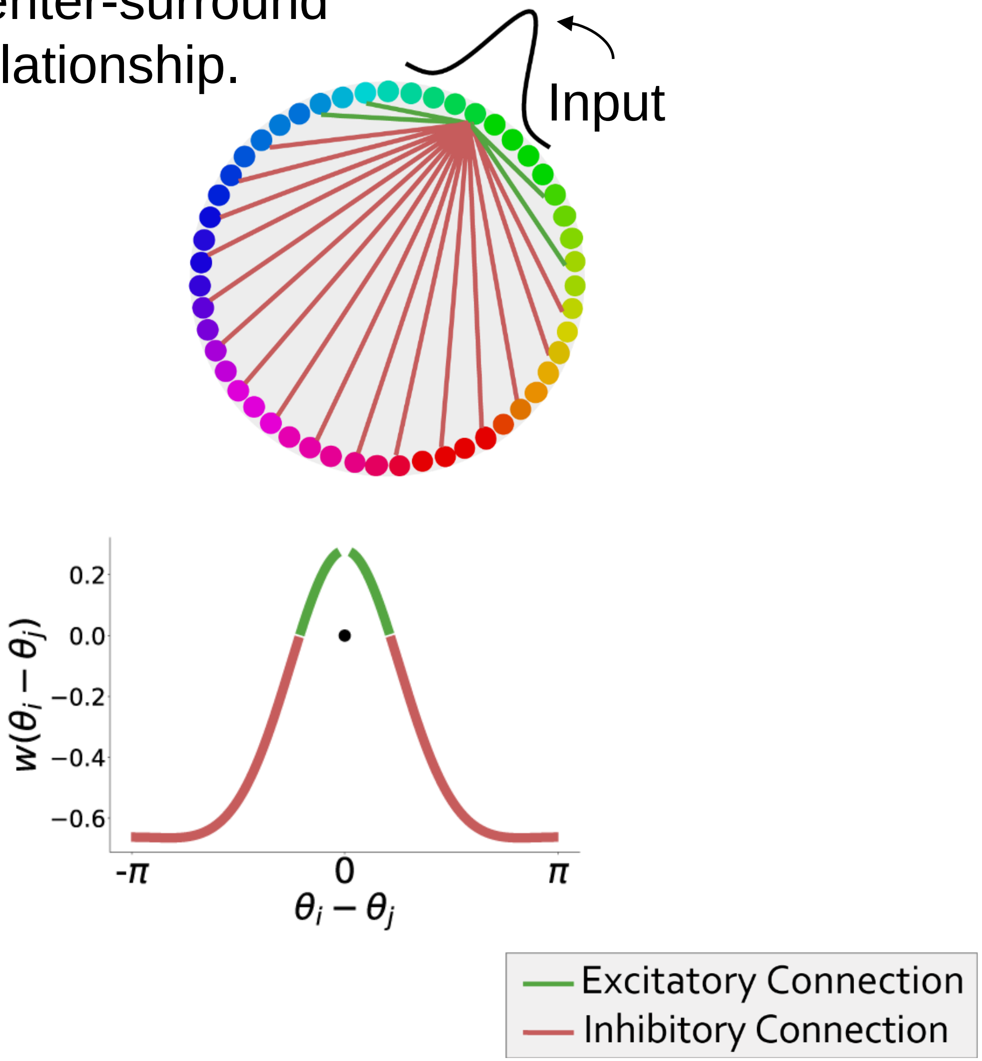
What color was the elephant?



What was it doing?

A simple model of random, reciprocal connections between a ‘random network’ and ‘sensory network’ can capture many working memory effects

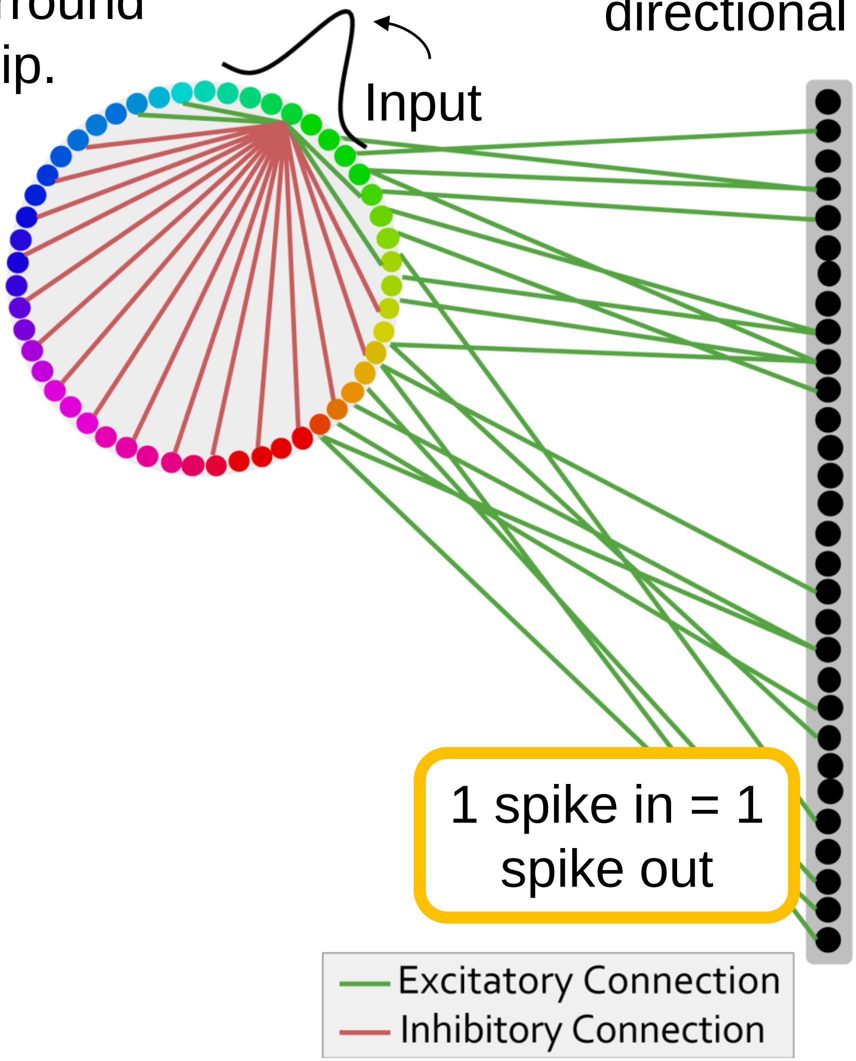
‘Sensory’ network with center-surround relationship.



A simple model of random, reciprocal connections between a 'random network' and 'sensory network' can capture many working memory effects

'Sensory' network with center-surround relationship.

'Random' network with *random*, bi-directional connections with sensory cortex.



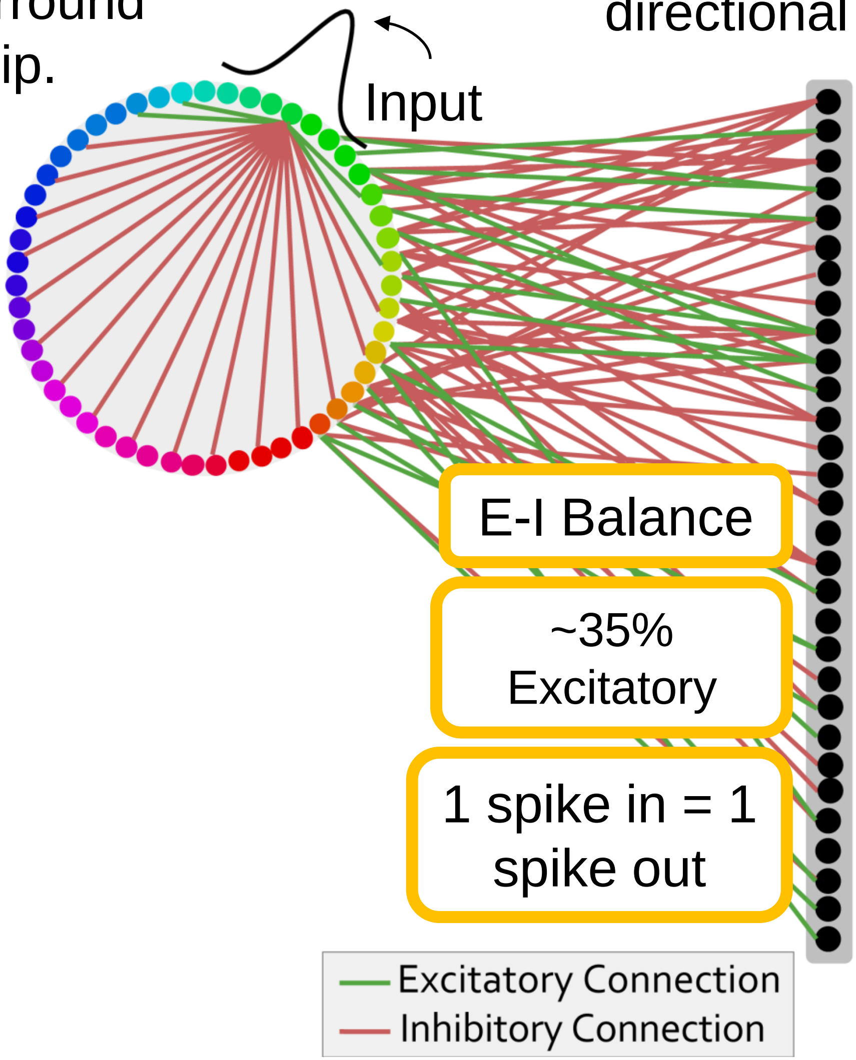
Sensory network cannot maintain representations alone; maintained only through interactions with random network.

Connection weight between sensory and random network is tuned such that 1 input spike results in 1 output spike (on average). This is the only tuned parameter.

A simple model of random, reciprocal connections between a 'random network' and 'sensory network' can capture many working memory effects

'Sensory' network with center-surround relationship.

'Random' network with *random*, bi-directional connections with sensory cortex.



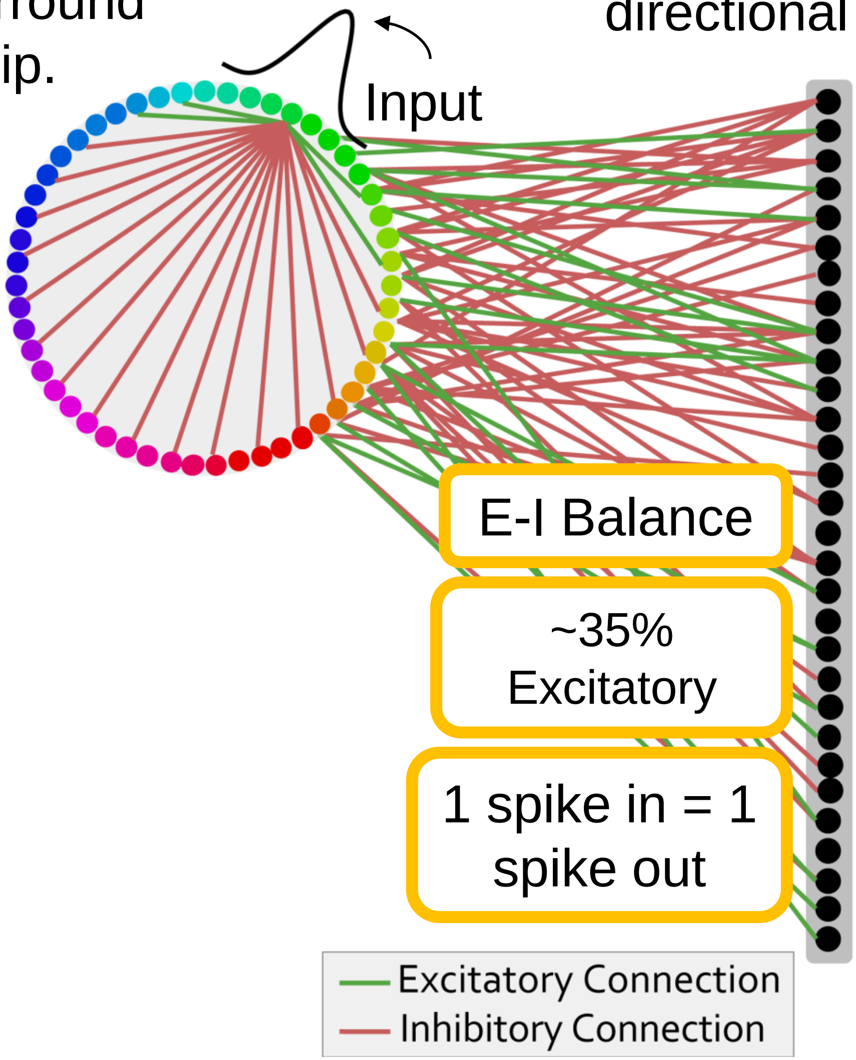
Roughly 35% of connections between sensory and random networks are excitatory. Remaining connections are inhibitory.

Neurons receive balanced excitation and inhibition. Thus, every neuron, on average, receives zero net synaptic input.

A simple model of random, reciprocal connections between a ‘random network’ and ‘sensory network’ can capture many working memory effects

‘Sensory’ network with center-surround relationship.

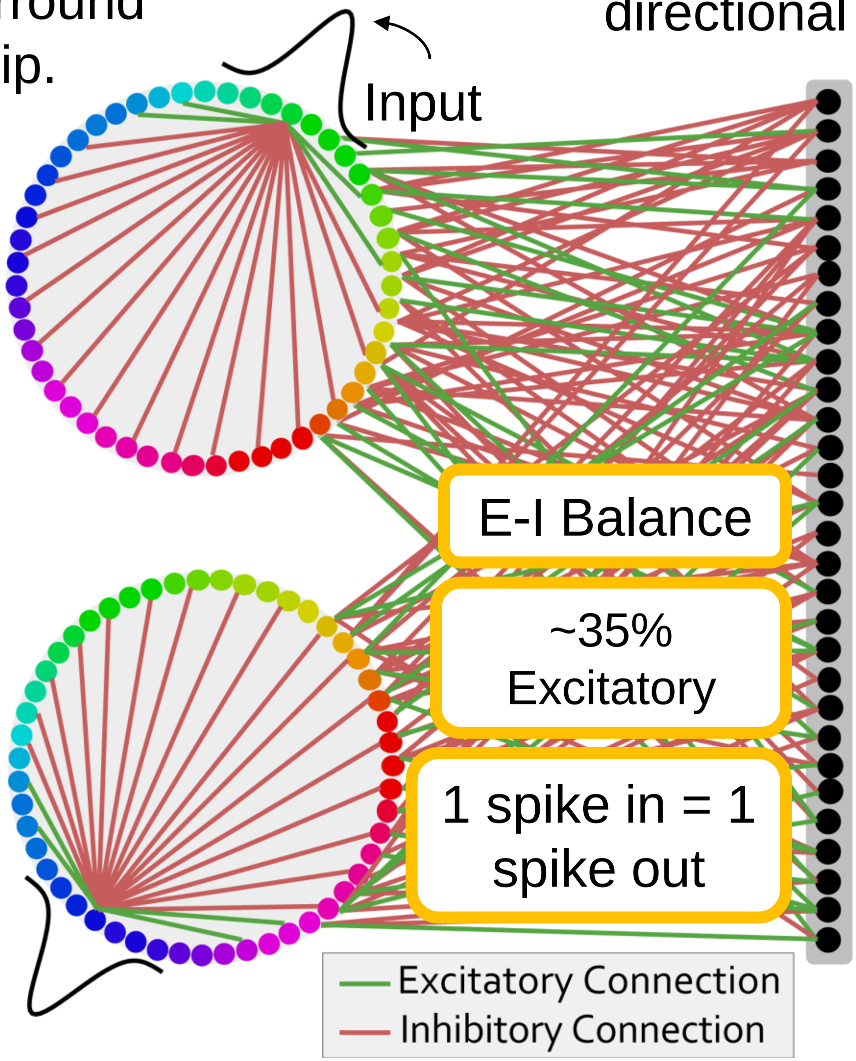
‘Random’ network with *random*, bi-directional connections with sensory cortex.



A simple model of random, reciprocal connections between a 'random network' and 'sensory network' can capture many working memory effects

'Sensory' network with center-surround relationship.

'Random' network with *random*, bi-directional connections with sensory cortex.

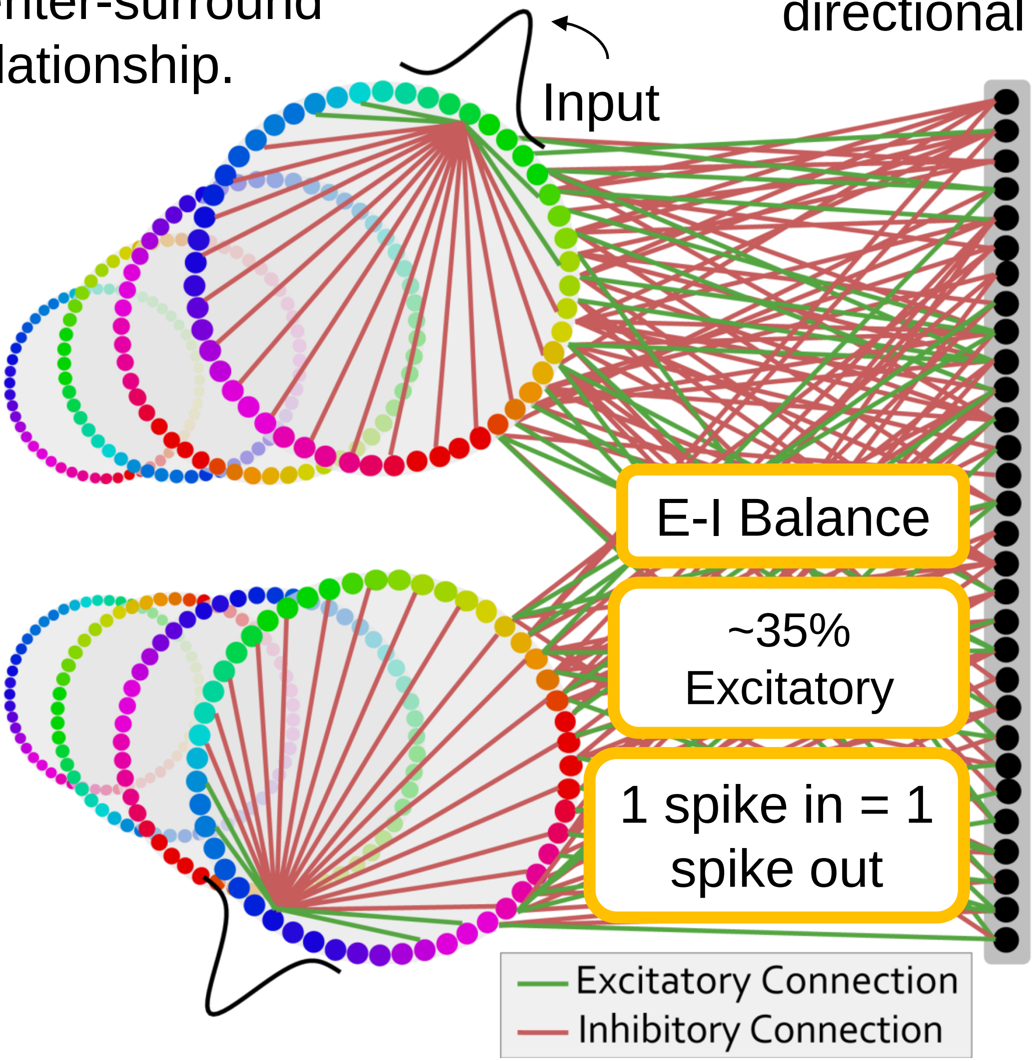


Multiple sensory inputs can be maintained simultaneously; all sensory networks project onto the same control networks.

A simple model of random, reciprocal connections between a 'random network' and 'sensory network' can capture many working memory effects

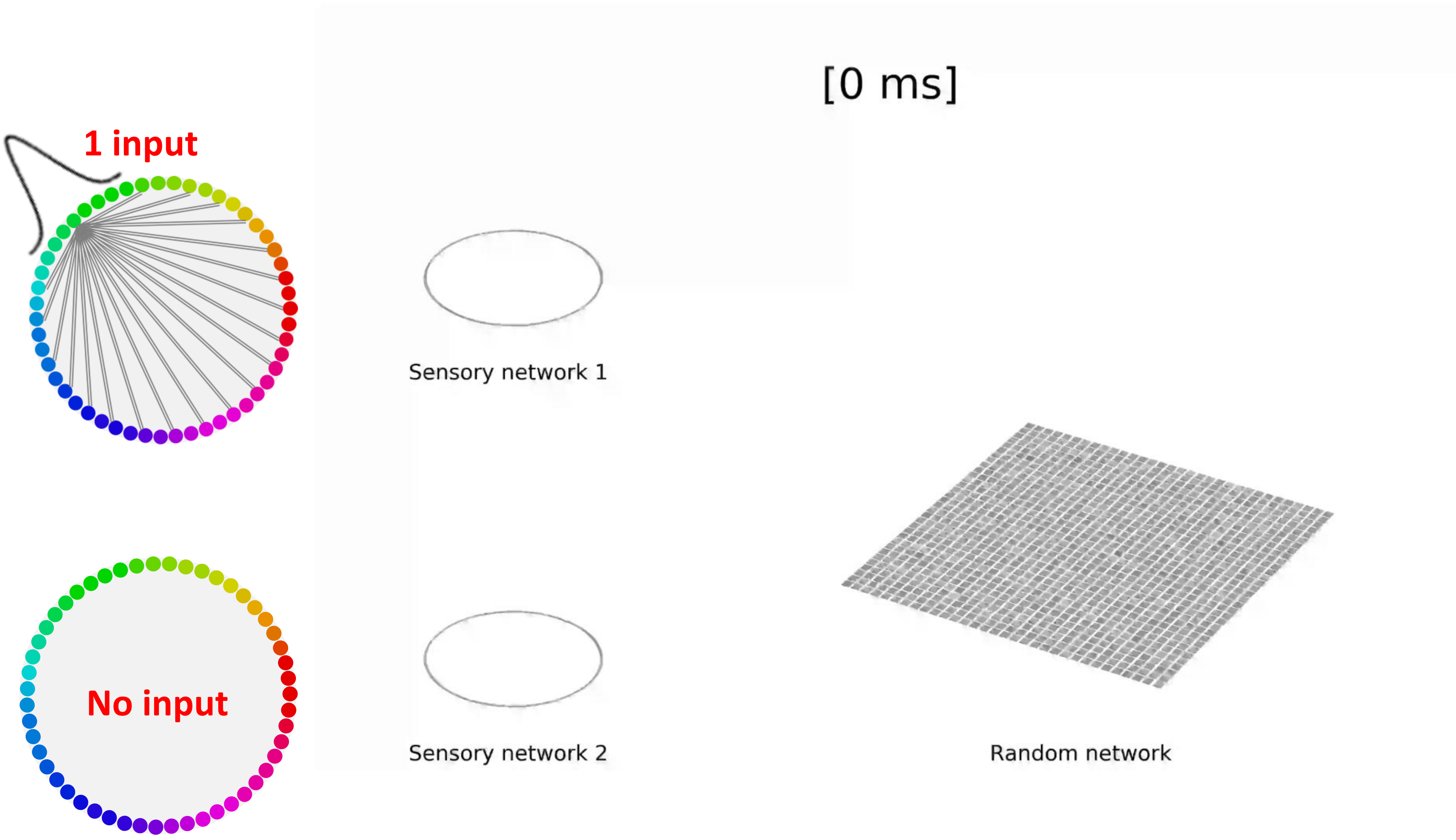
'Sensory' network with center-surround relationship.

'Random' network with *random*, bi-directional connections with sensory cortex.

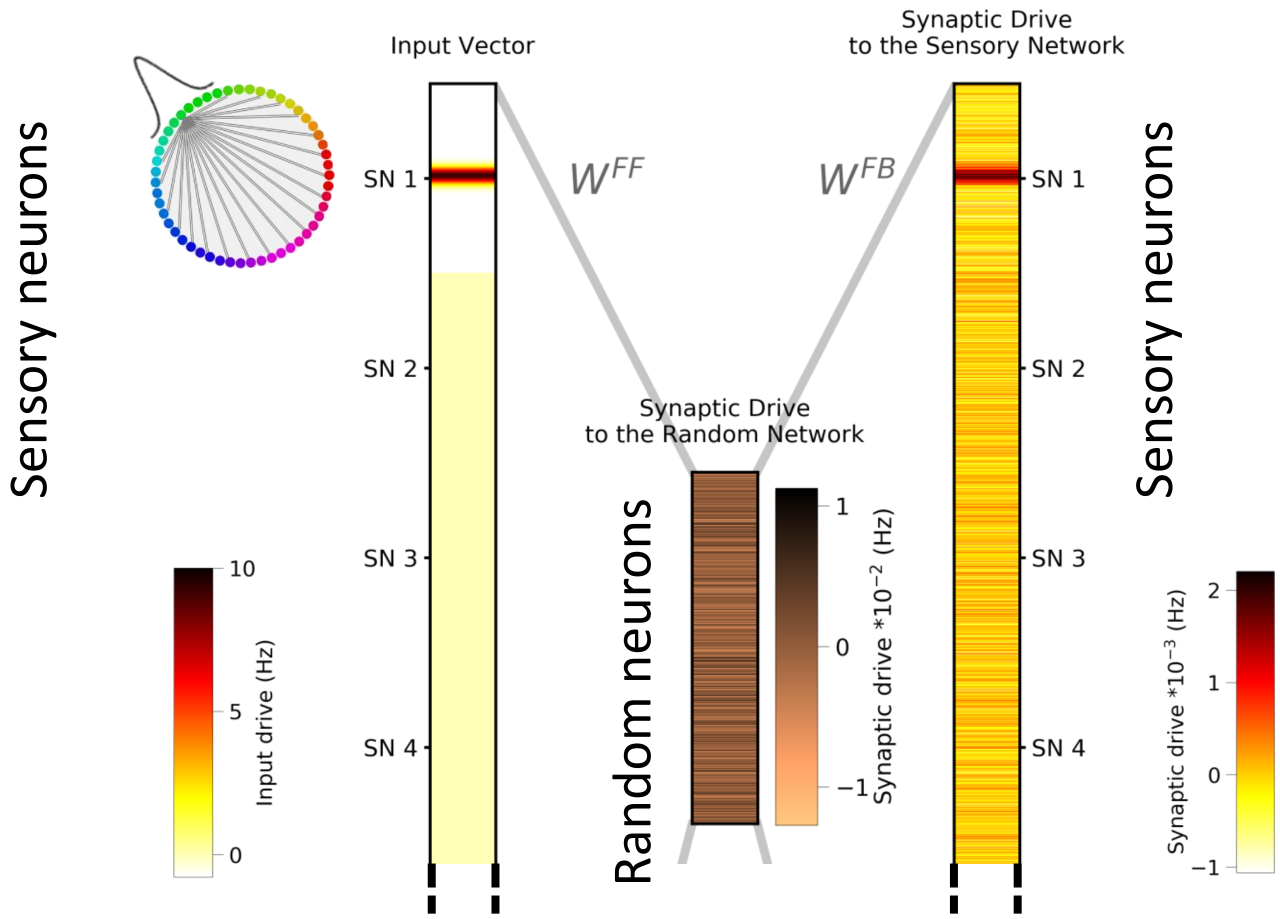


Multiple sensory inputs can be maintained simultaneously; all sensory networks project onto the same control networks.

Network is flexible; can maintain any input into the sensory network



Projections through random network maintain information, allowing memories to be sustained



Behavioral and Physiological Phenomena of Working memory

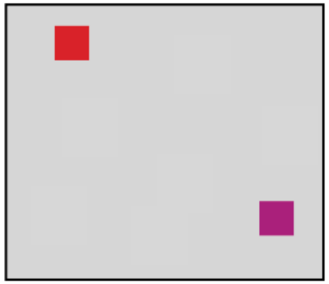
Decades of research has yielded a diversity of behavioral and physiological phenomena of working memory:

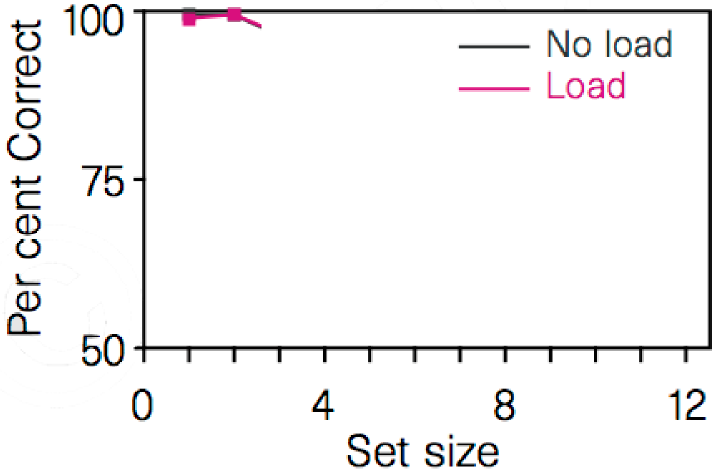
1. Flexibility of working memory (you can hold *anything* in mind).
2. Working memory has a capacity limit.

Behavioral and Physiological Phenomena of Working memory

Decades of research has yielded a diversity of behavioral and physiological phenomena of working memory:

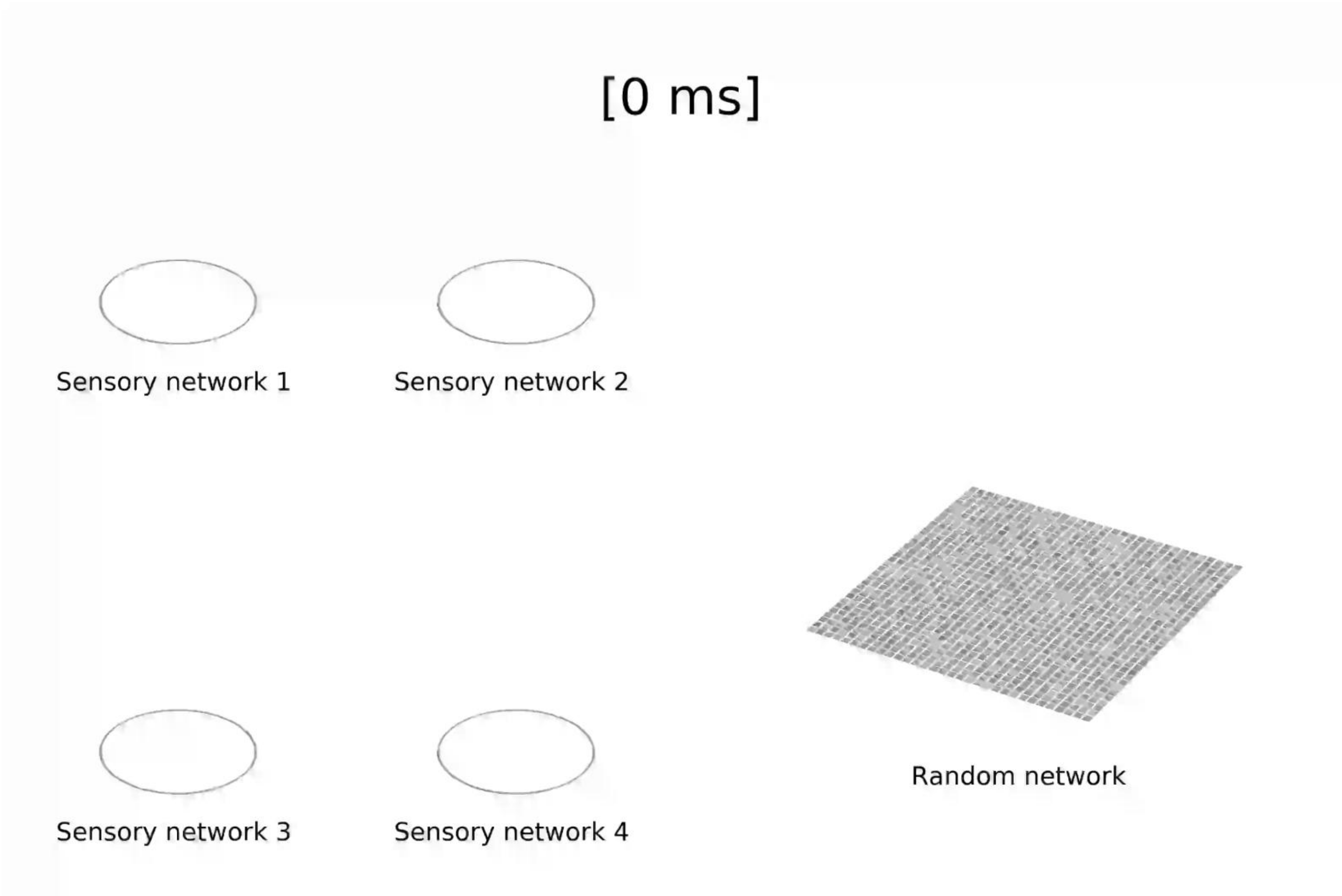
1. Flexibility of working memory (you can hold *anything* in mind).
2. Working memory has a capacity limit.





Luck and Vogel (1997)

Despite flexibility, the network has a strict capacity limit:
Network fails to maintain memory of all 6 items

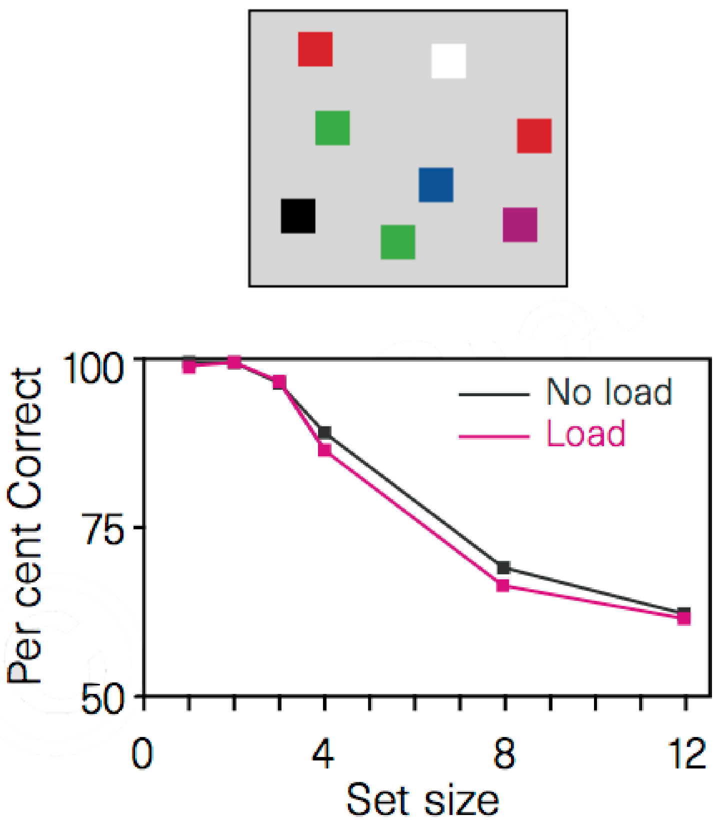


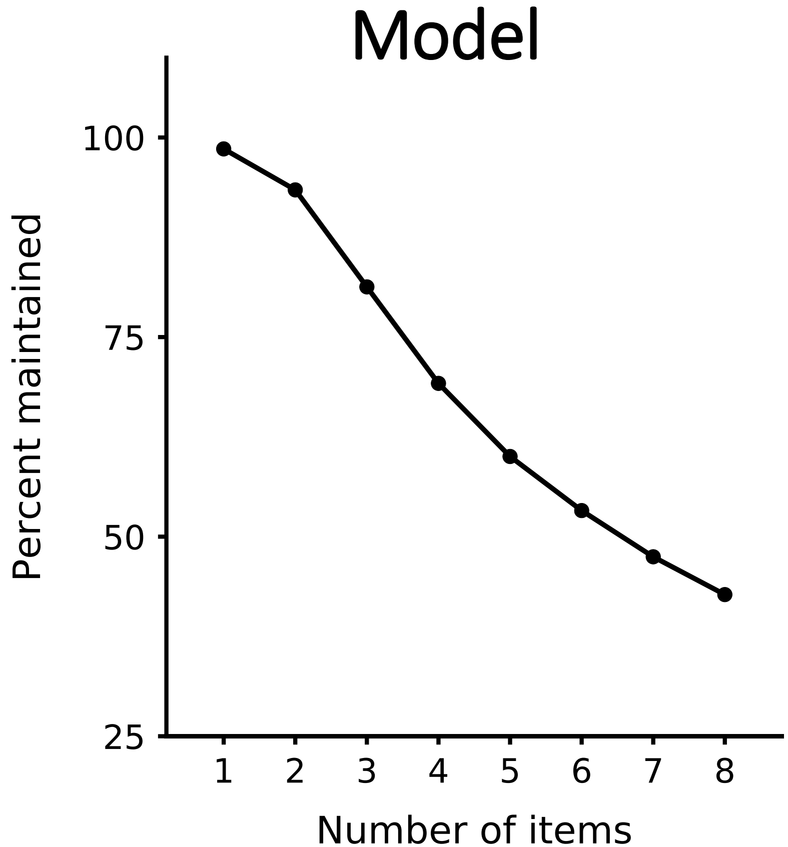
Six inputs into sensory networks are not all maintained.

Behavioral and Physiological Phenomena of Working memory

Decades of research has yielded a diversity of behavioral and physiological phenomena of working memory:

1. Flexibility of working memory (you can hold *anything* in mind).
2. Working memory has a capacity limit.



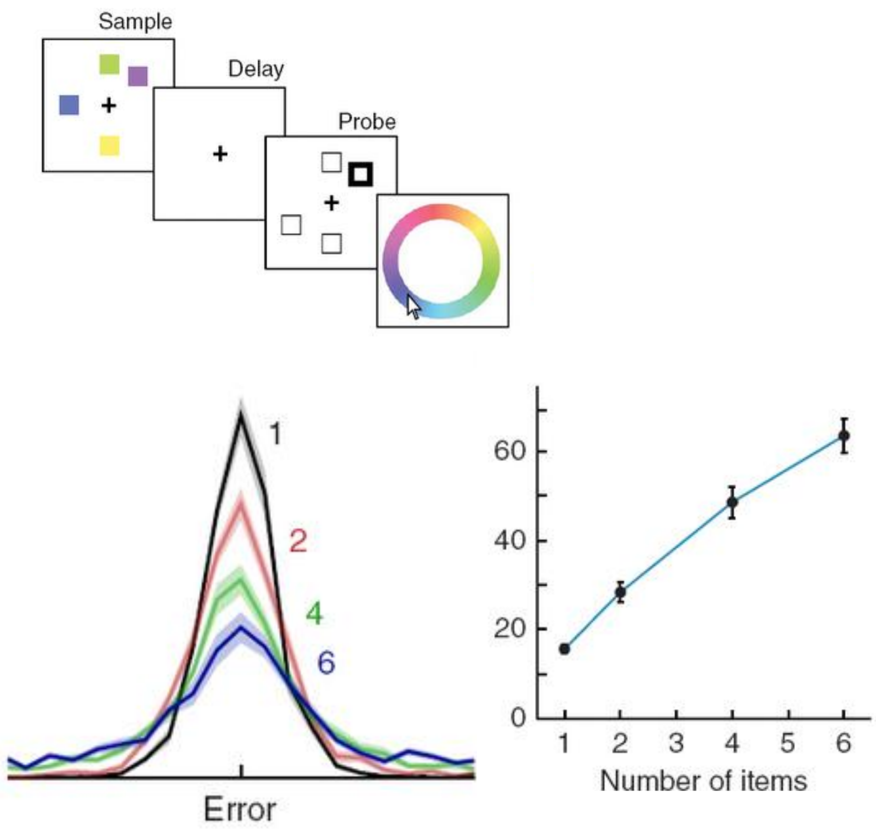


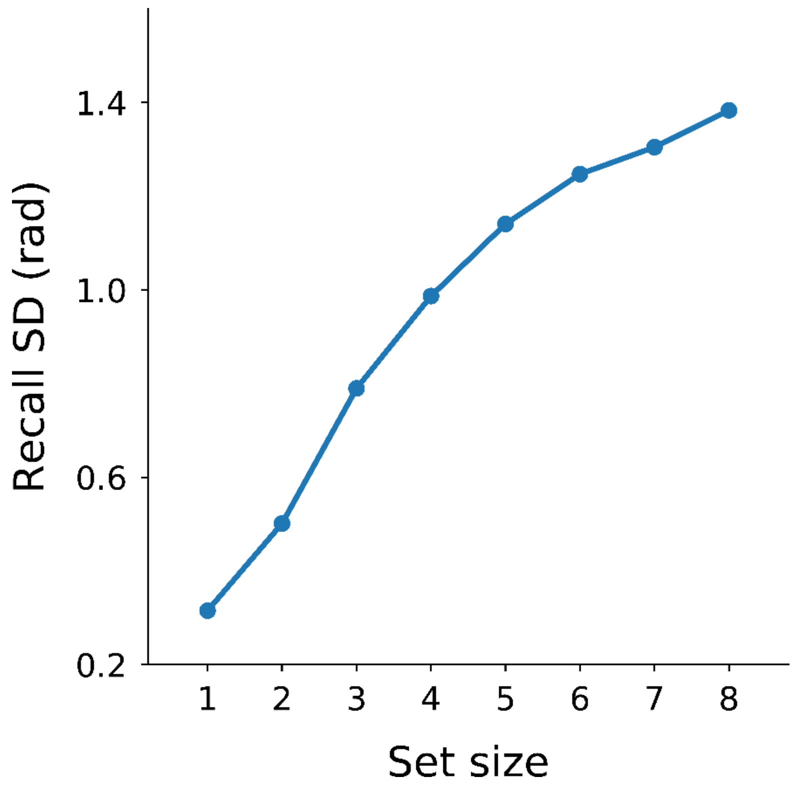
Luck and Vogel (1997)

Behavioral and Physiological Phenomena of Working memory

Decades of research has yielded a diversity of behavioral and physiological phenomena of working memory:

1. Flexibility of working memory (you can hold *anything* in mind).
2. Working memory has a capacity limit.

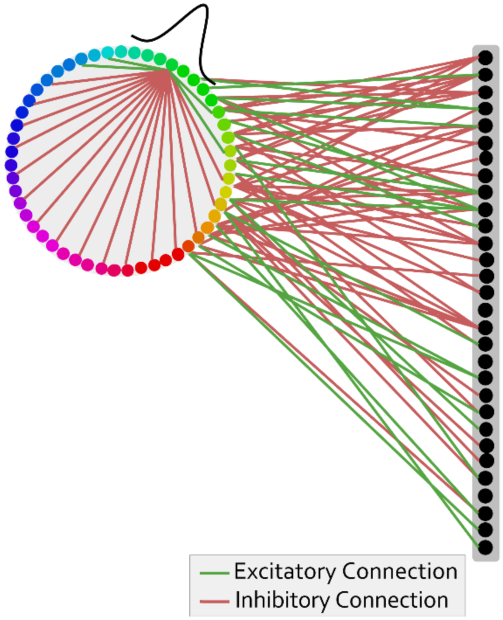




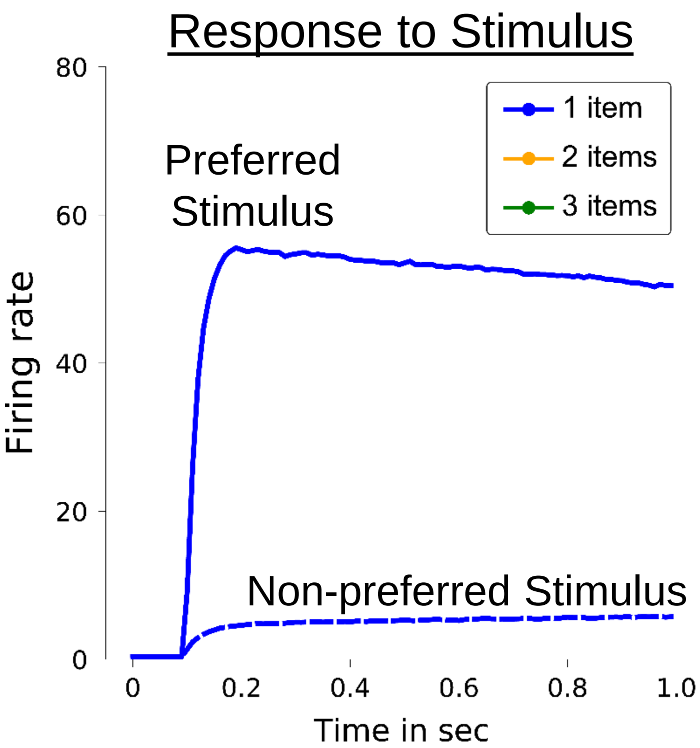
Wilken and Ma (2004)

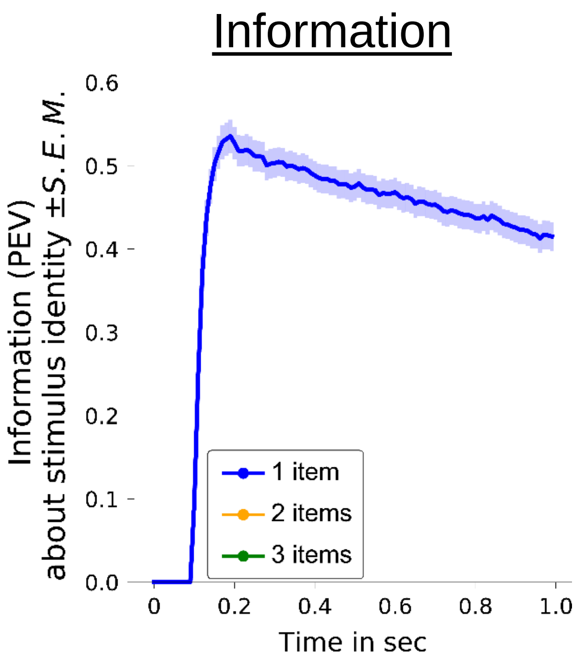
Capacity of working memory arises from interference between memories in the 'random' network

The random connections between sensory networks and random network means representations in the sensory network overlap in the random network.



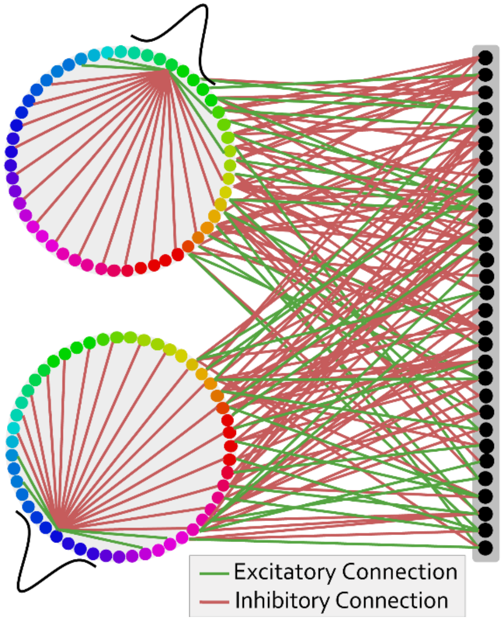
Given the excitation/inhibition balance, this results in a reduction in the response to a given input.



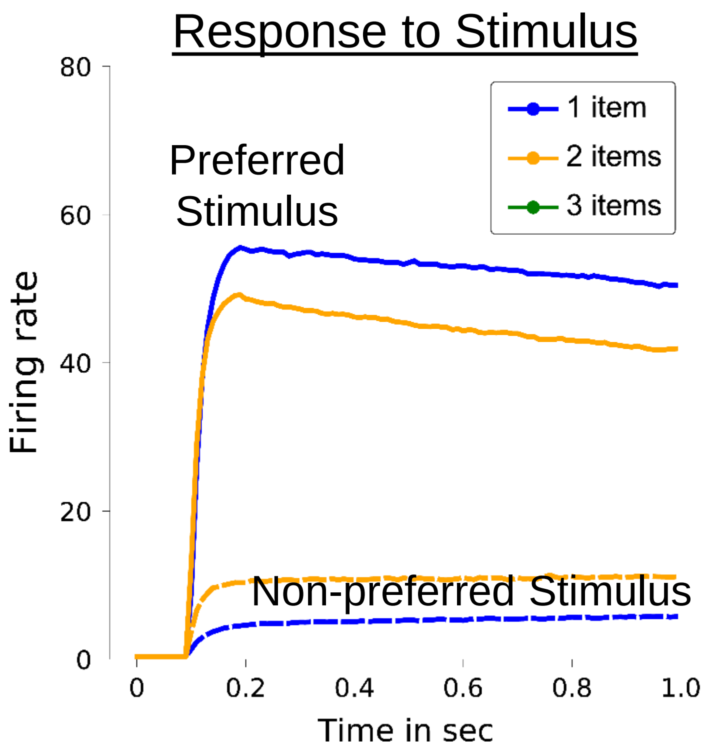


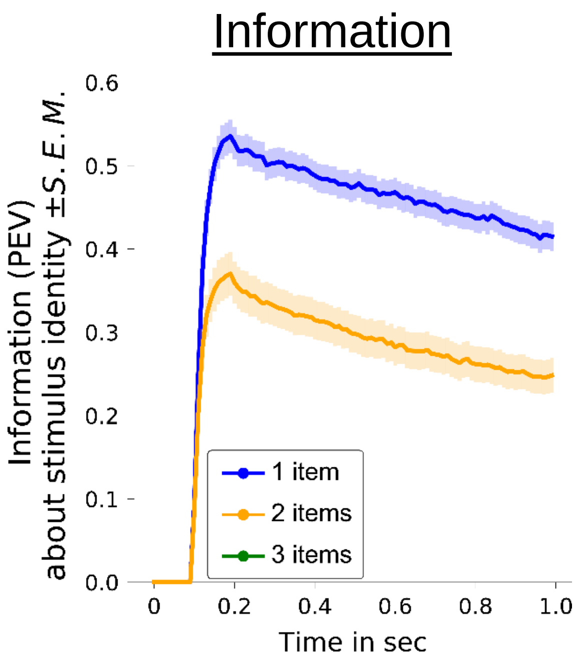
Capacity of working memory arises from interference between memories in the 'random' network

The random connections between sensory networks and random network means representations in the sensory network overlap in the random network.



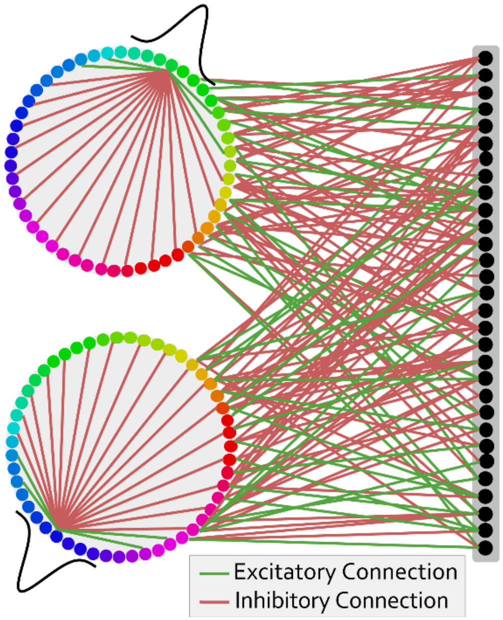
Given the excitation/inhibition balance, this results in a reduction in the response to a given input.



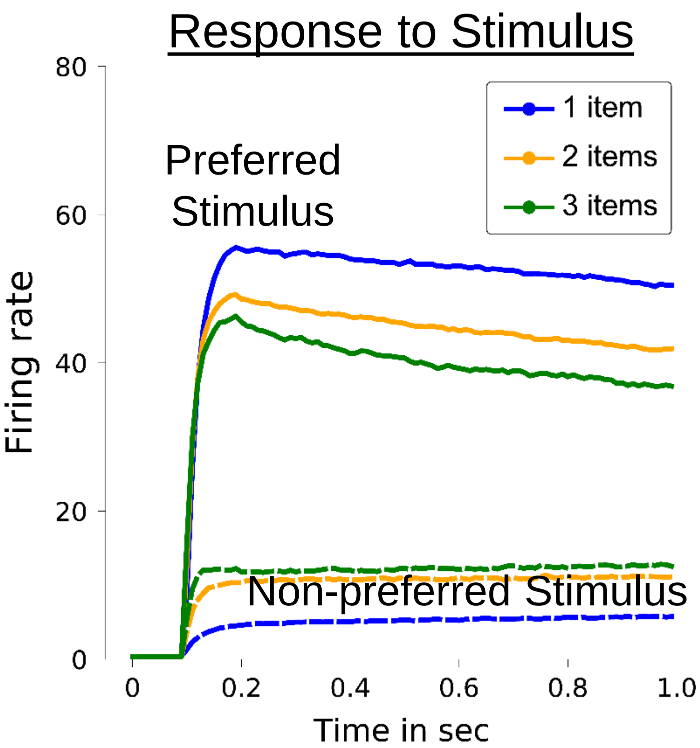


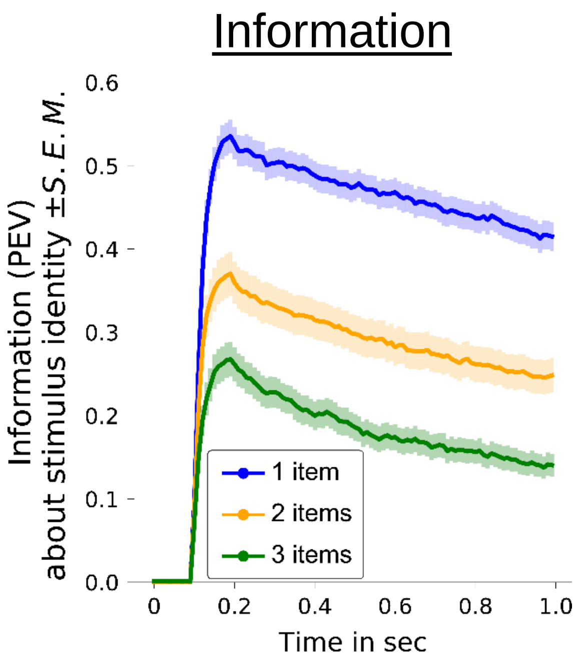
Capacity of working memory arises from interference between memories in the 'random' network

The random connections between sensory networks and random network means representations in the sensory network overlap in the random network.



Given the excitation/inhibition balance, this results in a reduction in the response to a given input.

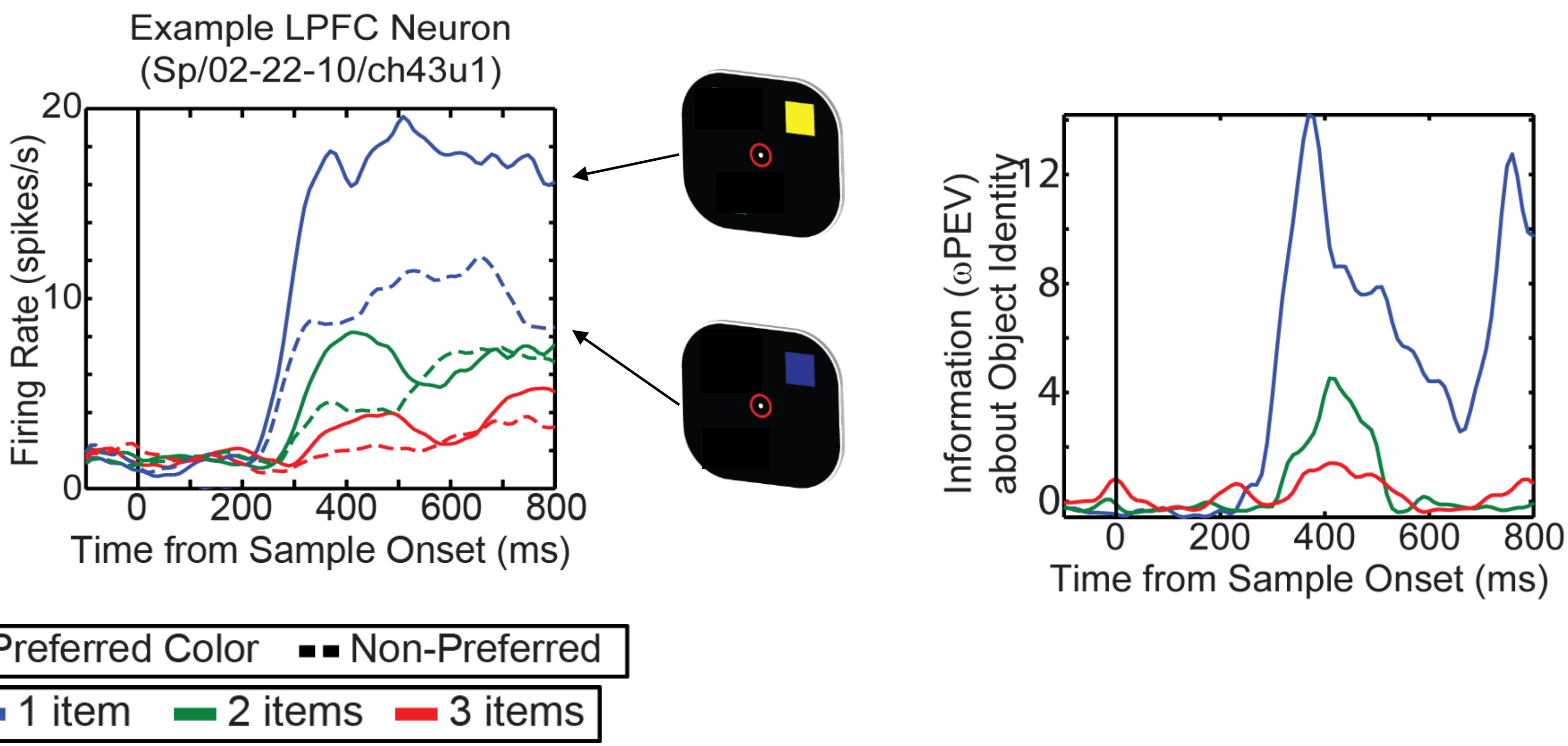




Behavioral and Physiological Phenomena of Working memory

Decades of research has yielded a diversity of behavioral and physiological phenomena of working memory:

1. Flexibility of working memory (you can hold *anything* in mind).
2. Working memory has a capacity limit; likely due to interference.
3. Neurons show divisive-normalization-like regularization of responses.

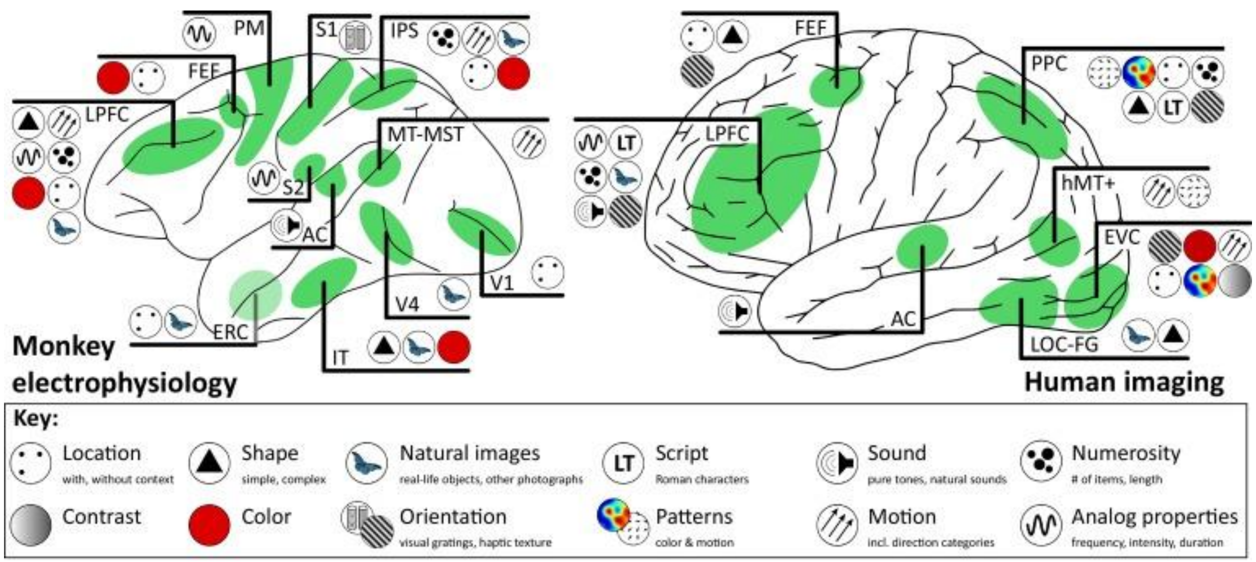


Buschman et al (2011)

Behavioral and Physiological Phenomena of Working memory

Decades of research has yielded a diversity of behavioral and physiological phenomena of working memory:

1. Flexibility of working memory (you can hold *anything* in mind).
2. Working memory has a capacity limit; likely due to interference.
3. Neurons show divisive-normalization-like regularization of responses.
4. Working memory representations are distributed across the brain, including in sensory and control regions.

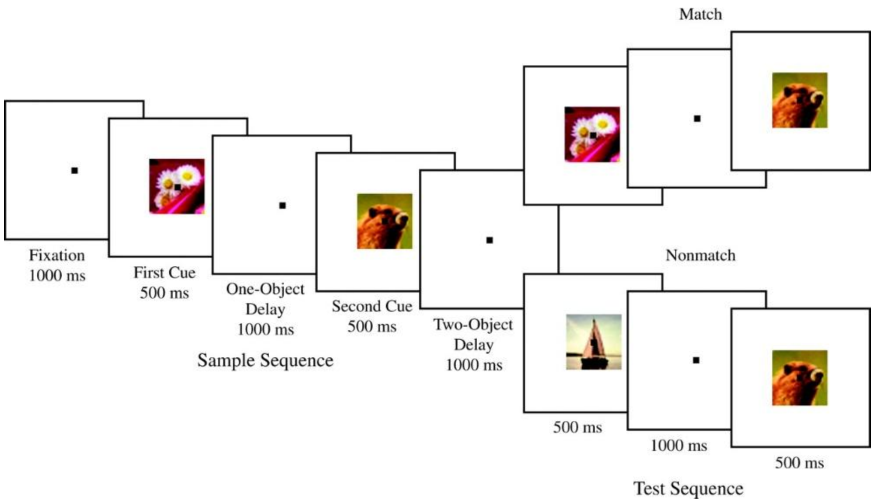


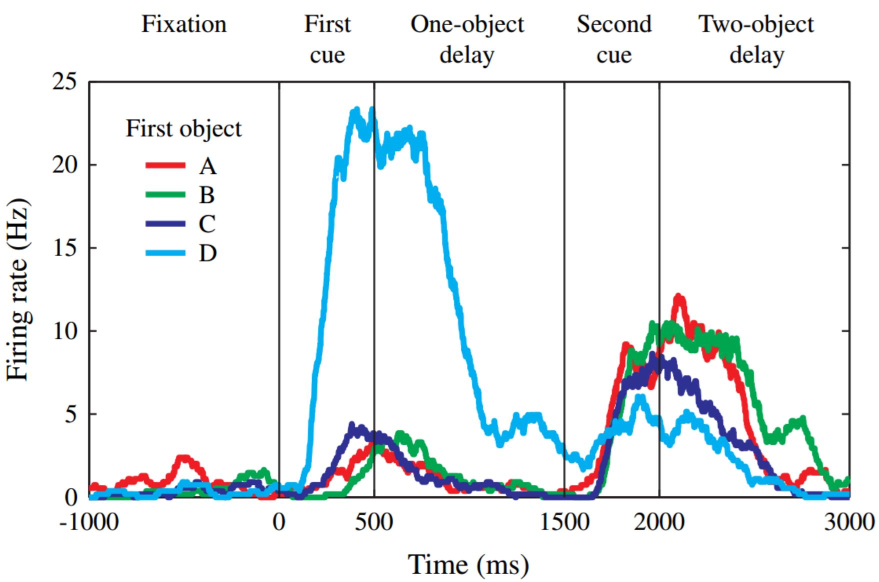
Christophel et al (2017)

Behavioral and Physiological Phenomena of Working memory

Decades of research has yielded a diversity of behavioral and physiological phenomena of working memory:

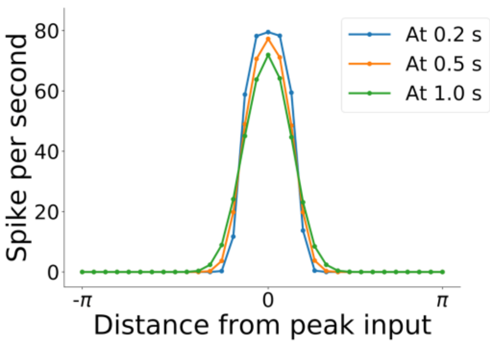
1. Flexibility of working memory (you can hold *anything* in mind).
2. Working memory has a capacity limit; likely due to interference.
3. Neurons show divisive-normalization-like regularization of responses.
4. Working memory representations are distributed across the brain, including in sensory and control regions.
5. Representations in ‘control’ networks (i.e. PFC) are complex (nonlinear).

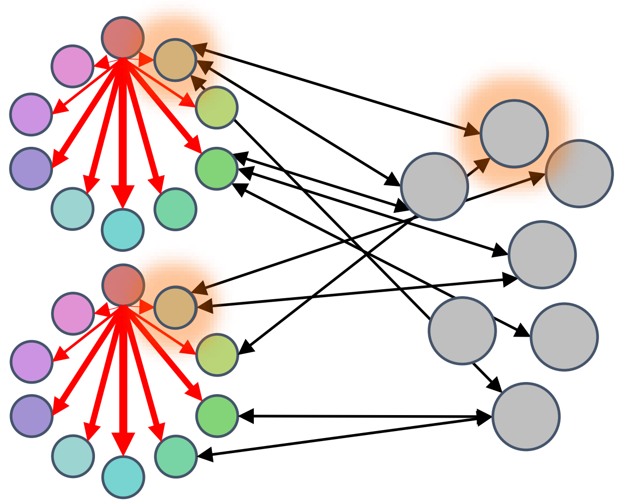


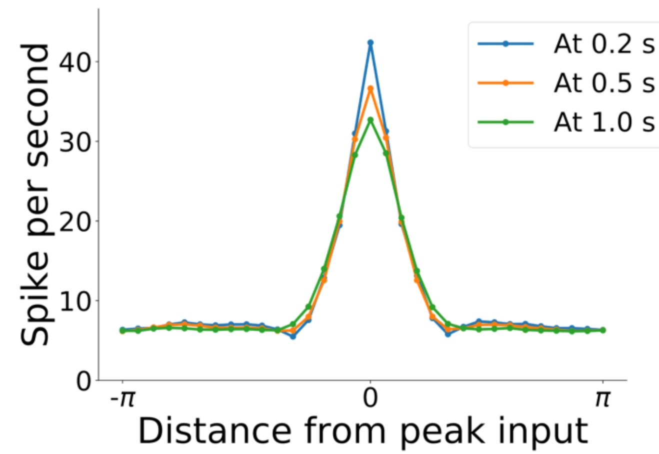


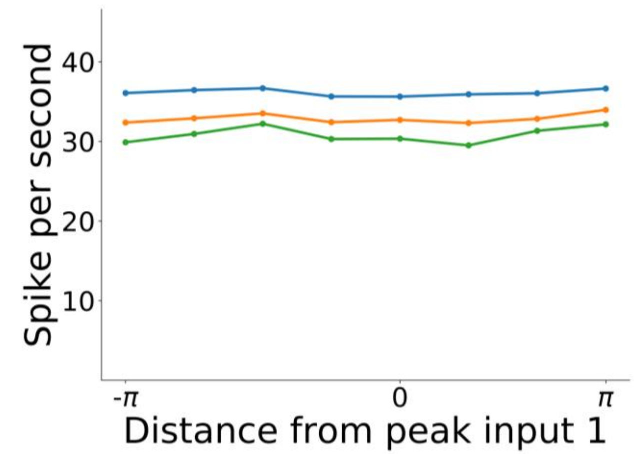
Warden and Miller (2007)

Model neurons in ‘control’ region show PFC like responses; ‘sensory’ network show stimulus tuning









Neurons in the randomly connected ‘control’ network show tuning curves to stimuli. However, they also show significant preference changes across locations. Both of these effects have been observed in monkey PFC.

Summary 2 – Random/Structured Network Captures Many Behavioral & Physiological Phenomena of Working Memory

Decades of research has yielded a diversity of behavioral and physiological phenomena of working memory:

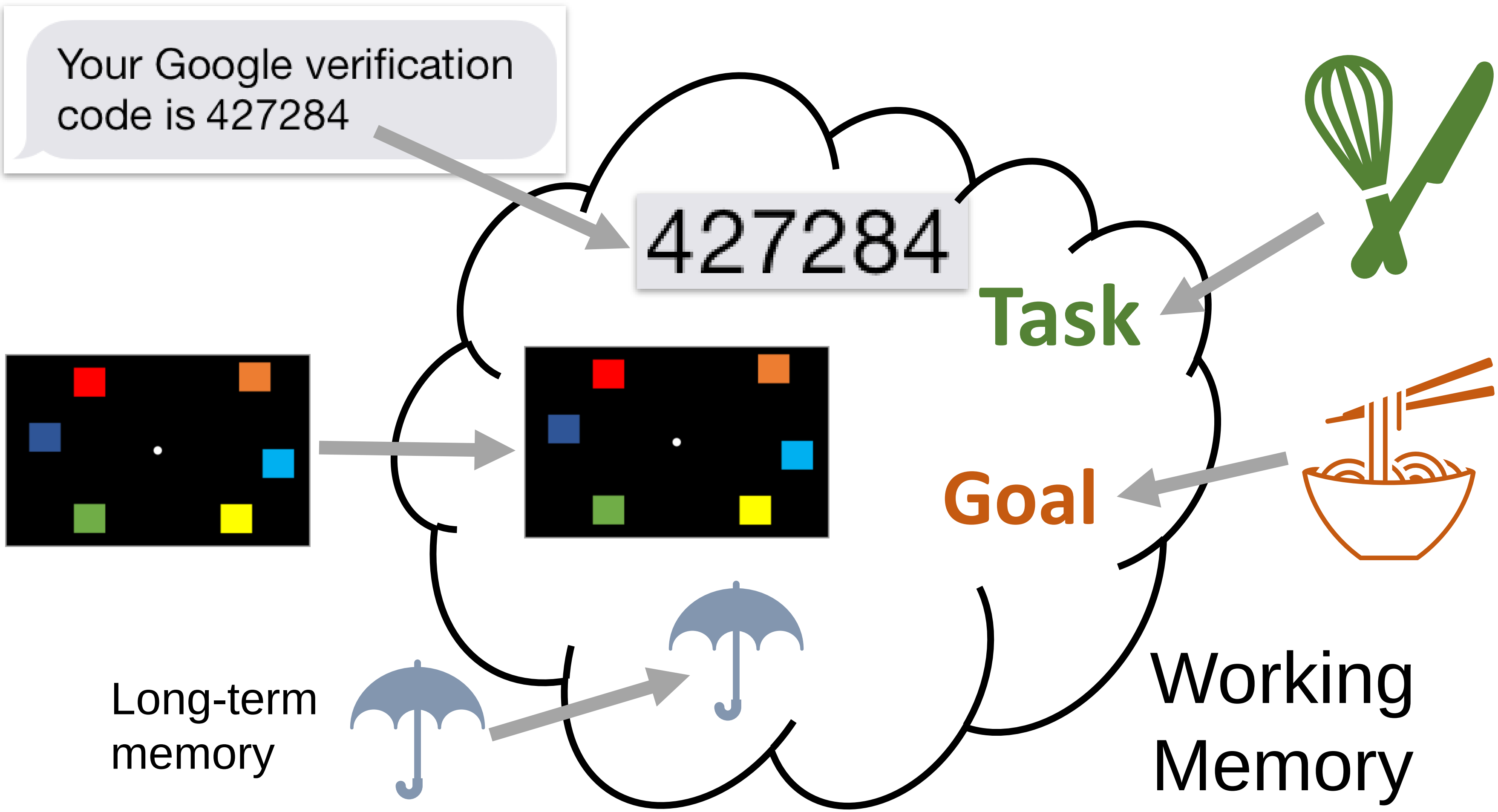
1. Flexibility of working memory (you can hold *anything* in mind).
2. Working memory has a capacity limit; likely due to interference.
3. Neurons show divisive-normalization-like regularization of responses.
4. Working memory representations are distributed across the brain, including in sensory and control regions.
5. Representations in ‘control’ networks (i.e. PFC) are complex (nonlinear).
6. Mnemonic activity is diverse; having both sustained and dynamic properties.
7. Recency-effect; more recent stimuli are better maintained.
8. Retro-cueing improves working memory accuracy.

Predicts:

1. Multiple items should increase noise in sensory cortex representations.
2. Forgotten items should be completely lost in sensory cortex.
3. Reduced interference should increase working memory capacity.

Working memory isn't just for remembering stimuli…

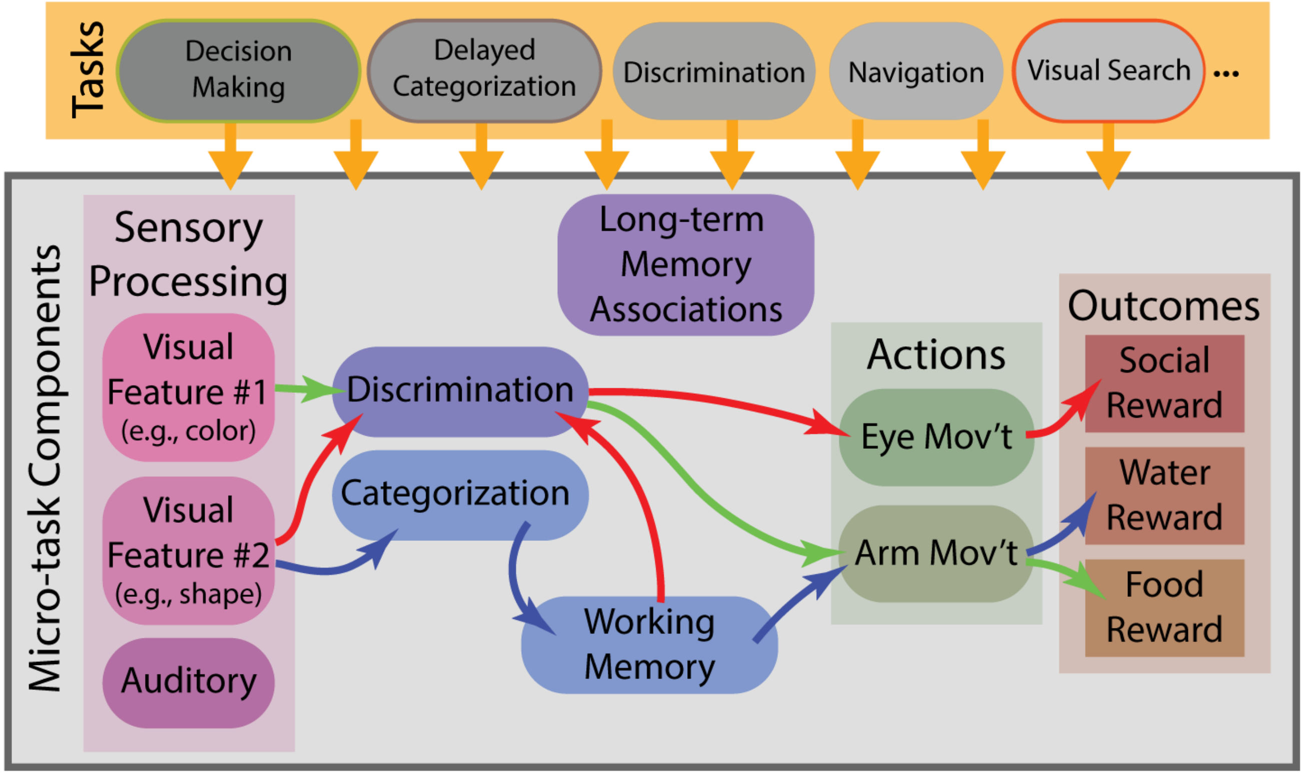
Working memory is your ability to hold things 'in mind'. It provides the workspace for higher cognitive functions, such as decision making, goal-directed behavior, and attention.

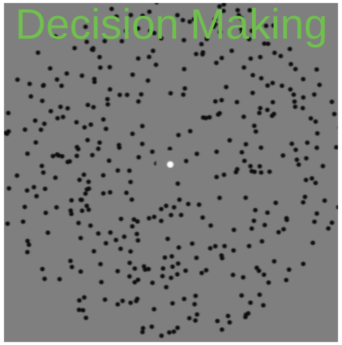


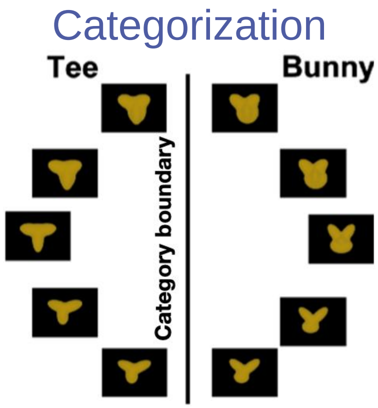
Working Memory for Cognitive Control

Cognition is remarkably flexible – humans and animals are excellent multi-task agents, able to perform a multitude of behaviors.

Cognitive control is the ability to select a goal-relevant, situationally-appropriate, behavior.







Learning attentional templates – a form of cognitive control

Caroline Jahn

How does the brain learn to control cognition?

Feature-based attention allows us to focus on the stimuli with task-relevant features (e.g., color, shape, motion).

An attentional template encapsulates the set of stimulus features that are relevant for the current situation.

Jahn et al, *Cell*, 2024

Learning attentional templates – a form of cognitive control

Caroline Jahn

How does the brain learn to control cognition?

Feature-based attention allows us to focus on the stimuli with task-relevant features (e.g., color, shape, motion).

An attentional template encapsulates the set of stimulus features that are relevant for the current situation.

Jahn et al, *Cell*, 2024

Learning attentional templates – a form of cognitive control

Caroline Jahn

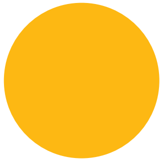
How does the brain learn to control cognition?

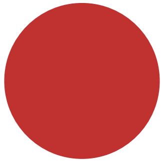
When the situation (or your goals) changes, you must adapt by learning a new contextually-appropriate template.

New York City

Costa Rica

Berlin



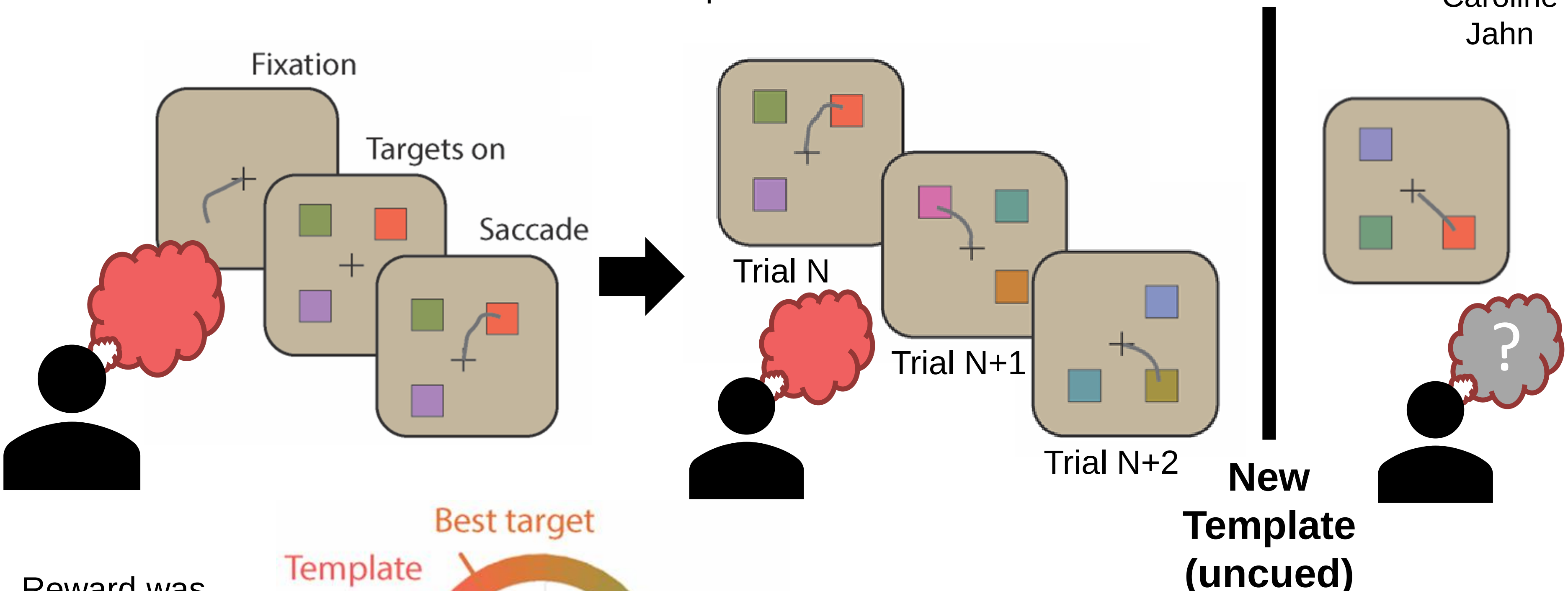


Jahn et al, *Cell*, 2024

Monkeys performed a visual search task while learning new attentional templates

Caroline Jahn

Monkeys performed a typical visual search task: find the stimulus that was closest to the attentional template.



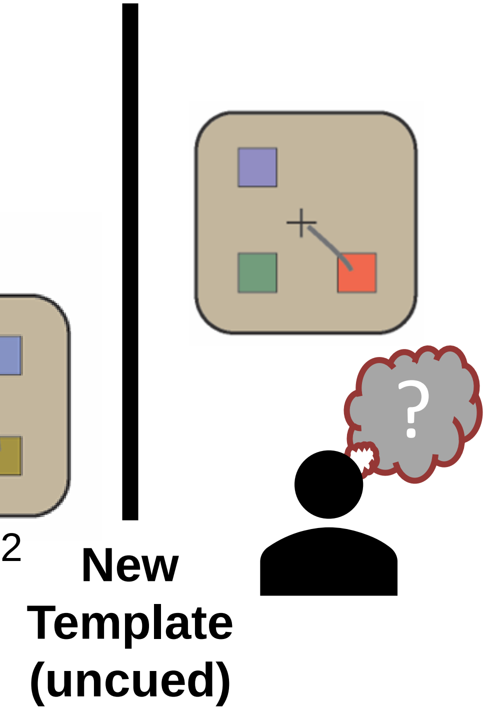
Reward was proportional to the distance in color space between the chosen target and the template.

Jahn et al, *Cell*, 2024

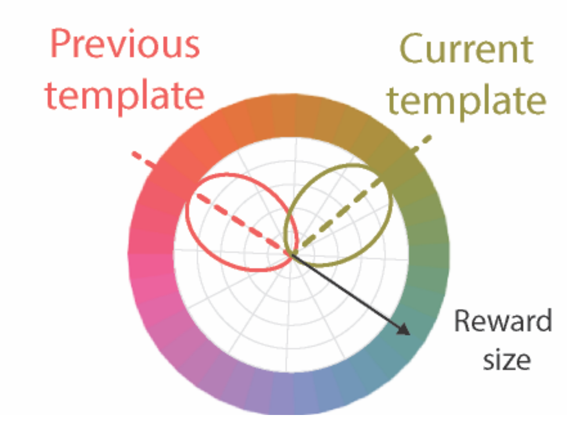
Monkeys performed a visual search task while learning new attentional templates

Caroline Jahn

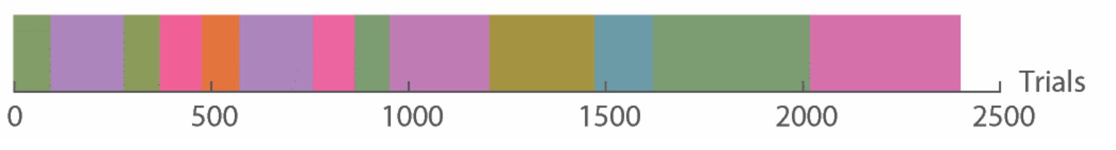
Monkeys performed a typical visual search task: find the stimulus that was closest to the attentional template.



The reward function changes with the attentional template.



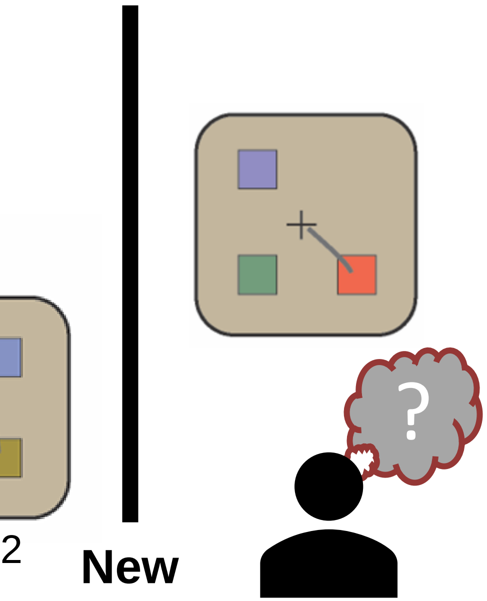
To study learning, the template changed multiple times during each day:



Jahn et al, *Cell*, 2024

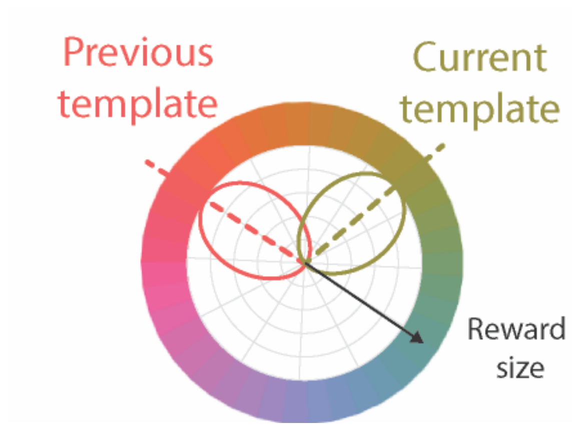
Monkeys performed a visual search task while learning new attentional templates

Monkeys performed a typical visual search task: find the stimulus that was closest to the attentional template.

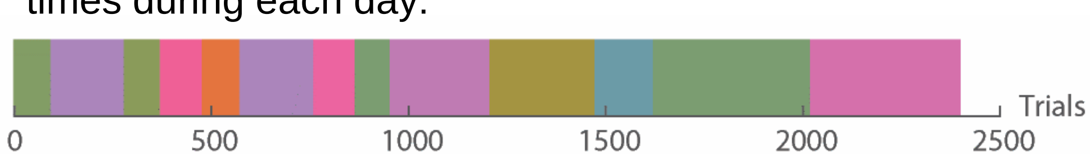


New Template (uncued)

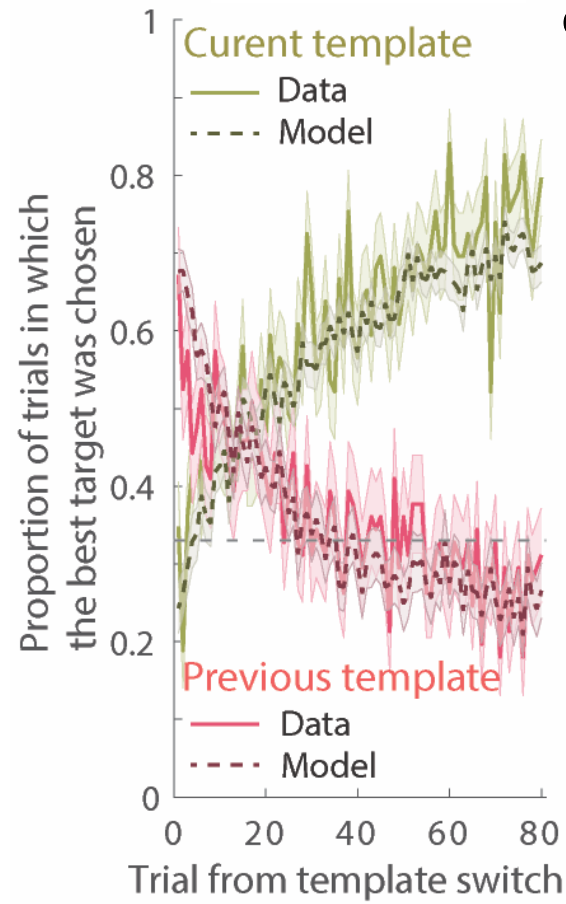
The reward function changes with the attentional template.



To study learning, the template changed multiple times during each day:



Learning Curve



Behavior was modeled by a Q-learning with function approximation model.

Caroline Jahn

Jahn et al, *Cell*, 2024

Neural recordings were performed in frontal and parietal cortex

Caroline
Jahn

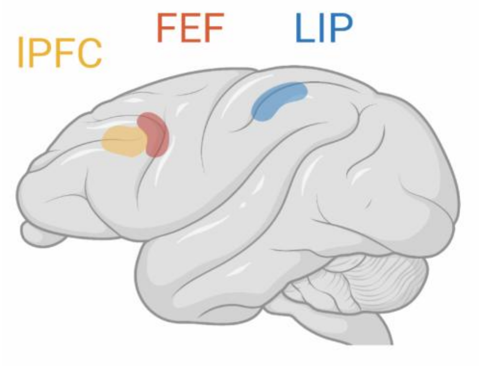
We were interested in answering three questions:

1. How are attentional templates represented?

2. How are new attentional templates learned?

3. How does the animal make decisions across multiple templates?

Neural recordings were performed across prefrontal and parietal cortex – regions known to be involved in directing attention.

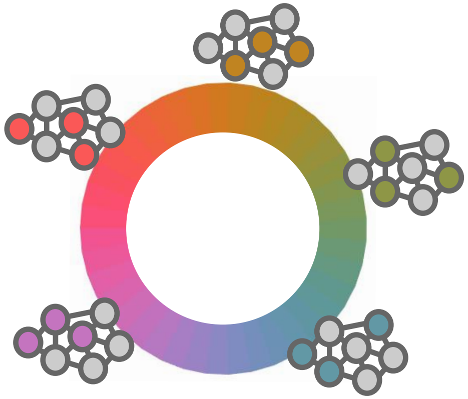


Jahn et al, *Cell*, 2024

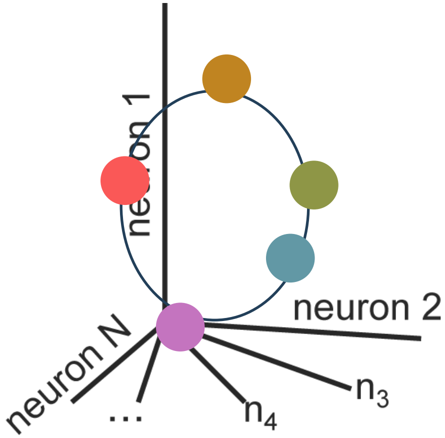
Attentional templates are represented in a structured manner

Caroline Jahn

How are templates represented?

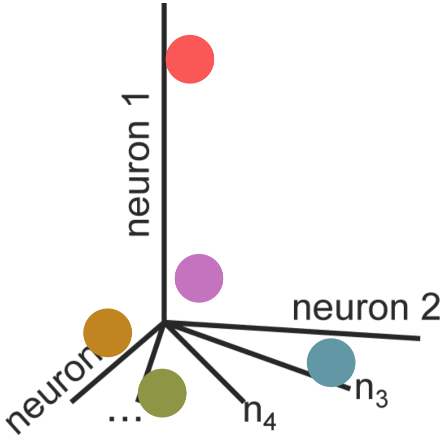


Hypothesis #1: Control representations are structured, with semantically similar templates represented in similar ways in the neural population.



Advantages:
+ Interpolation/Generalization
+ Match sensory representations

Hypothesis #2: Control representations are high-dimensional and unique to each task.



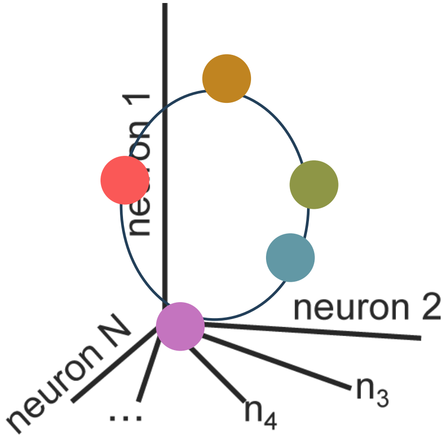
Advantages:
+ Linearly separable
+ Reduced interference

Jahn et al, *Cell*, 2024

Attentional templates are represented in a structured manner

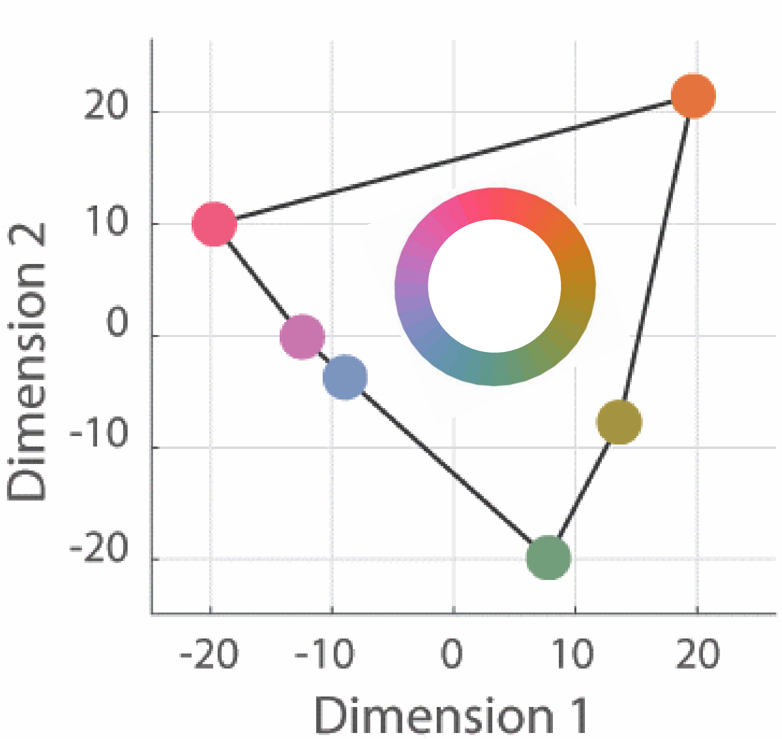
How are templates represented?

Hypothesis #1: Control representations are structured, with semantically similar templates represented in similar ways in the neural population.



Caroline Jahn

Multi-Dimensional Scaling
Of Neural Population Response



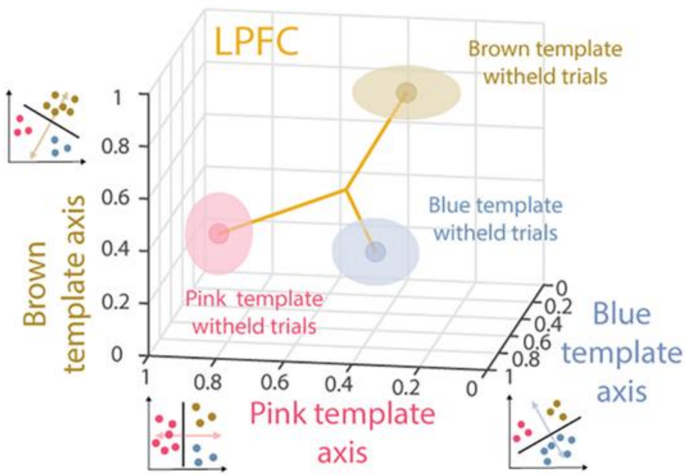
Jahn et al, *Cell*, 2024

Attentional templates are maintained throughout the task

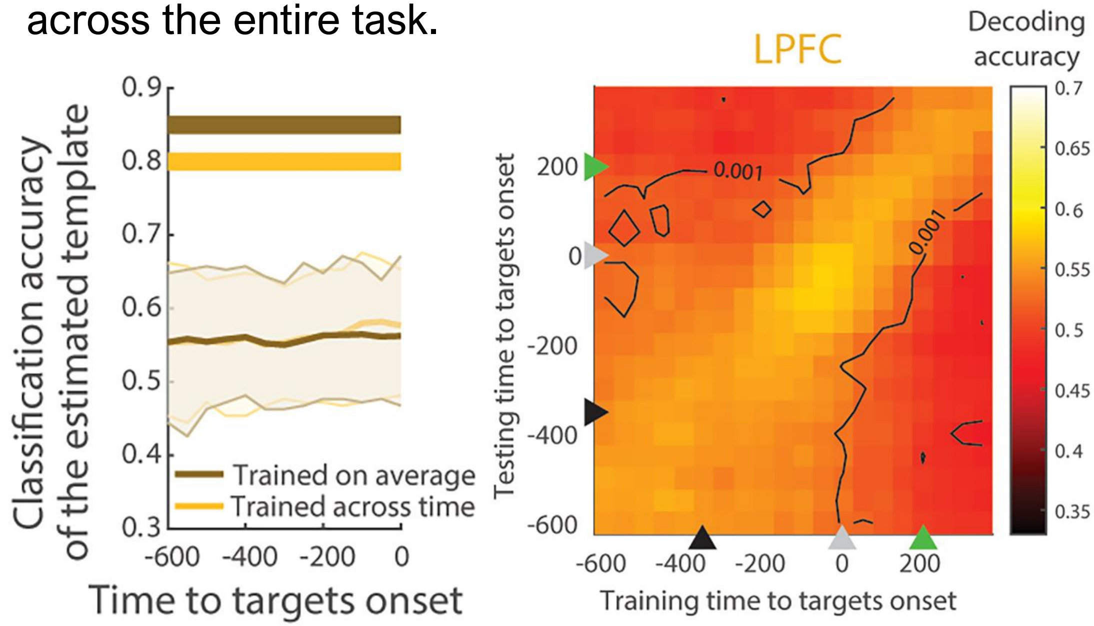

Caroline Jahn

Template representations can be decoded from the neural population in LPFC:

Estimated template population representation



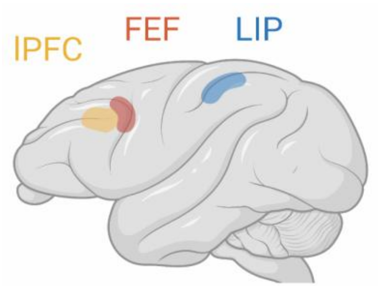
These template representations are <u>stably maintained</u> across the entire task.



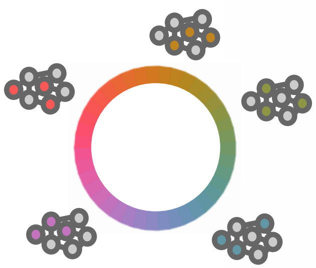
Jahn et al, *Cell*, 2024

Summary of Learning to Control Attention

Caroline Jahn

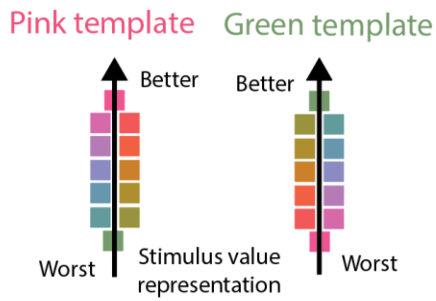


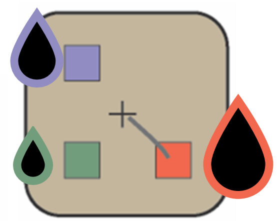
Fronto-parietal network



Template

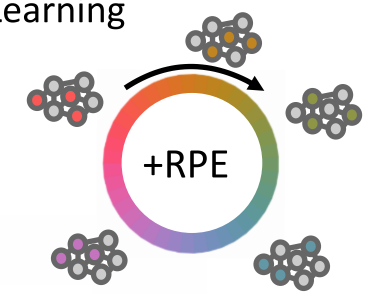
Fig 2

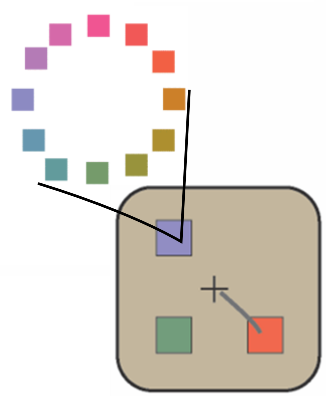




Sensory Input

Learning



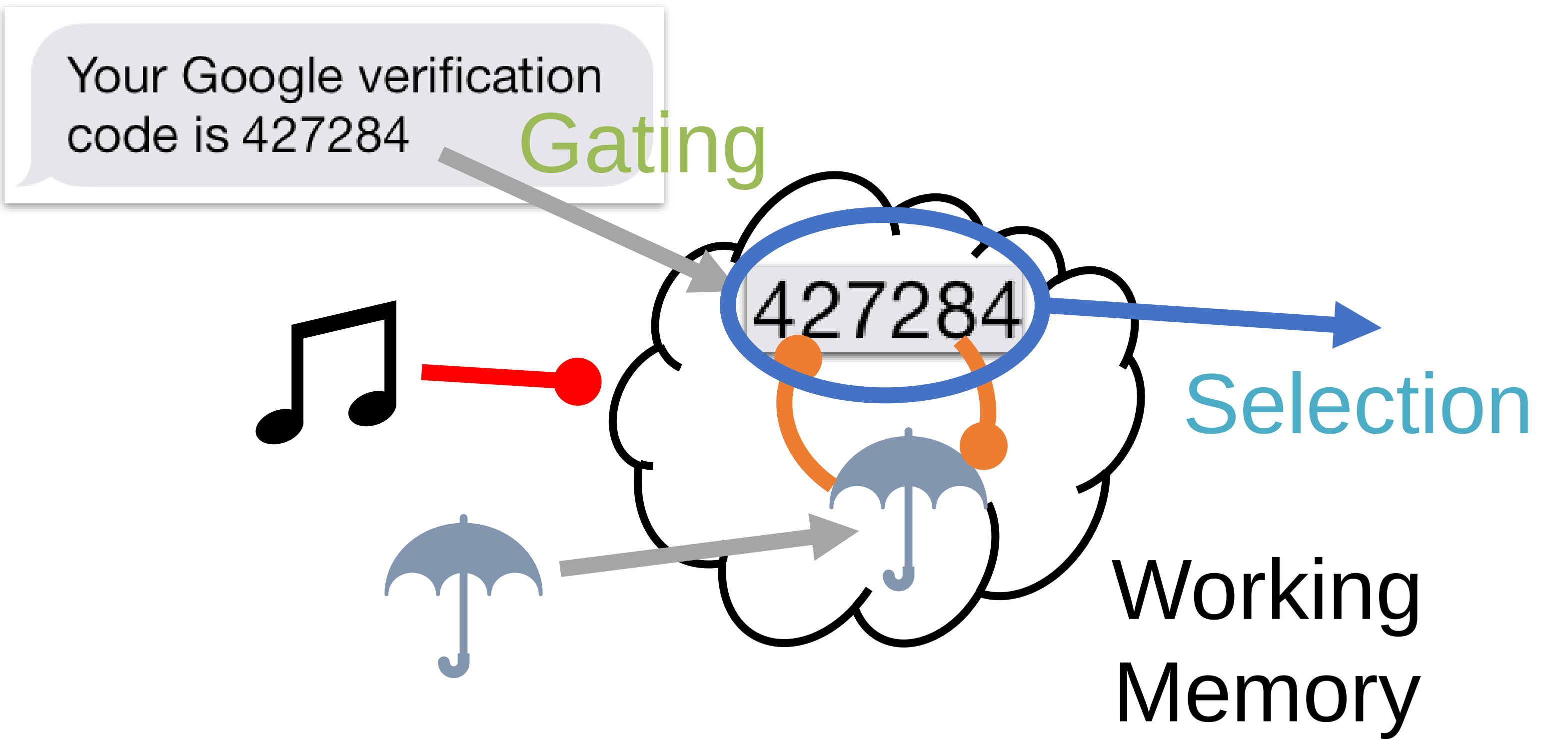


Sensory Input

Cognitive Control of Working Memory

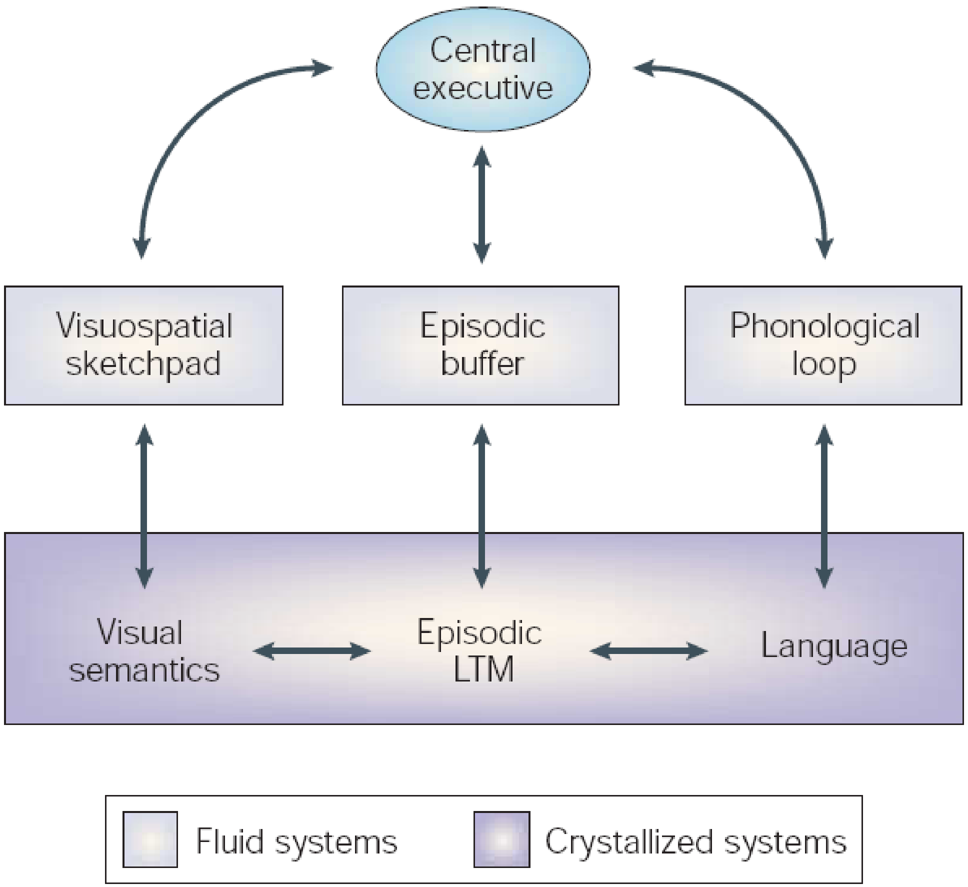
Control of Working Memory Compensates for Limited Capacity

Working memory is your ability to hold things 'in mind'. It provides the workspace for higher cognitive functions, such as decision making, goal-directed behavior, and attention.

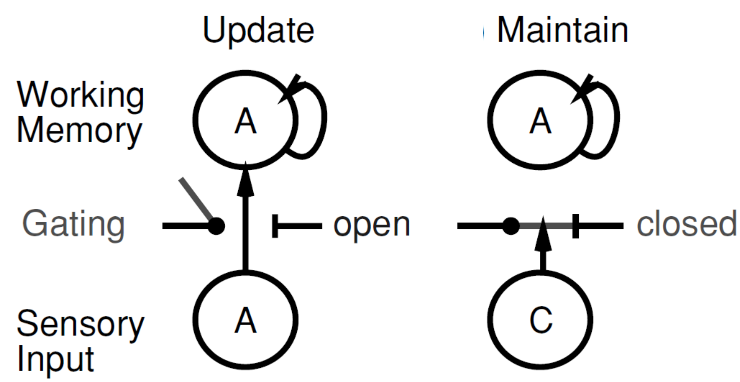


To compensate for the limited capacity of working memory, the brain has developed mechanisms to tightly control the contents of working memory.

Baddeley's Multicomponent Model of WM



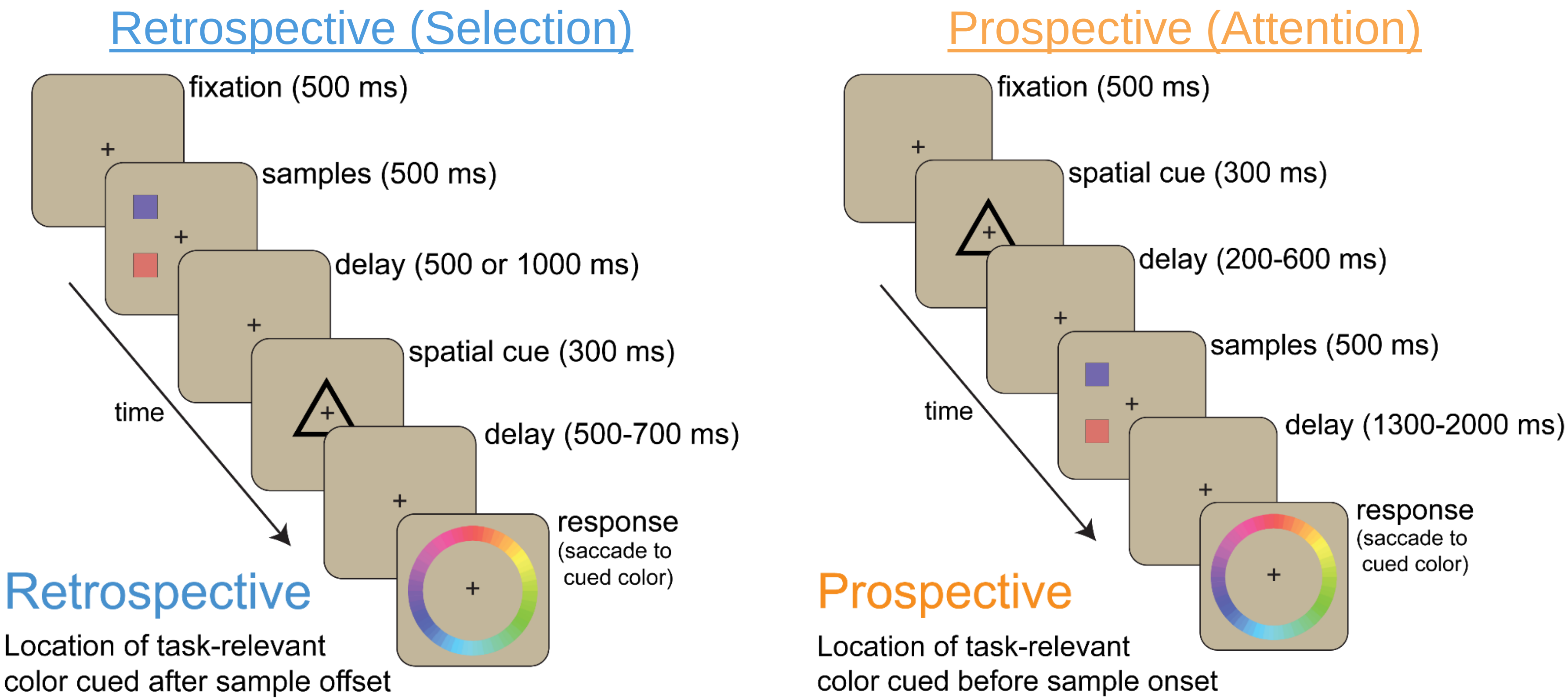
<u>Gating the Contents of Working Memory</u>



- Central executive regulates limited attentional resources and limited storage capacity
- Items move "in" or "out" of visual or verbal buffers (analogous to RAM)
- Active maintenance/ rehearsal of items in buffers

Baddeley (2003)

Retrospective selection from working memory; prospective attention to visual stimuli

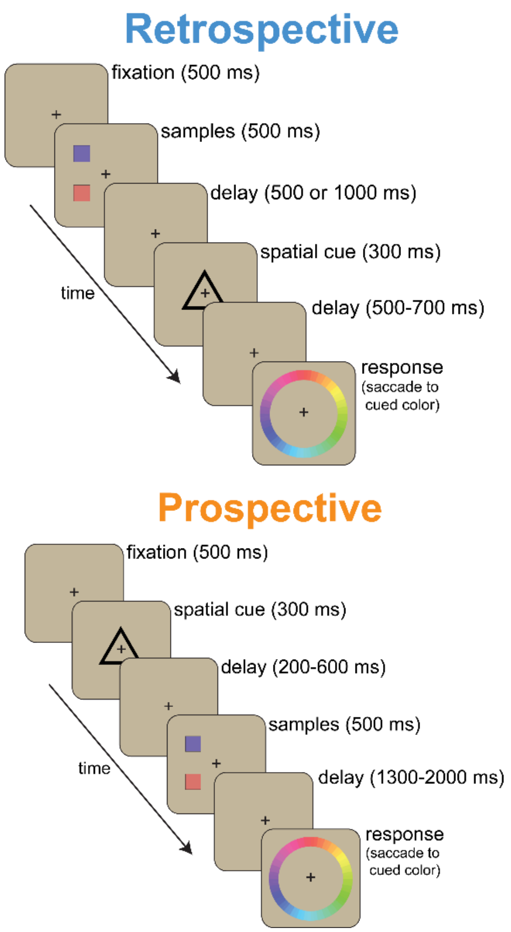


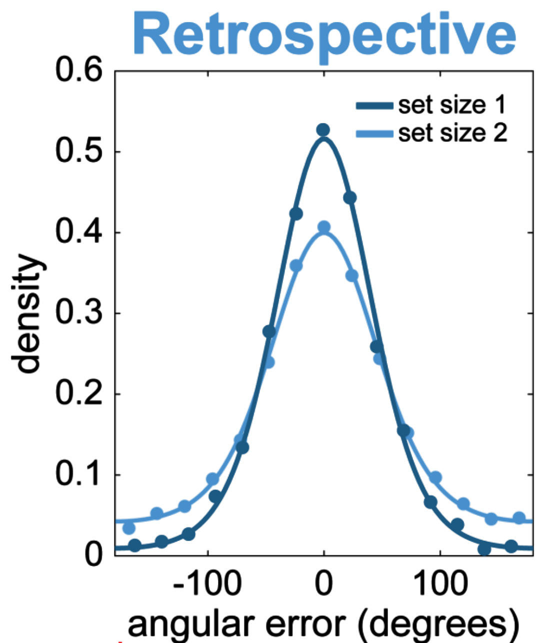
Monkeys performed the continuous working memory task in both a selection condition (with a retro-cue) and an attention condition (cue before stimulus).

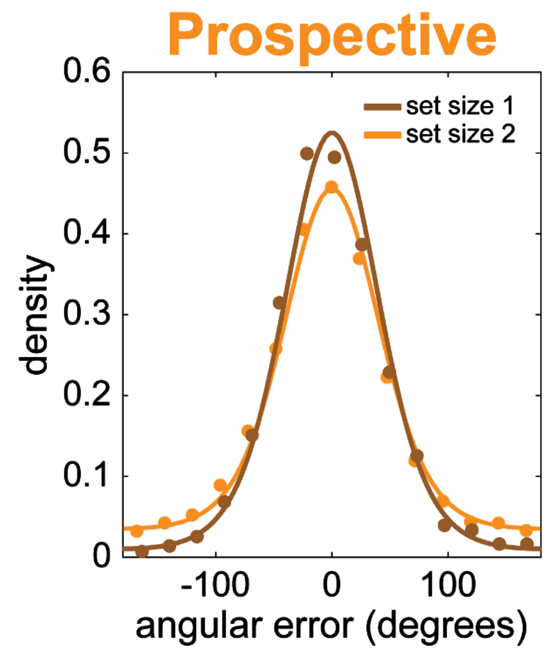
Panichello and Buschman, *Nature* 2021

Selection a single item from working memory improves memory performance

Holding two items in memory impairs working memory accuracy.



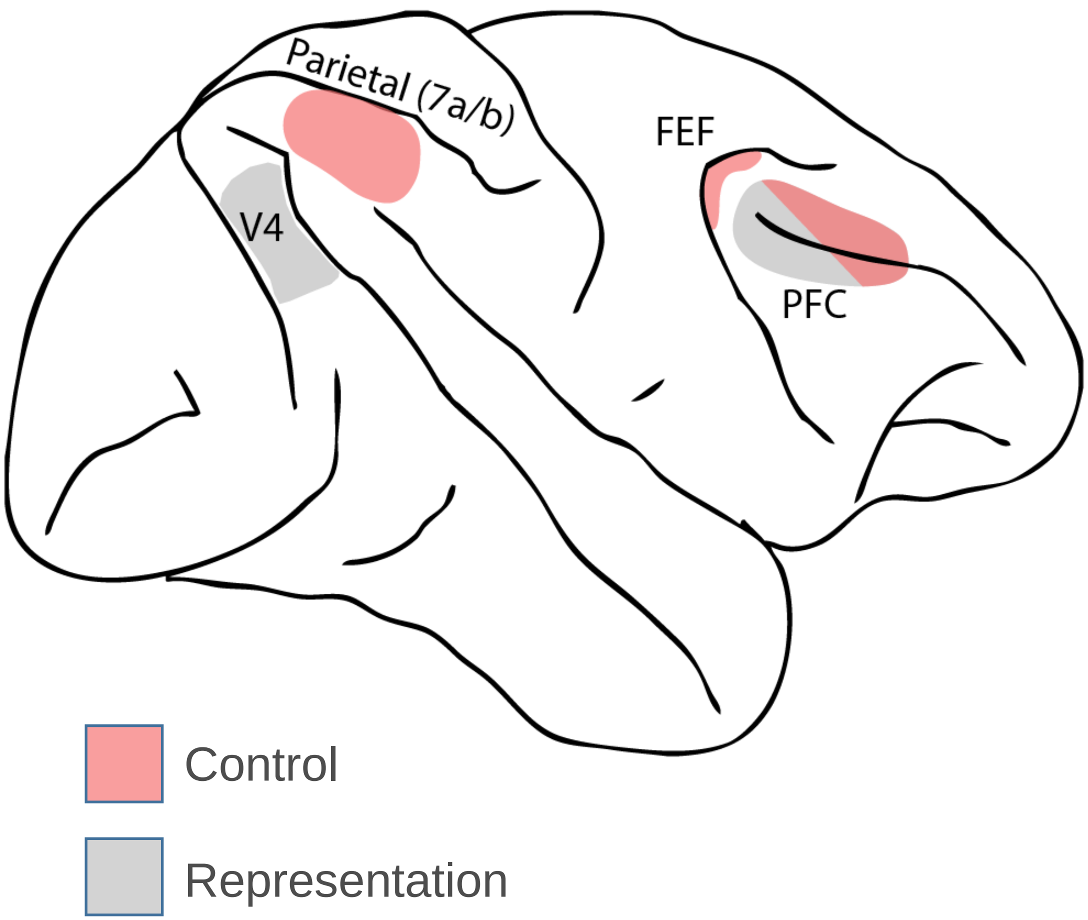




Panichello and Buschman, *Nature* 2021

Large-scale electrophysiological recordings across multiple brain regions involved in working memory

Recordings in two monkeys performing continuous working memory task.
Simultaneous recordings across visual, parietal and frontal cortex:



Over 1500 electrodes across two animals, yielded:

- 682 neurons in LPFC
- 187 neurons in FEF
- 331 neurons in 7a/b
- 341 neurons in V4/PIT
- 163 neurons in STG/TPoT
- 351 neurons in PMC

Panichello and Buschman, *Nature* 2021

Neural Mechanisms of Selection

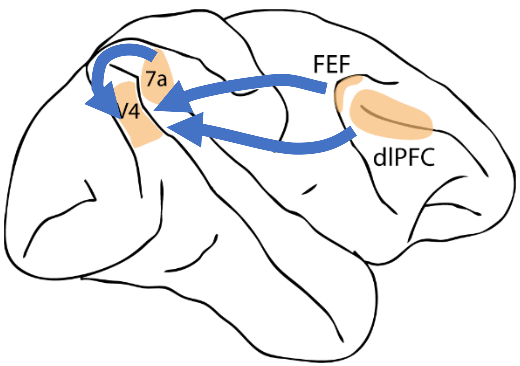
We are interested in addressing two questions:

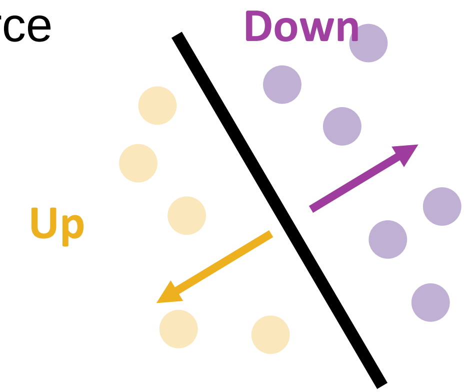
1) What are the neural mechanisms controlling the selection of items in working memory?

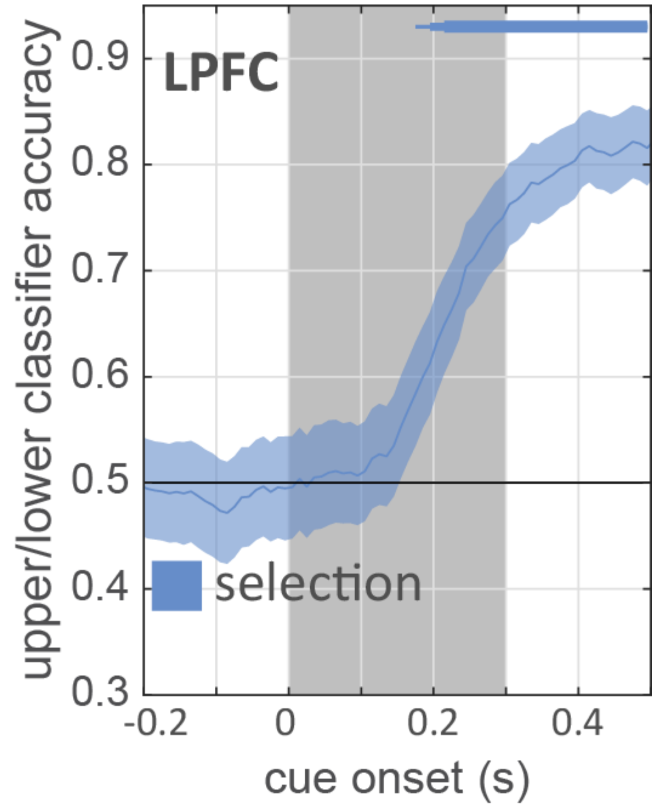
2) How does selection act on working memory representations?

Selection is directed from frontal cortex

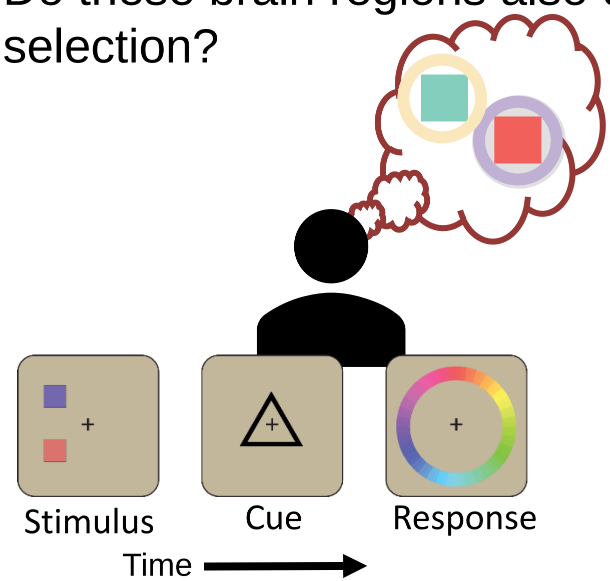
Previous work has shown frontal and parietal cortex are the source of internal control of attention.

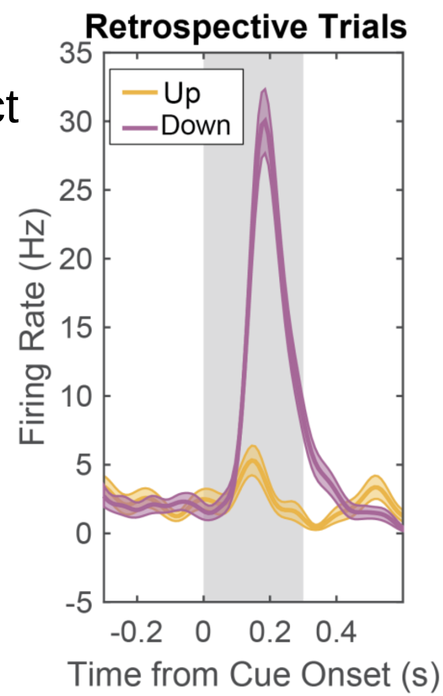






Do these brain regions also direct selection?

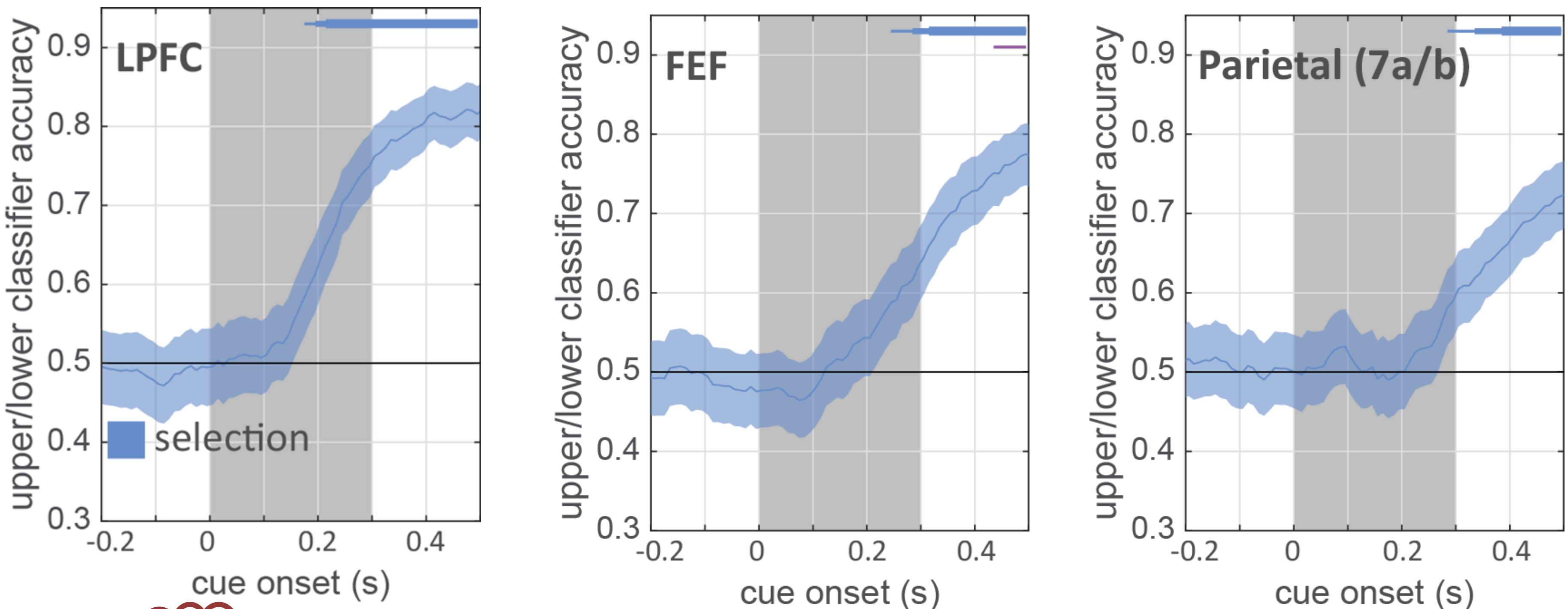


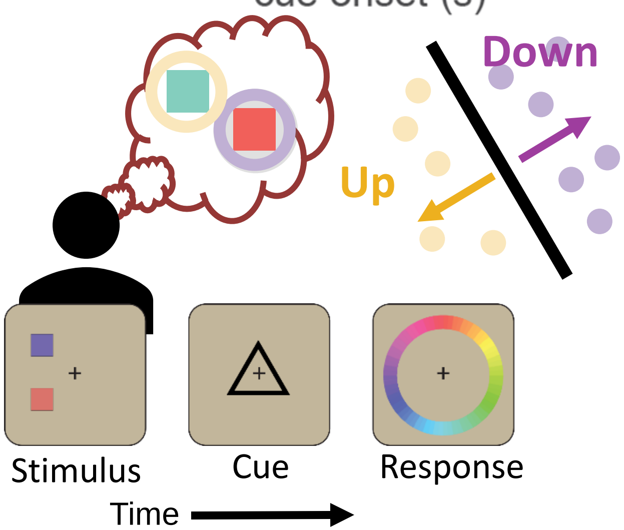


PFC neurons encode the direction of selection.

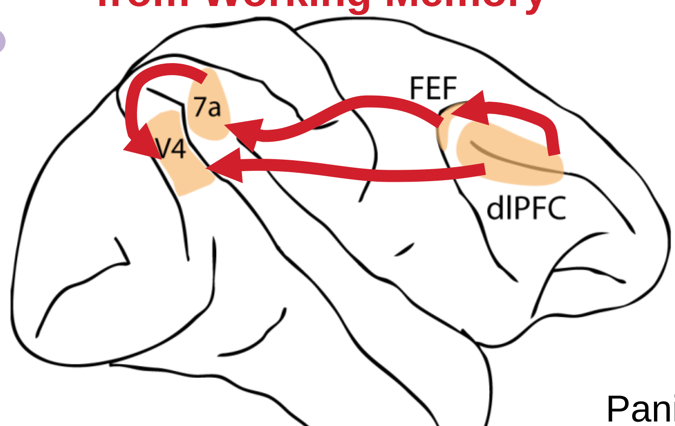
Panichello and Buschman, *Nature* 2021

Selection is directed from frontal cortex

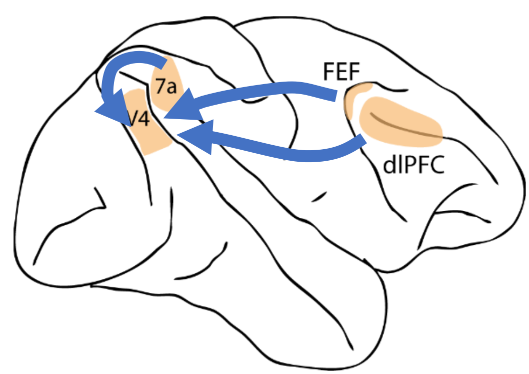




Control of Selection from Working Memory

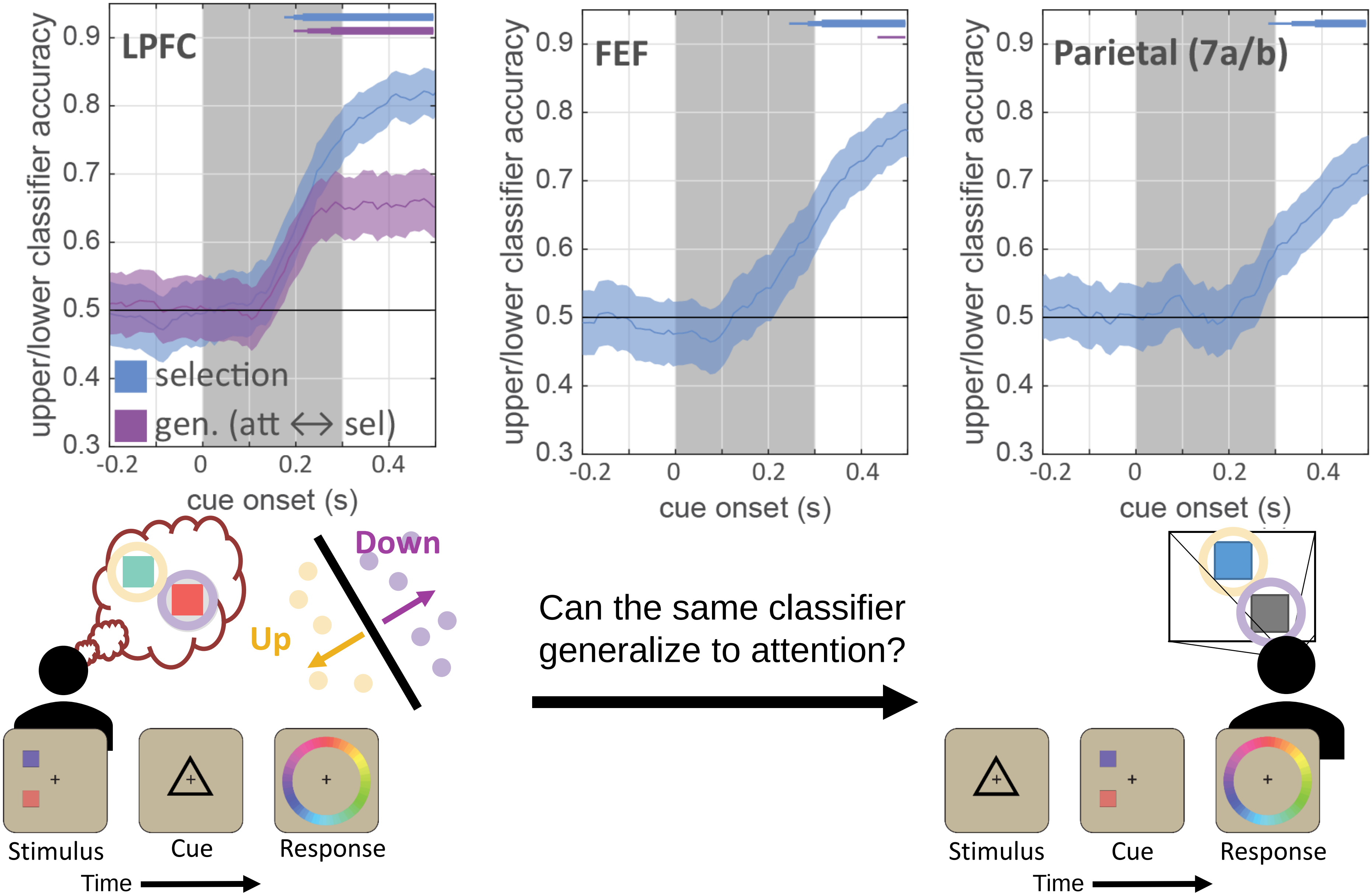


Control of Attention



Panichello and Buschman, *Nature* 2021

Selection is directed from frontal cortex; PFC has generalized control representation

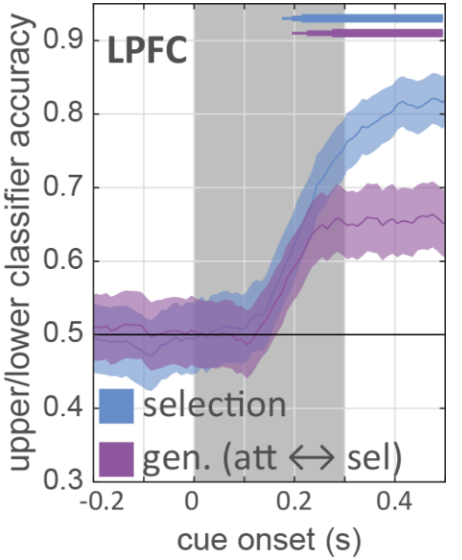


Can the same classifier generalize to attention?

Summary: Selection from Working Memory Relies on Similar Neural Mechanisms that Control Attention

1) What are the neural mechanisms controlling the selection of items in working memory?
   - Selective control of working memory originates in prefrontal cortex and flows back to parietal cortex.
   - Selection overlaps with attention in LPFC, suggesting PFC may be a domain general controller.

2) How does selection alter working memory representations?

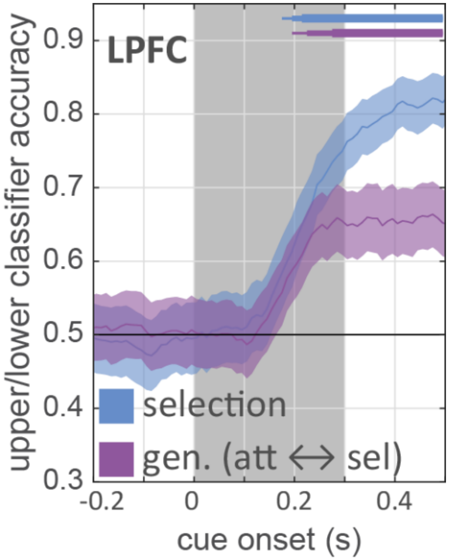


Summary: Selection from Working Memory Relies on Similar Neural Mechanisms that Control Attention

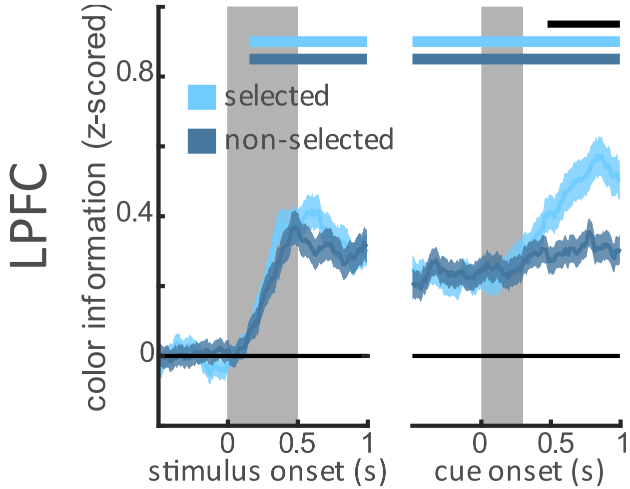
Working memory has a severe capacity limitation. Control of working memory helps to compensate for this capacity limitation.

Our results suggest that:

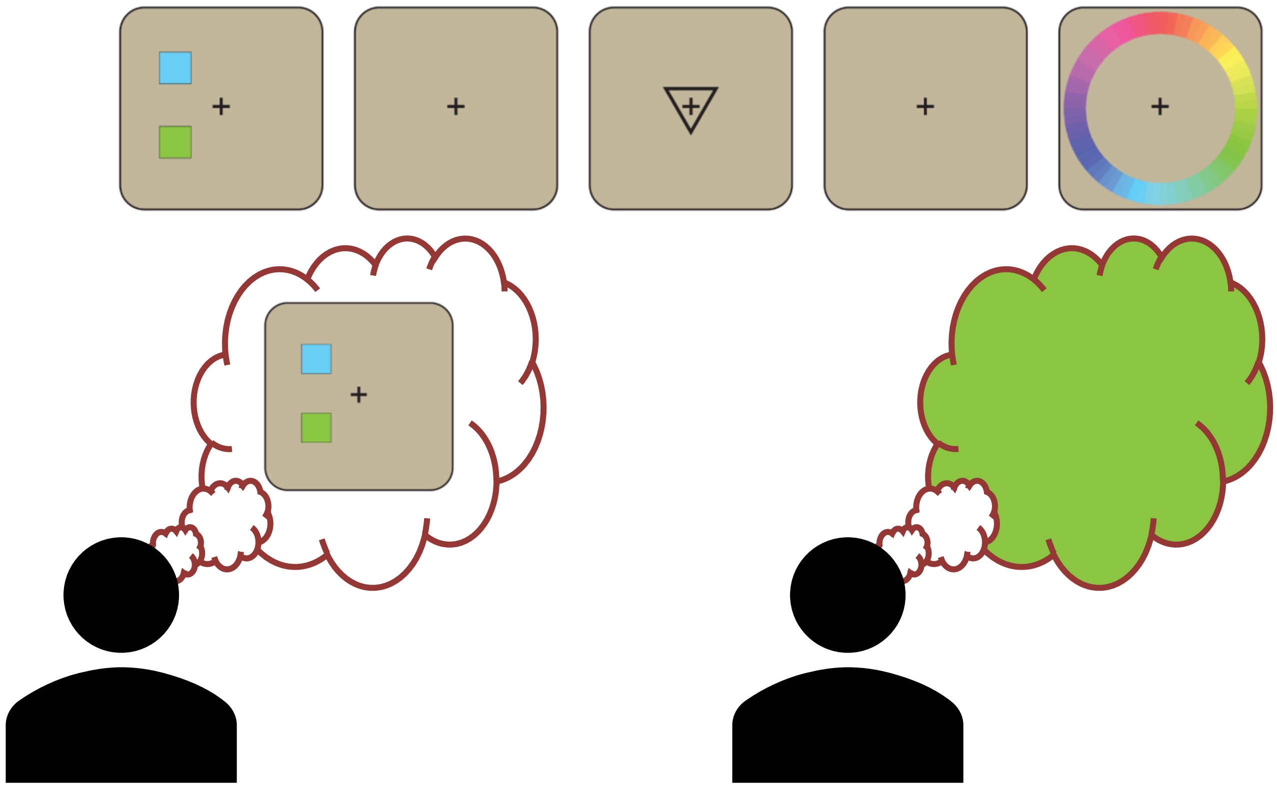
1) Selective control of working memory originates in prefrontal cortex and overlaps with control of attention, suggesting PFC may be a domain general controller.



2) Selection amplifies neural response of selected memory representation, similar to attention.



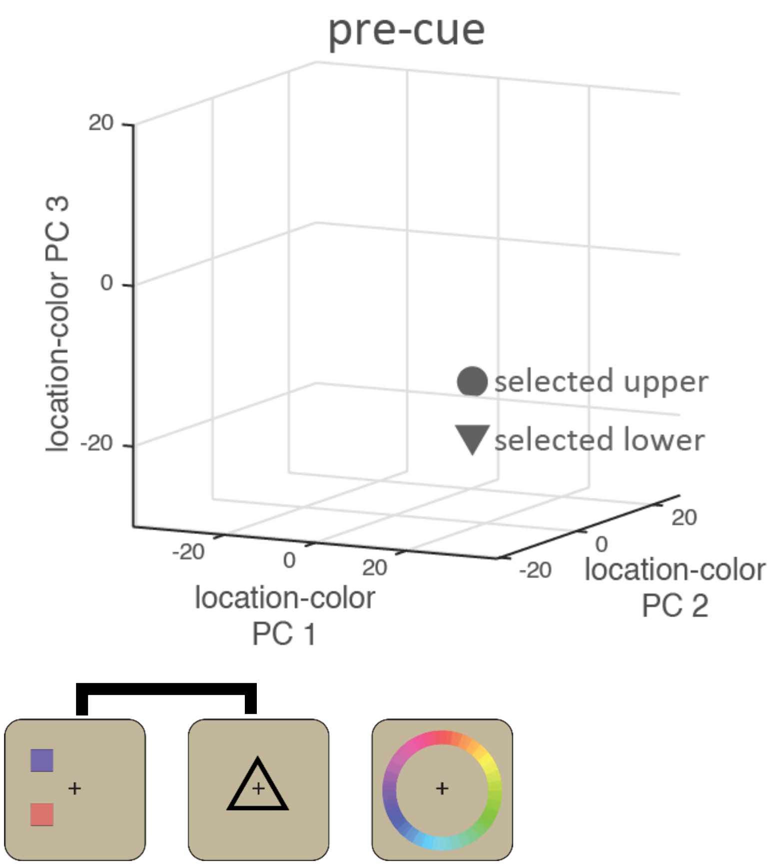
How items in working memory are going to be used changes during the task



Before the cue: maintain the color and location of the stimuli in working memory.

After the cue: maintain the color of the selected item and prepare to do visual search.

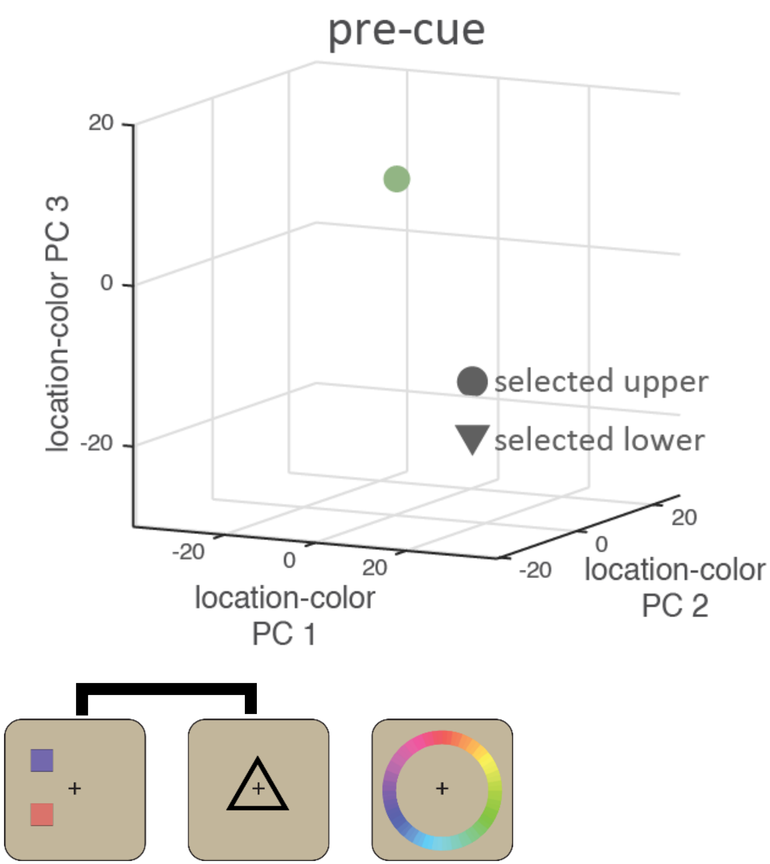
Before selection, items are maintained in independent subspaces in working memory.



Projecting high-dimensional ‘pseudo-population’ representations into low-dimensional PC space.

Panichello and Buschman, *Nature* 2021

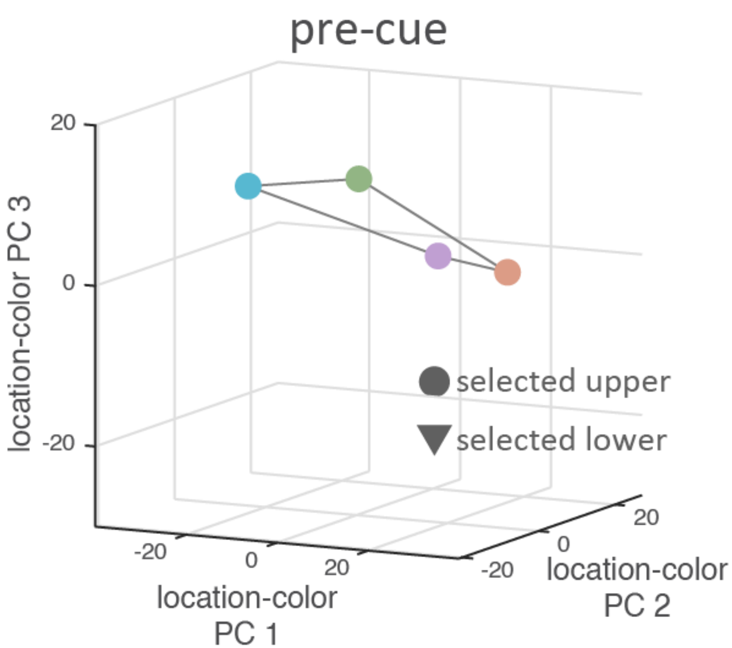
Before selection, items are maintained in independent subspaces in working memory.

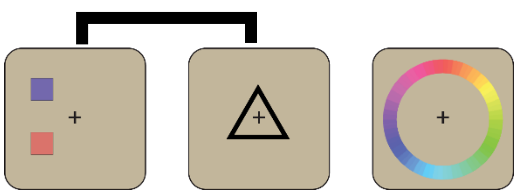


Projecting high-dimensional ‘pseudo-population’ representations into low-dimensional PC space.

Panichello and Buschman, *Nature* 2021

Before selection, items are maintained in independent subspaces in working memory.

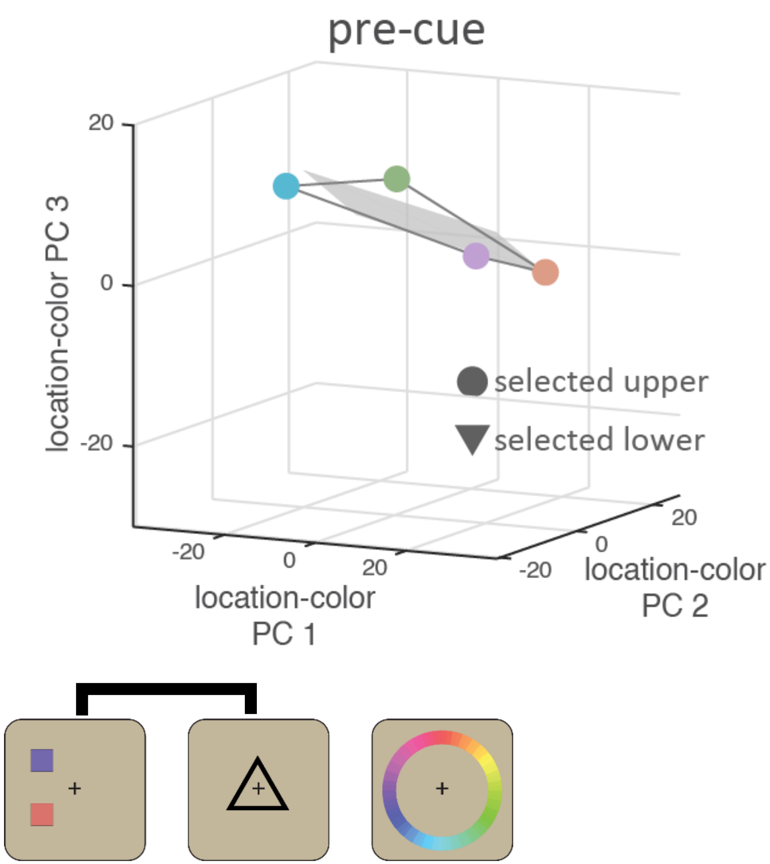




Projecting high-dimensional ‘pseudo-population’ representations into low-dimensional PC space.

Panichello and Buschman, *Nature* 2021

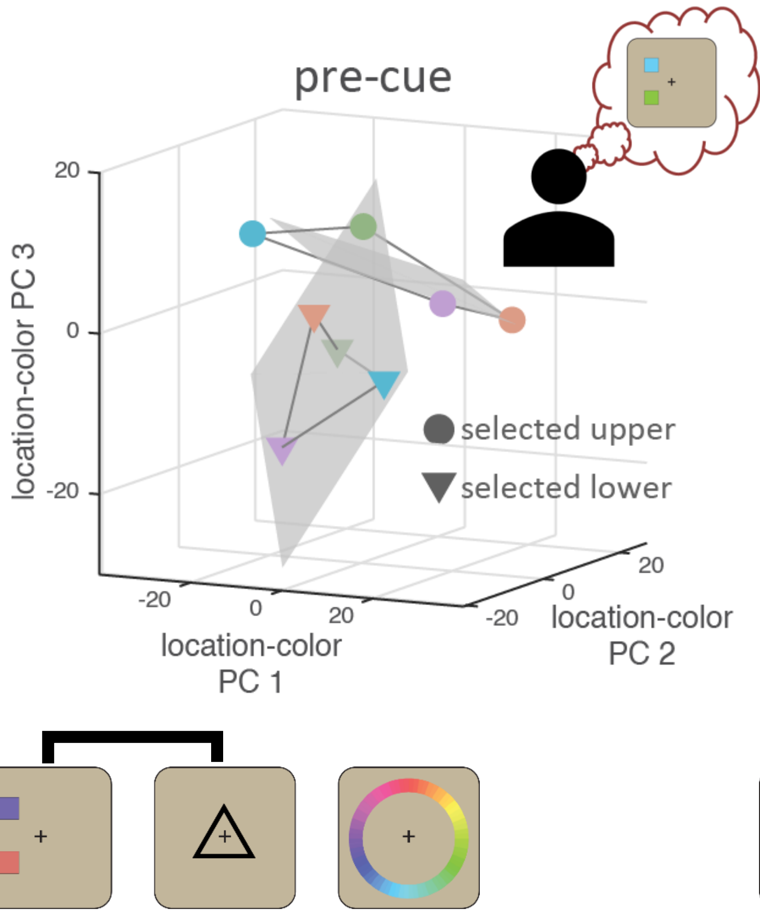
Before selection, items are maintained in independent subspaces in working memory.



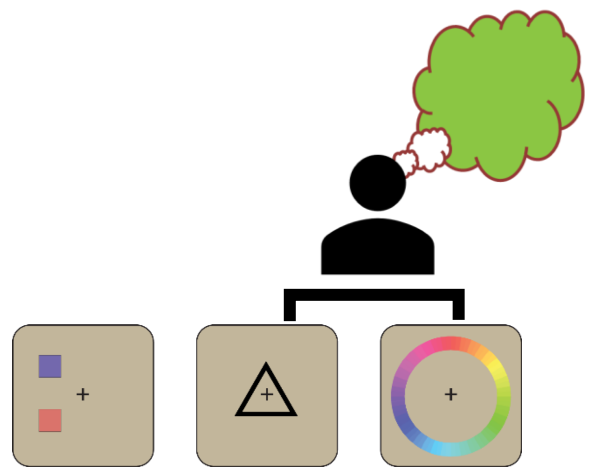
Projecting high-dimensional ‘pseudo-population’ representations into low-dimensional PC space.

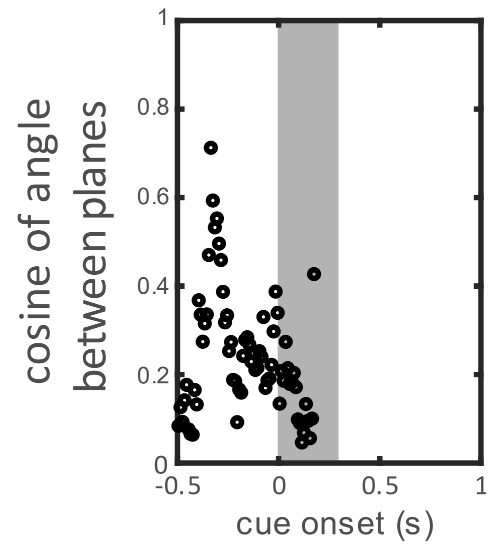
Panichello and Buschman, *Nature* 2021

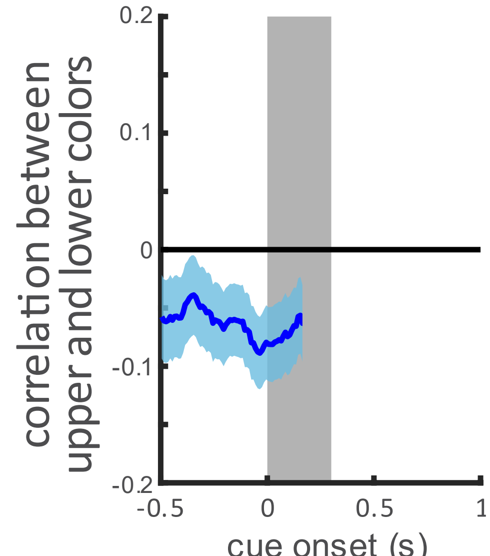
Before selection, items are maintained in independent subspaces in working memory.



What happens after selection?



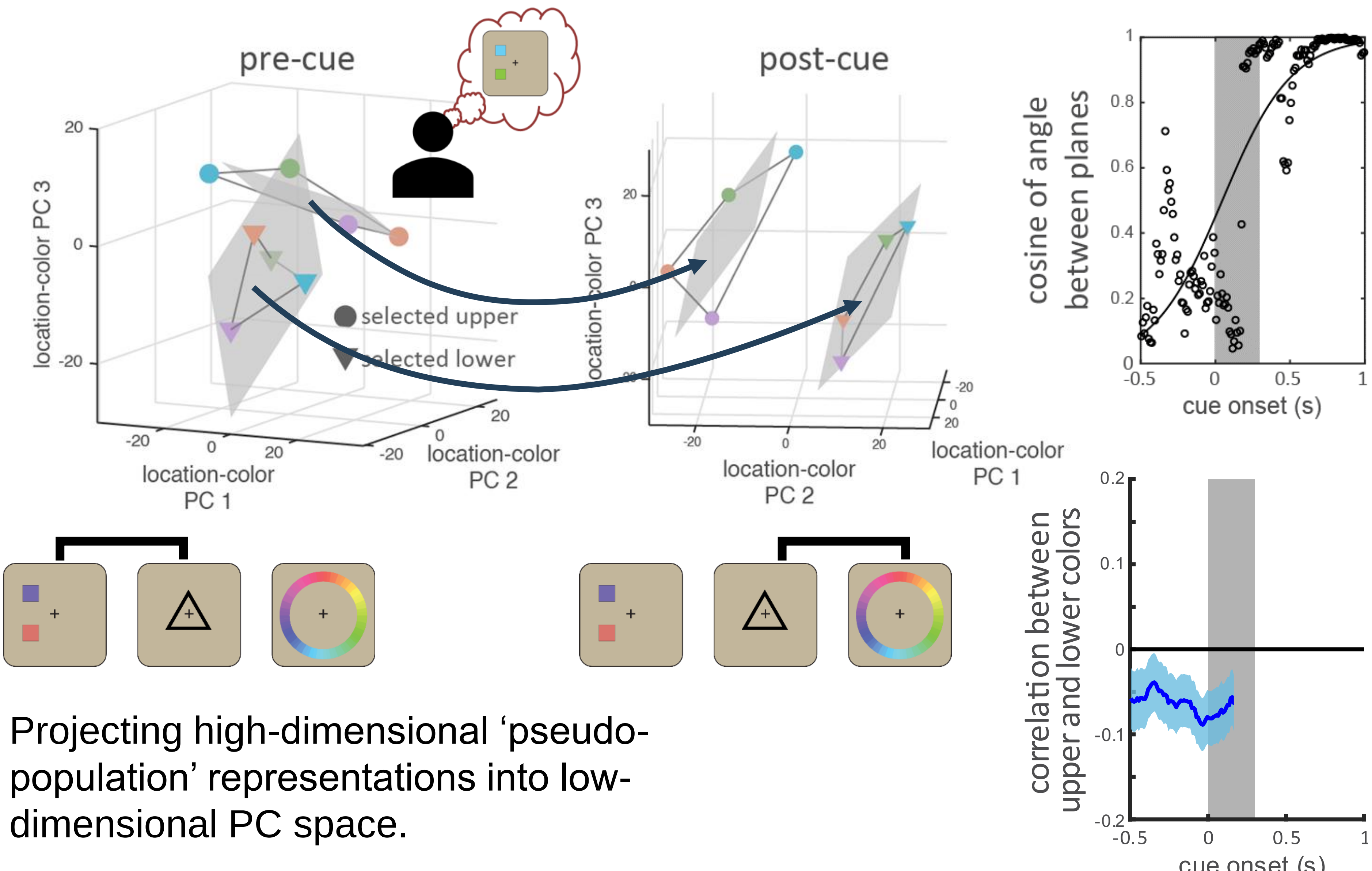




Projecting high-dimensional ‘pseudo-population’ representations into low-dimensional PC space.

Panichello and Buschman, *Nature* 2021

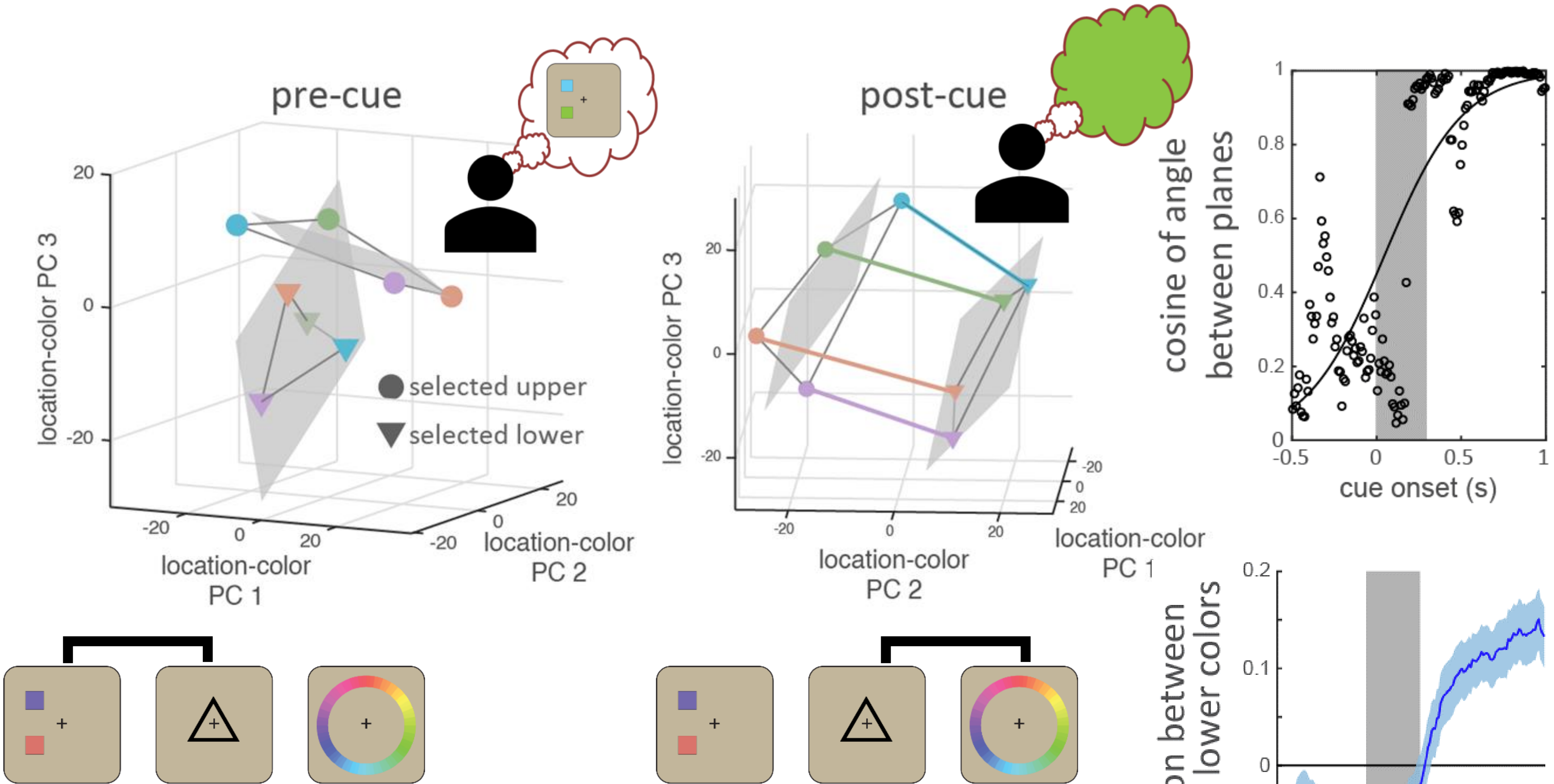
Before selection, items are in independent subspaces. After selection, representations are aligned.

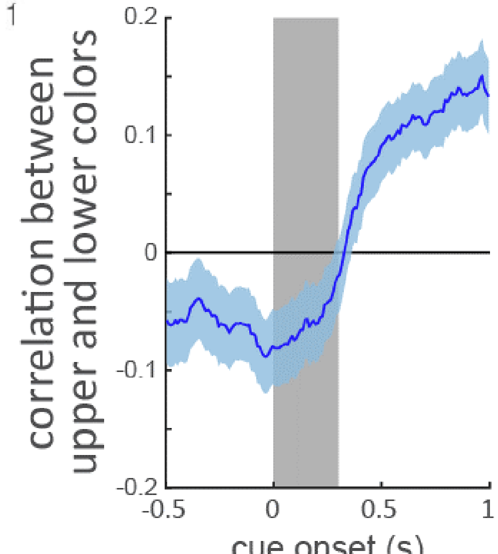


Projecting high-dimensional ‘pseudo-population’ representations into low-dimensional PC space.

Panichello and Buschman, *Nature* 2021

Before selection, items are in independent subspaces. After selection, representations are aligned.

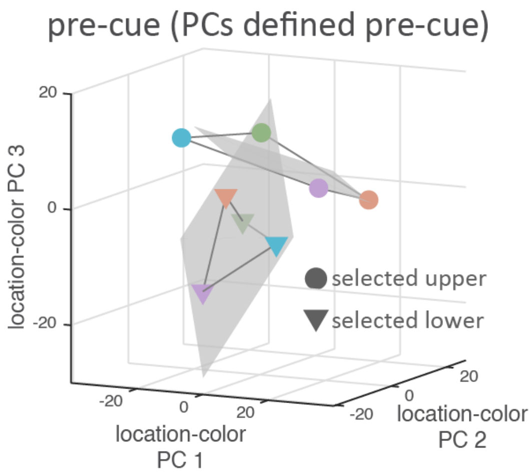


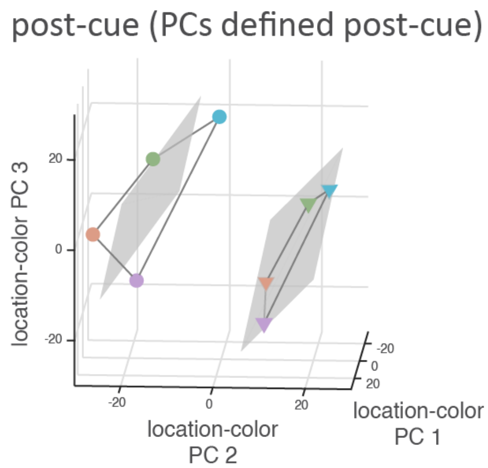


Projecting high-dimensional ‘pseudo-population’ representations into low-dimensional PC space.

Panichello and Buschman, *Nature* 2021

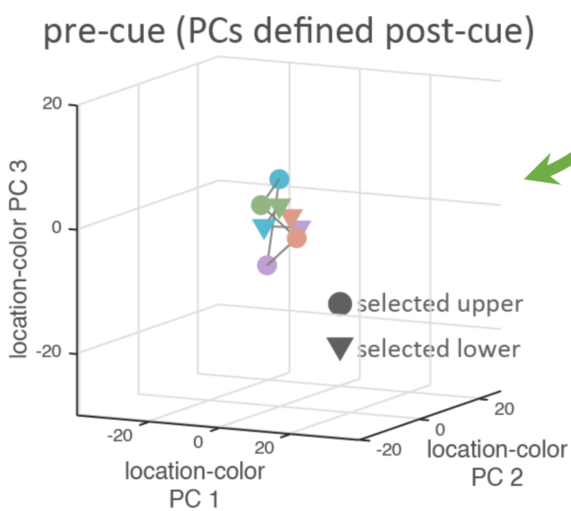
Selection rotates representation of selected item into an actionable subspace.





These subspaces have been defined independently for each time period.

Are pre-cue and post-cue representations in the same or different subspace?



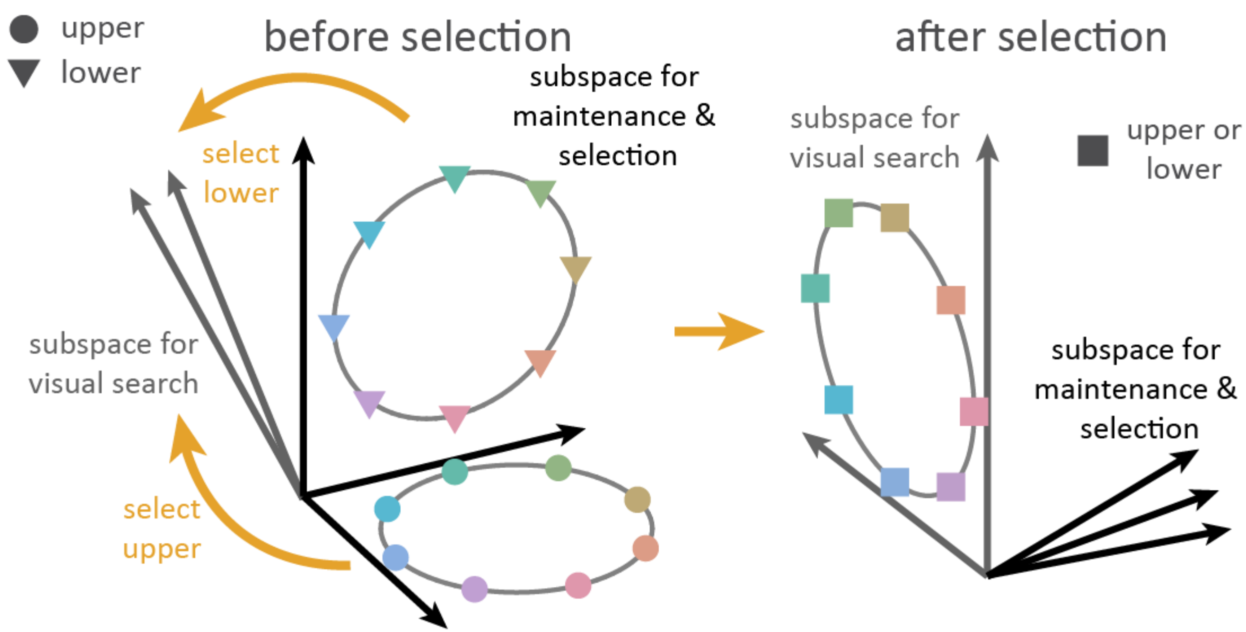
Use these PCs!

To test, we can use the post-cue subspace to try to decode pre-cue representations.

Panichello and Buschman, *Nature* 2021

Summary: Selection from Working Memory Dynamically Changes the Geometry of Neural Representations

- Selection dynamically transforms memory representations, facilitating ‘read-out’ of task-relevant information.
- This may be a mechanism of cognitive control, allowing the brain to control how information is read-out by task-specific networks.



Panichello and Buschman, *Nature* 2021