Scene representation and visual search

09.09.2021

Raul Grieben

Introduction

- Most **DFT models** of **higher cognition** share two core **sub-networks** that are crucial for object-oriented **interaction** with the **environment**.

Introduction

- Most DFT models of higher cognition share two core sub-networks that are crucial for object-oriented interaction with the environment.
- The **first** is the ***visual search sub-network***, that consists of a **bottom-up** feed-forward feature-extraction path and a **top-down** guidance path.

Introduction

- Most DFT models of higher cognition share two core sub-networks that are crucial for object-oriented interaction with the environment.
- The first is the *visual search sub-network*, that consists of a feed-forward feature-extraction path and a top-down guidance path.
- The **second** is the ***scene memory sub-network***, that **autonomously** builds **working memory** feature representations of **previously attended objects**.

Introduction

- Here I am going to present a **neural dynamic process model** that builds on these two core sub-networks to account for the **difference** between **feature** and **conjunctive search**.

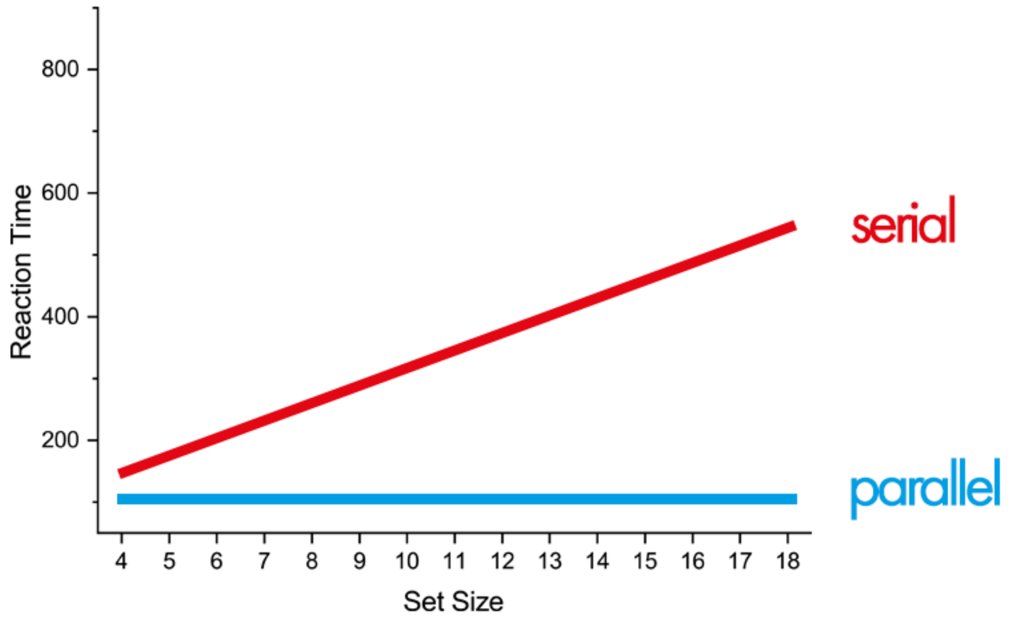
Introduction

- Here I am going to present a neural dynamic process model that builds on these two core sub-networks to account for the difference between feature and conjunctive search.
- In this context, I will **address the question** of whether both the overall **speed** and the **efficiency** of conjunctive visual **search** can be **improved** by scene **memory**.

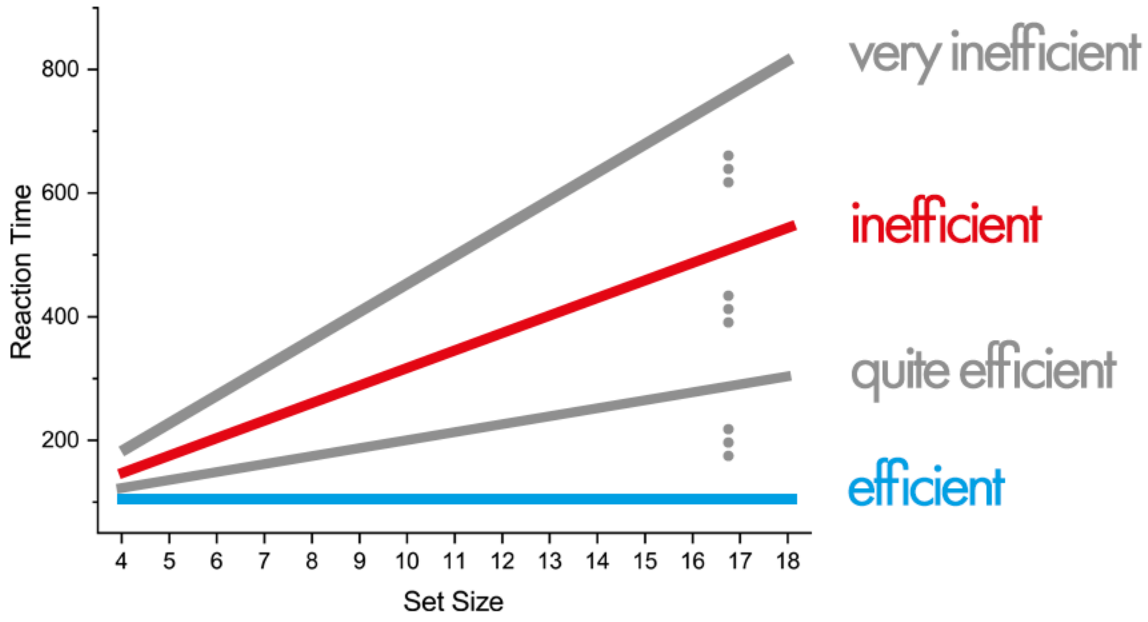
Introduction

- Here I am going to present a neural dynamic process model that builds on these two core sub-networks to account for the difference between feature and conjunctive search.
- In this context, I will address the question of whether both the overall speed and the efficiency of conjunctive visual search can be improved by scene memory.
- I will also explain how we **extended** this **model** to understand the **interplay** between **bottom-up** processing **and top-down** guidance in visual **search**, an issue in need of theoretical resolution (Proulx, 2007).

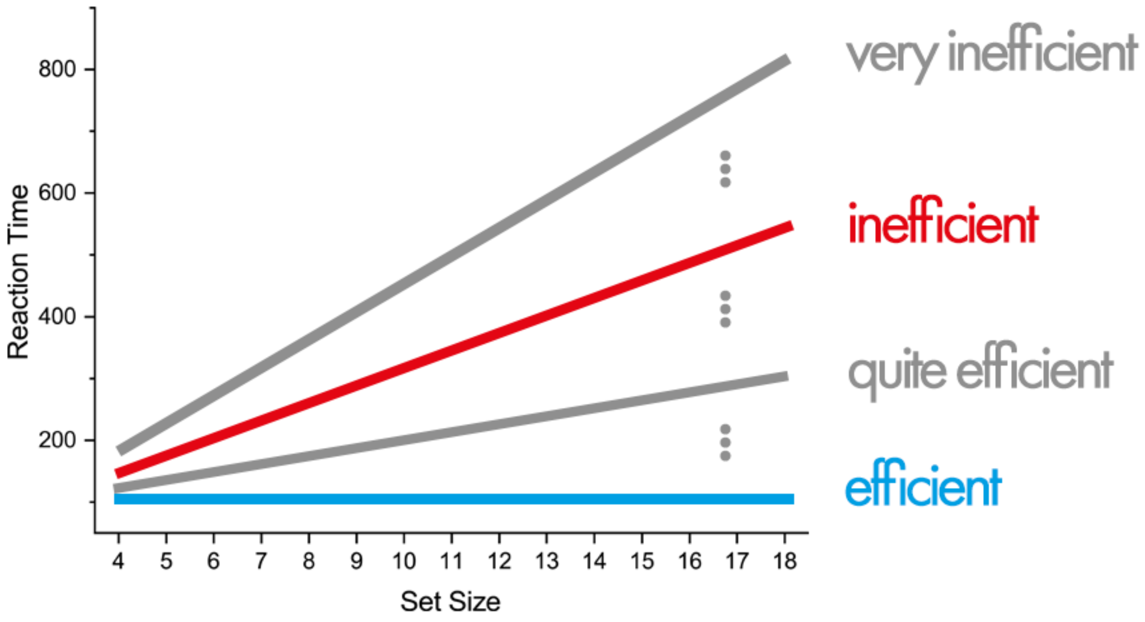
Proulx. Bottom-up guidance in visual search for conjunctions. JEP: Human Perception and Performance (2007)



In the classical view of Anne **Treisman**, visual **search** was either **parallel** or **serial**.

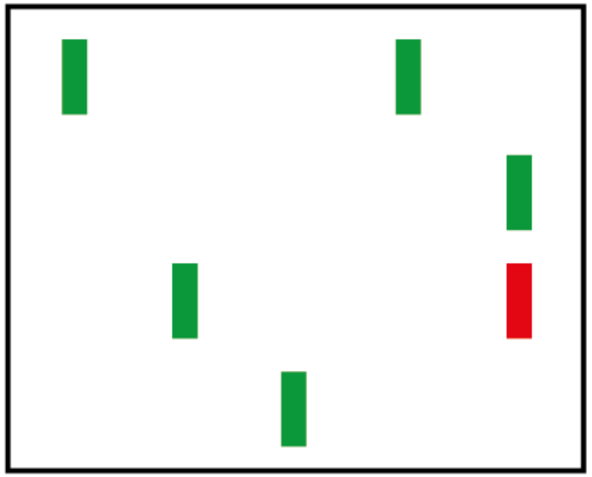


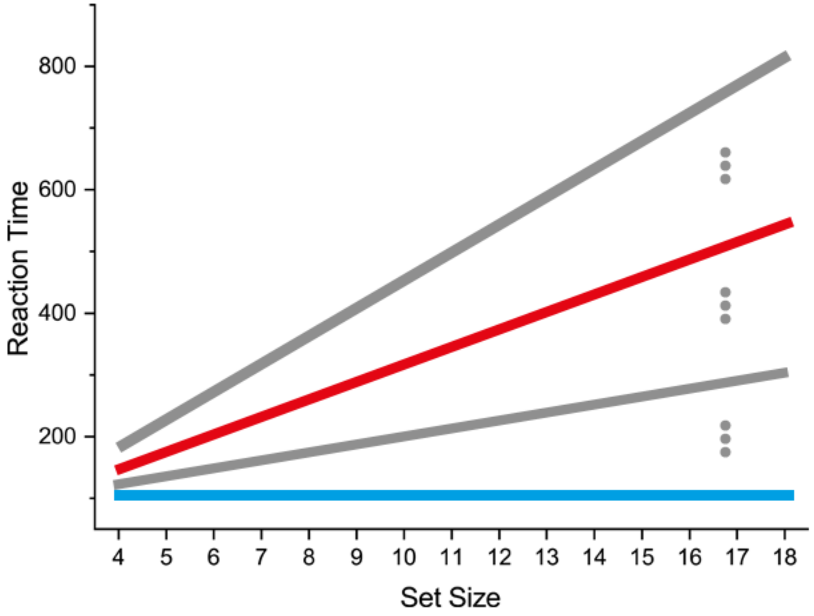
Jeremy **Wolfe**, on the other hand, described the **efficiency** of visual **search** as forming a **continuum**.



He defined the **slope** of the RT against set size function as the **measure** of **efficiency**.

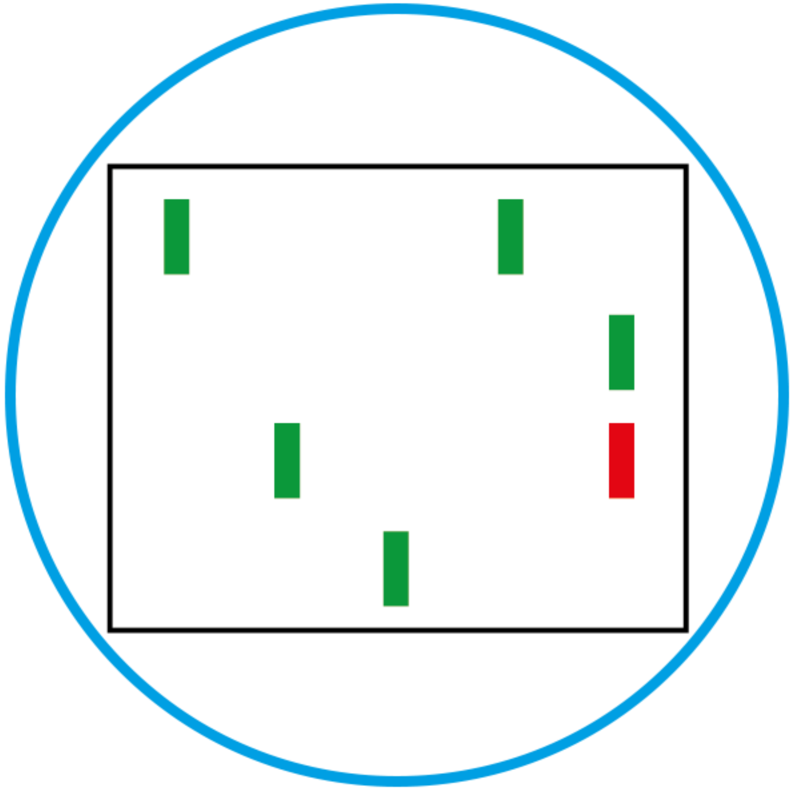
single feature search



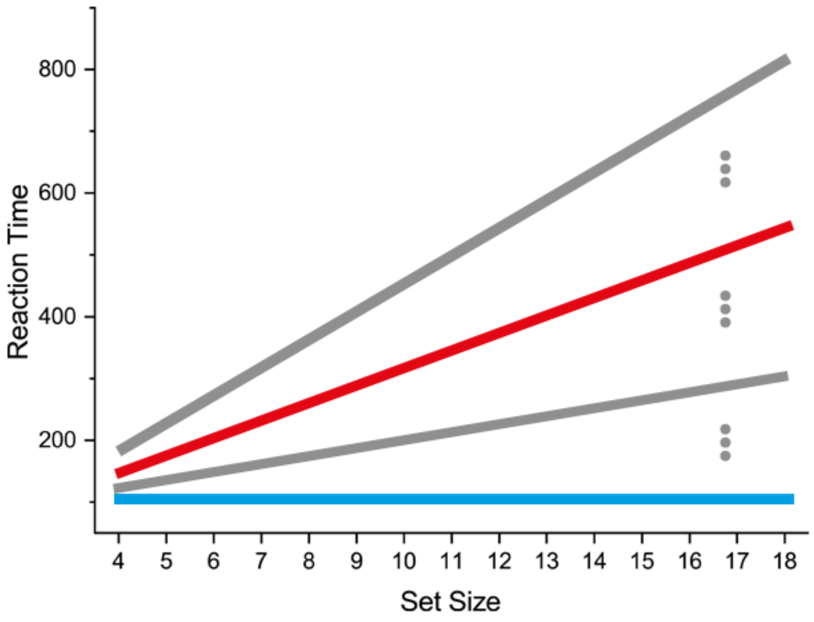


By this measure, single **feature search** is **efficient** as the reaction times are **independent** of **set size**.

single feature search



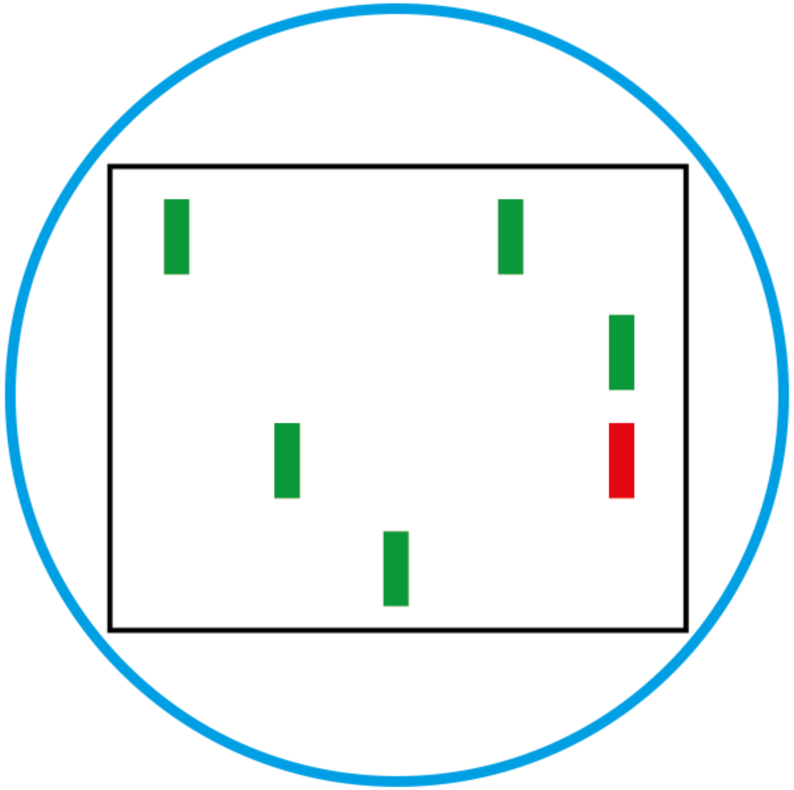
efficient

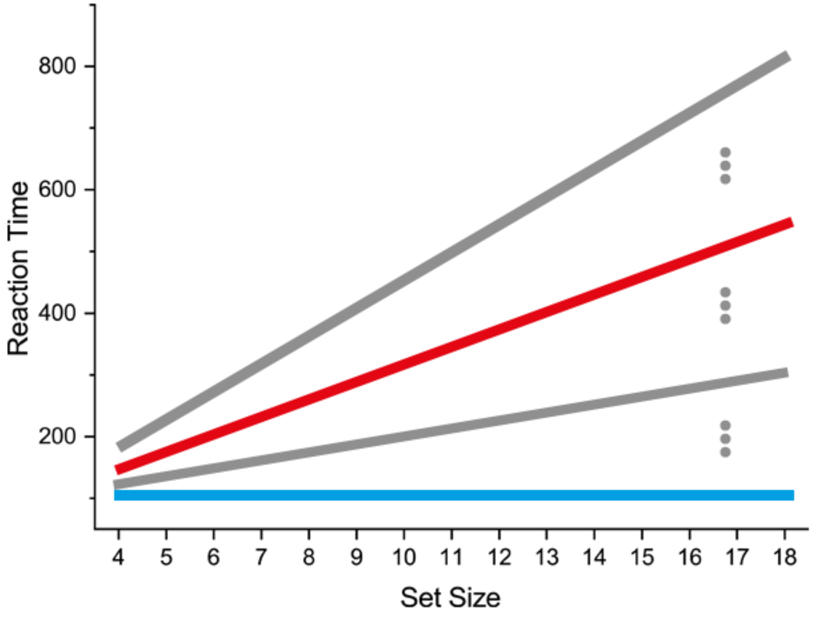


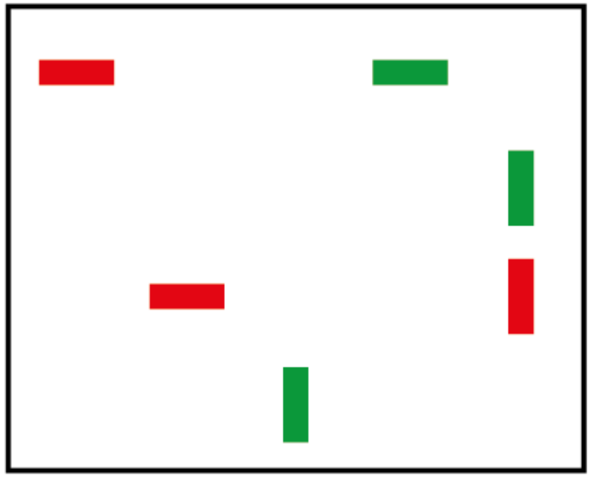
The **target pops out**.

single feature search

conjunctive search



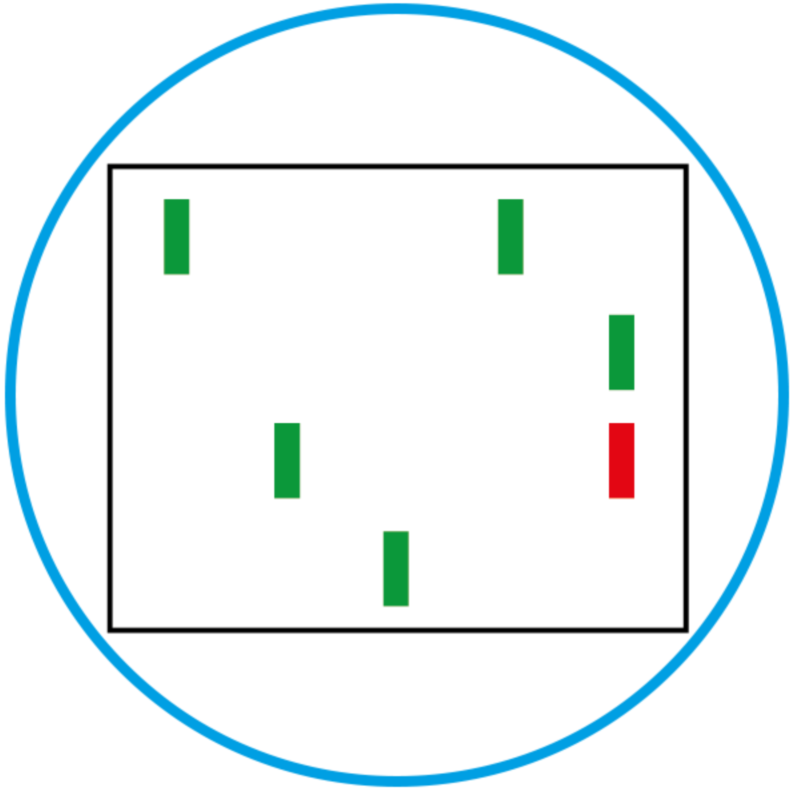




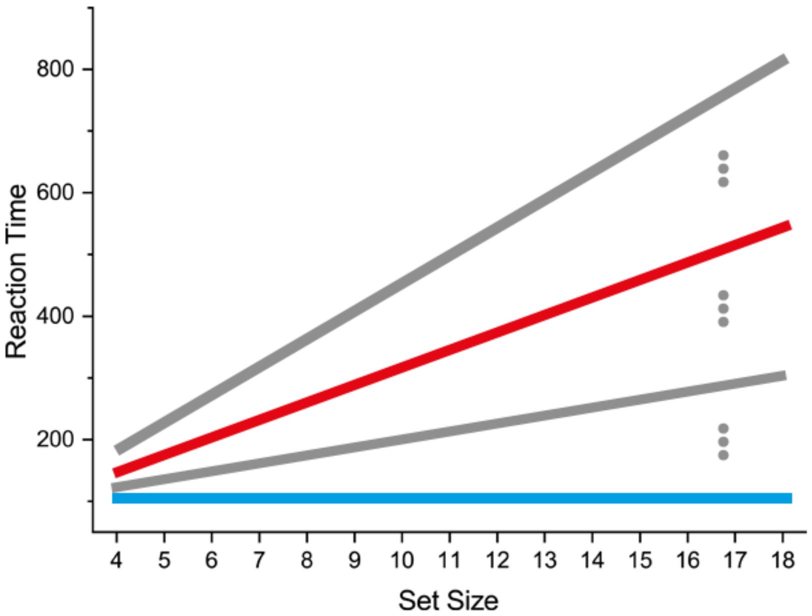
efficient

In the **conjunctive** condition RTs are **proportional** to the number of **distractor items**.

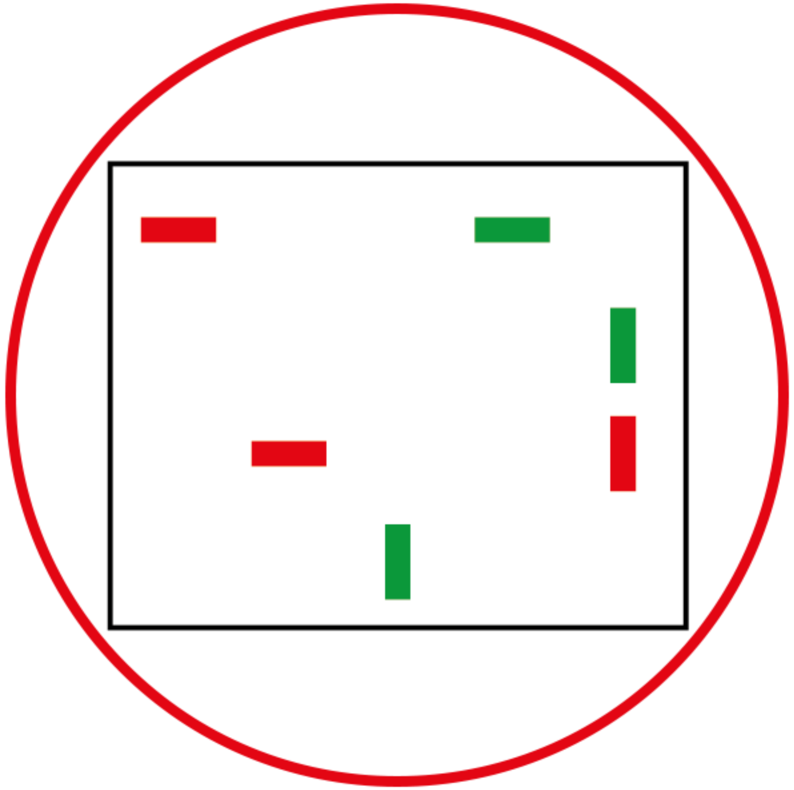
single feature search



efficient



conjunctive search



inefficient

Conjunctive search is, therefore, considered **inefficient**.

Visual search and scene memory

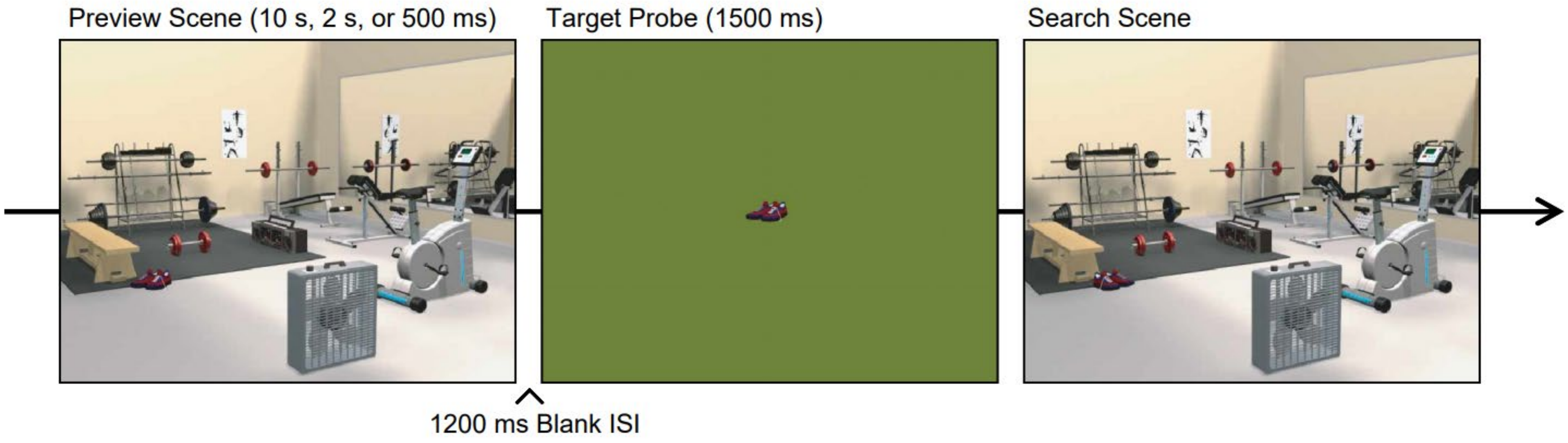
- The **role** of **memory** in visual **search** has been **intensely studied** in a variety of experimental paradigms.

Visual search and scene memory

- The role of memory in visual search has been intensely studied in a variety of experimental paradigms.
- A prominent paradigm is the **preview paradigm**.

Visual search and scene memory

- The role of memory in visual search has been intensely studied in a variety of experimental paradigms.
- A prominent paradigm is the preview paradigm.



Hollingworth. Two forms of scene memory guide visual search. Visual Cognition (2009)

Visual search and scene memory

- The role of memory in visual search has been intensely studied in a variety of experimental paradigms.
- A prominent paradigm is the preview paradigm.
- Using this paradigm in a **naturalistic** setting, **Hollingworth** found **benefits** of scene preview.

Hollingworth. Two forms of scene memory guide visual search. Visual Cognition (2009)

Visual search and scene memory

- **Hillstrom** and colleagues **extended** this work by showing that information on the **gist** of **scene** can **improve** search **efficiency**.

Hillstrom et al. The effect of the first glimpse at a scene on eye movements during search. Psychon Bull Rev (2012)

Visual search and scene memory

- Hillstrom and colleagues extended this work by showing that information on the gist of scene can improve search efficiency.
- These **effects** were **not found** for **randomly ordered** search **arrays**, indicating that it is **specific** to **naturalistic scenes**.

Visual search and scene memory

- Hillstrom and colleagues extended this work by showing that information on the gist of scene can improve search efficiency.
- These effects were not found for randomly ordered search arrays, indicating that it is specific to naturalistic scenes.
- A **common finding** in the **preview paradigm** is that **mean RTs** are **reduced** if a preview of the search array is provided.

Visual search and scene memory

- **Becker** and **Pashler** argued that this provides **strong evidence** for **guidance** of attention **by VWM**.

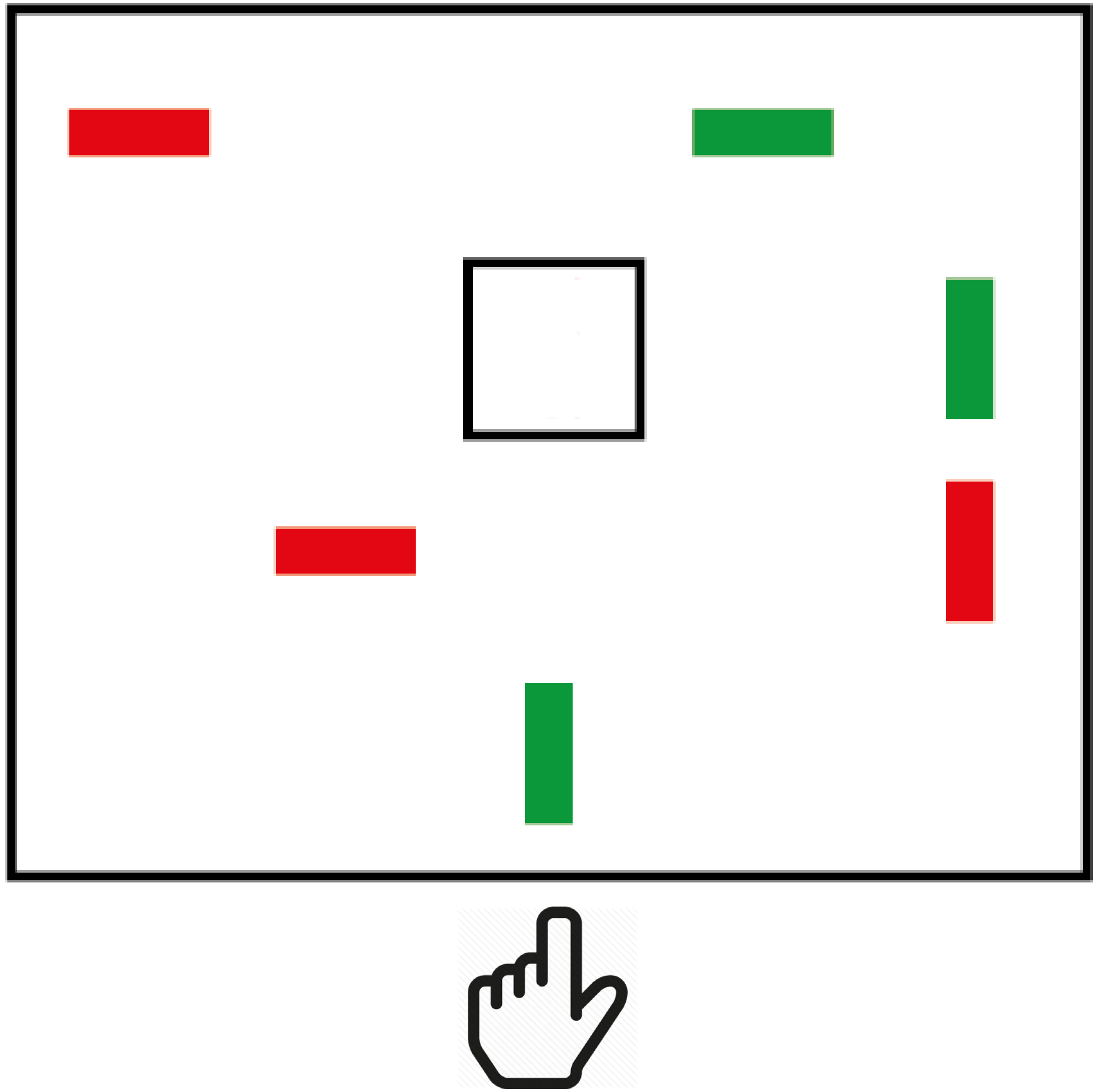
Becker and Pashler. Awareness of the continuously visible. Perception & Psychophysics (2005)

Visual search and scene memory

- Becker and Pashler argued that this provides strong evidence for guidance of attention by VWM.
- In their experiments, **efficiency** was **not increased** by preview, however.

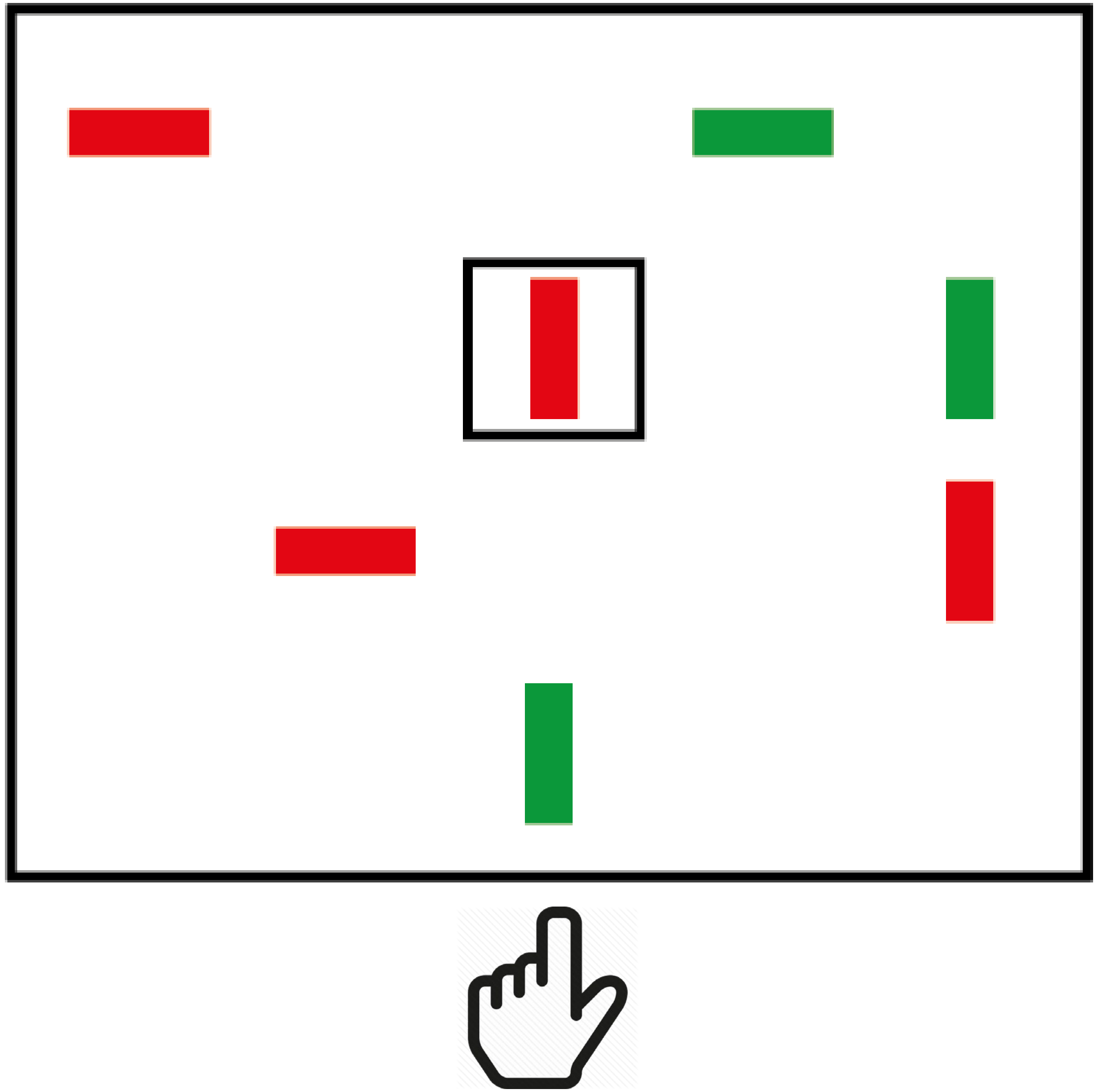
Becker and Pashler. Awareness of the continuously visible. Perception & Psychophysics (2005)

Scenario



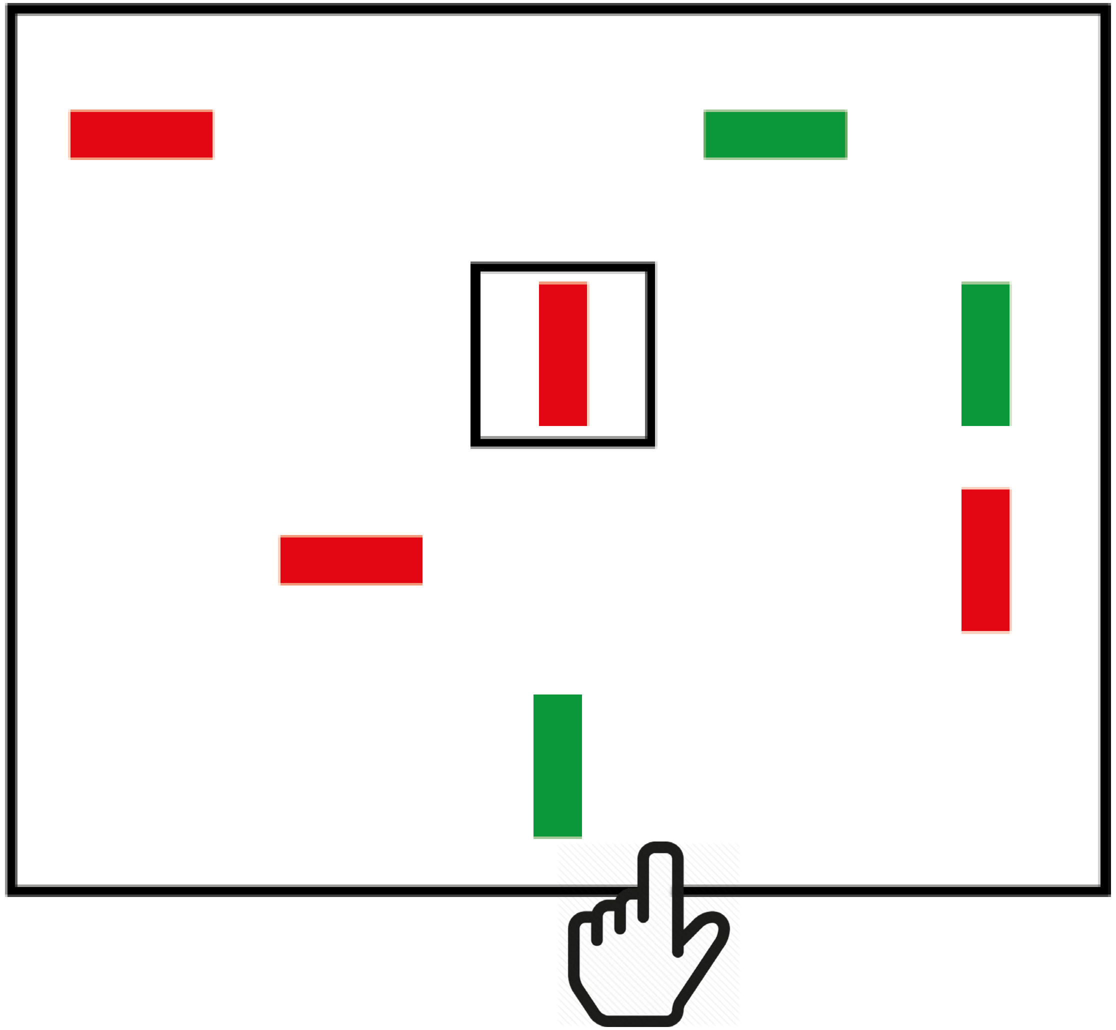
- Both **experiments** and **model** simulations are **based** on a **scenario**, in which participants **explore** a visual **scene**

Scenario



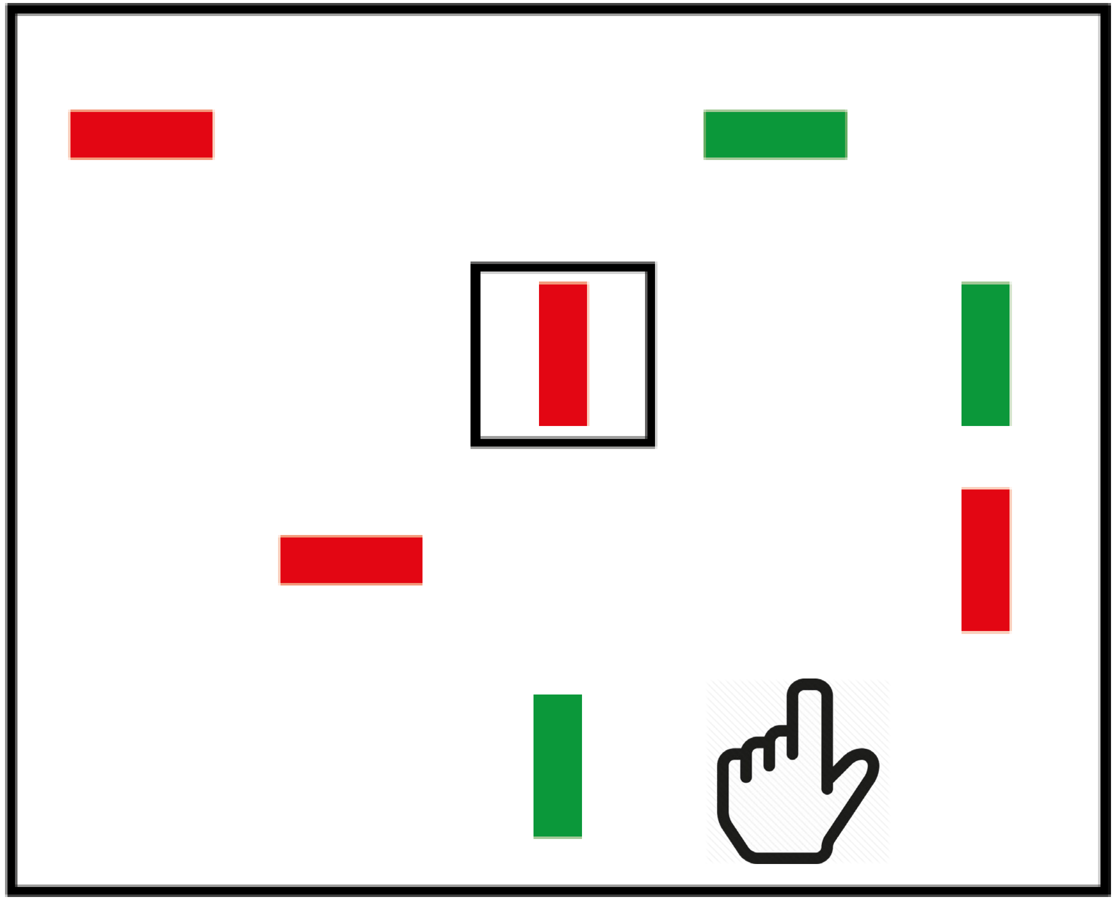
- Both experiments and model simulations are based on a scenario, in which participants explore a visual scene, are **cued** at some point to a visual **search task** by a sample **target** object that **appears** in the visual **array**

Scenario



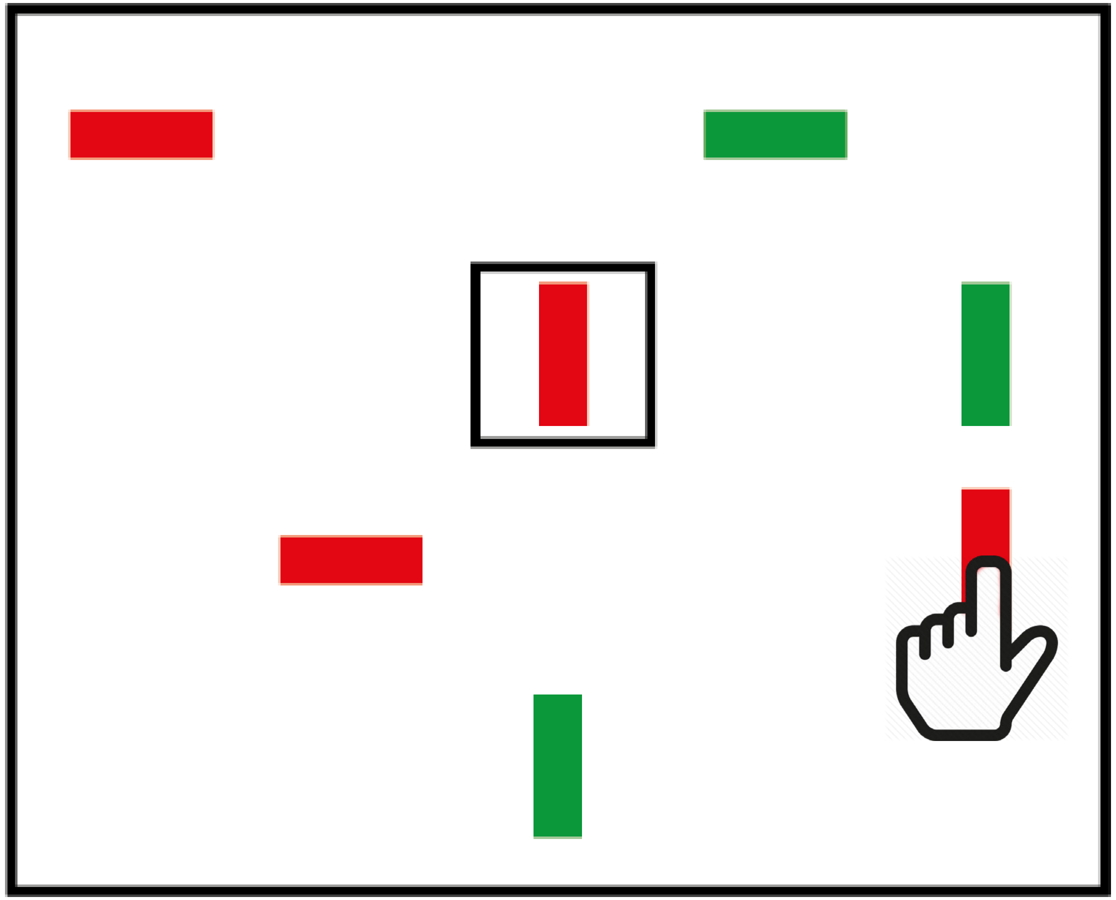
- Both experiments and model simulations are based on a scenario, in which participants explore a visual scene, are cued at some point to a visual search task by a sample target object that appears in the visual array, and then **respond** by **indicating** the **location** of a **matching** visual **object**.

Scenario



- Both experiments and model simulations are based on a scenario, in which participants explore a visual scene, are cued at some point to a visual search task by a sample target object that appears in the visual array, and then **respond** by **indicating** the **location** of a **matching** visual **object**.

Scenario



- Both experiments and model simulations are based on a scenario, in which participants explore a visual scene, are cued at some point to a visual search task by a sample target object that appears in the visual array, and then **respond** by **indicating** the **location** of a **matching** visual **object**.

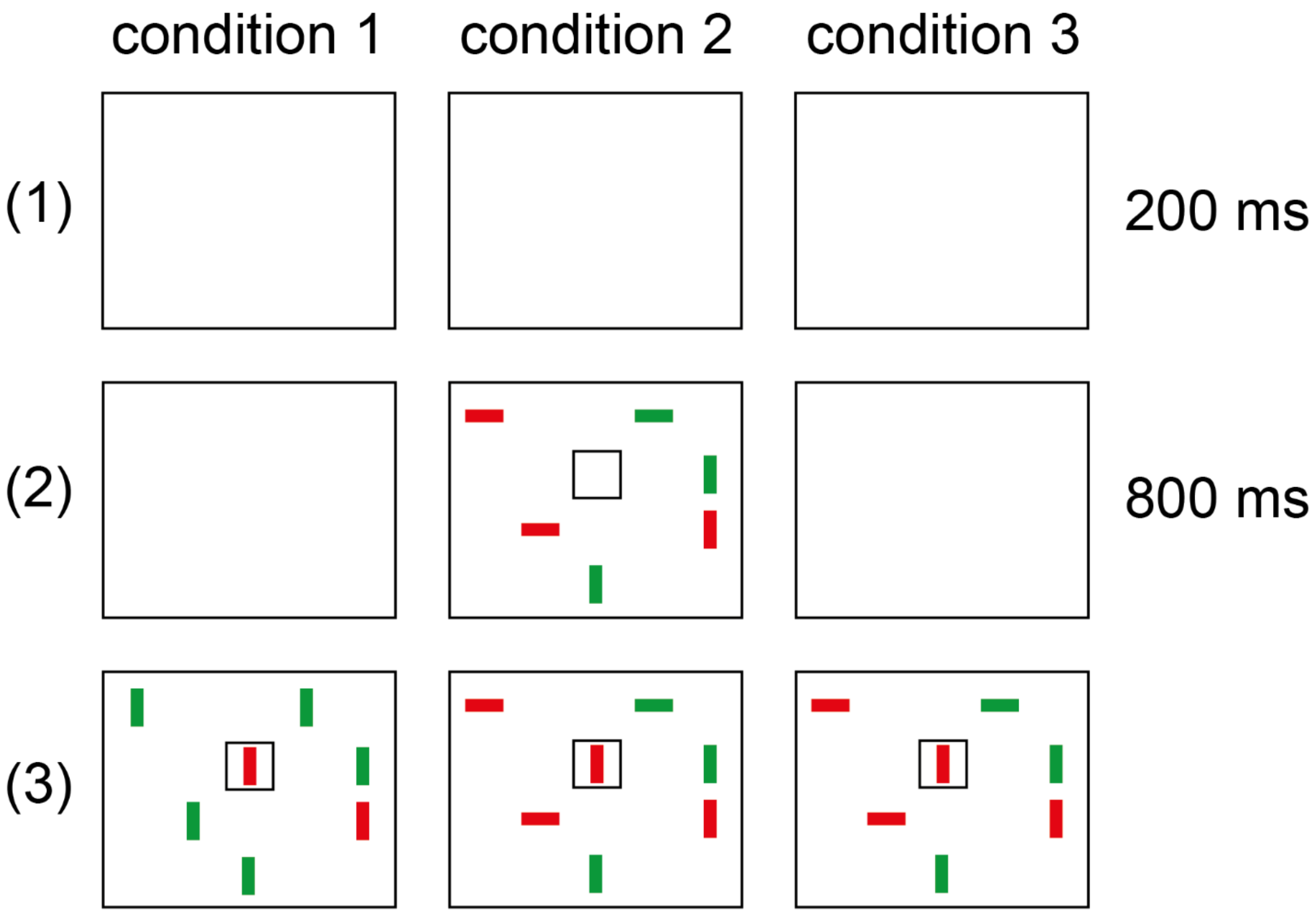
Scenario

- We **chose** the scene **preview paradigm** as a key behavioral **task** to address with the DFT **model**.

Scenario

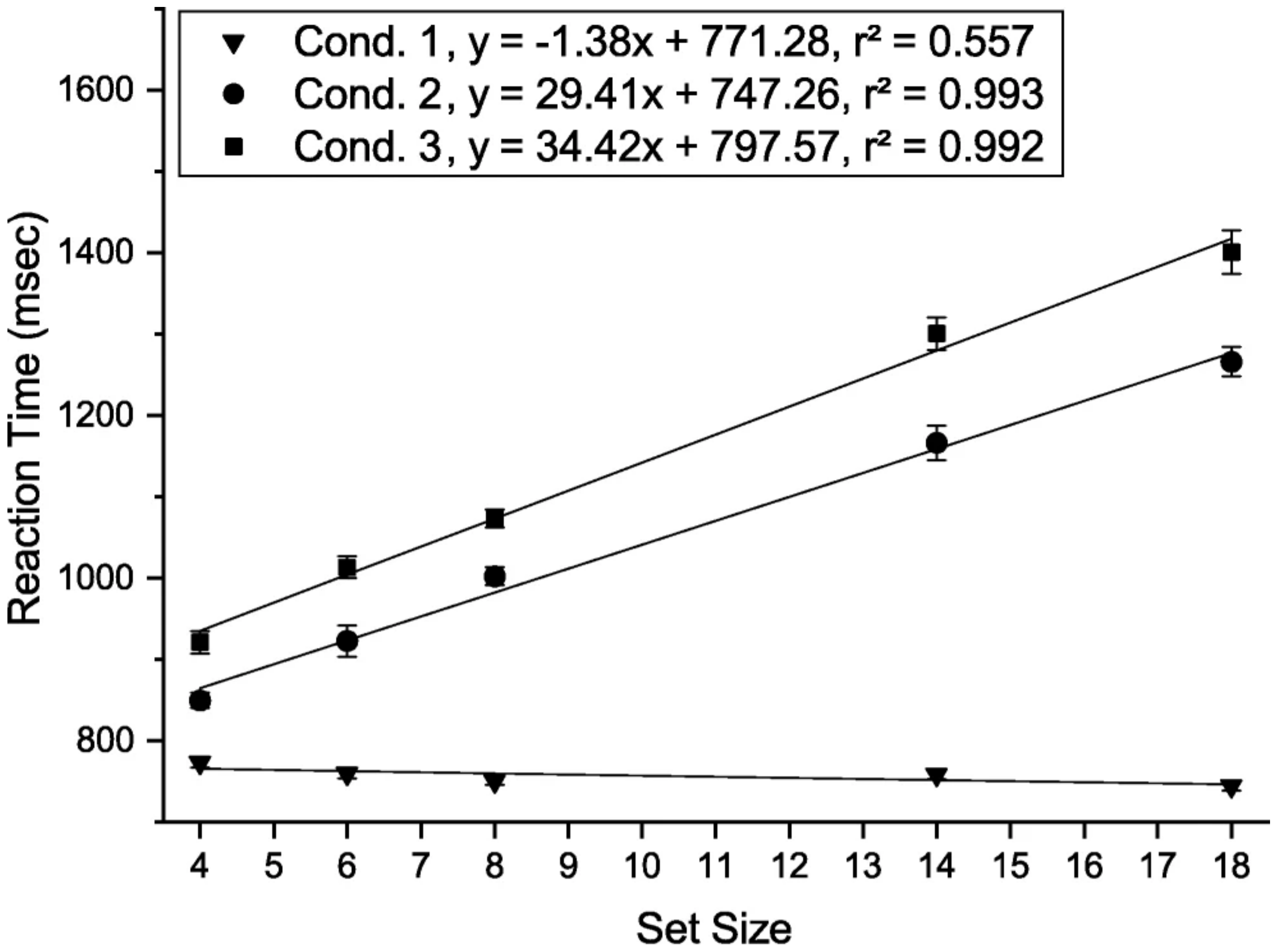
- We chose the scene preview paradigm as a key behavioral task to address with the DFT model.
- We **specifically addressed** the **question**, why **preview** benefits observed for **natural** scenes did **not generalize** to **randomly** arranged search arrays.

Experiment



Grieben et al. Scene memory and spatial inhibition in visual search. Atten Percept Psychophys (2020)

Experiment - Results



Grieben et al. Scene memory and spatial inhibition in visual search. Atten Percept Psychophys (2020)

Model

- The **model** captures **three** fundamental **processes** of **visual cognition**:

Model

- The model captures three fundamental processes of visual cognition:
  - **Exploring** the visual array through **sequences** of attentional **selection decisions**, and at each attended location, **committing** the perceived feature values to scene **memory**.

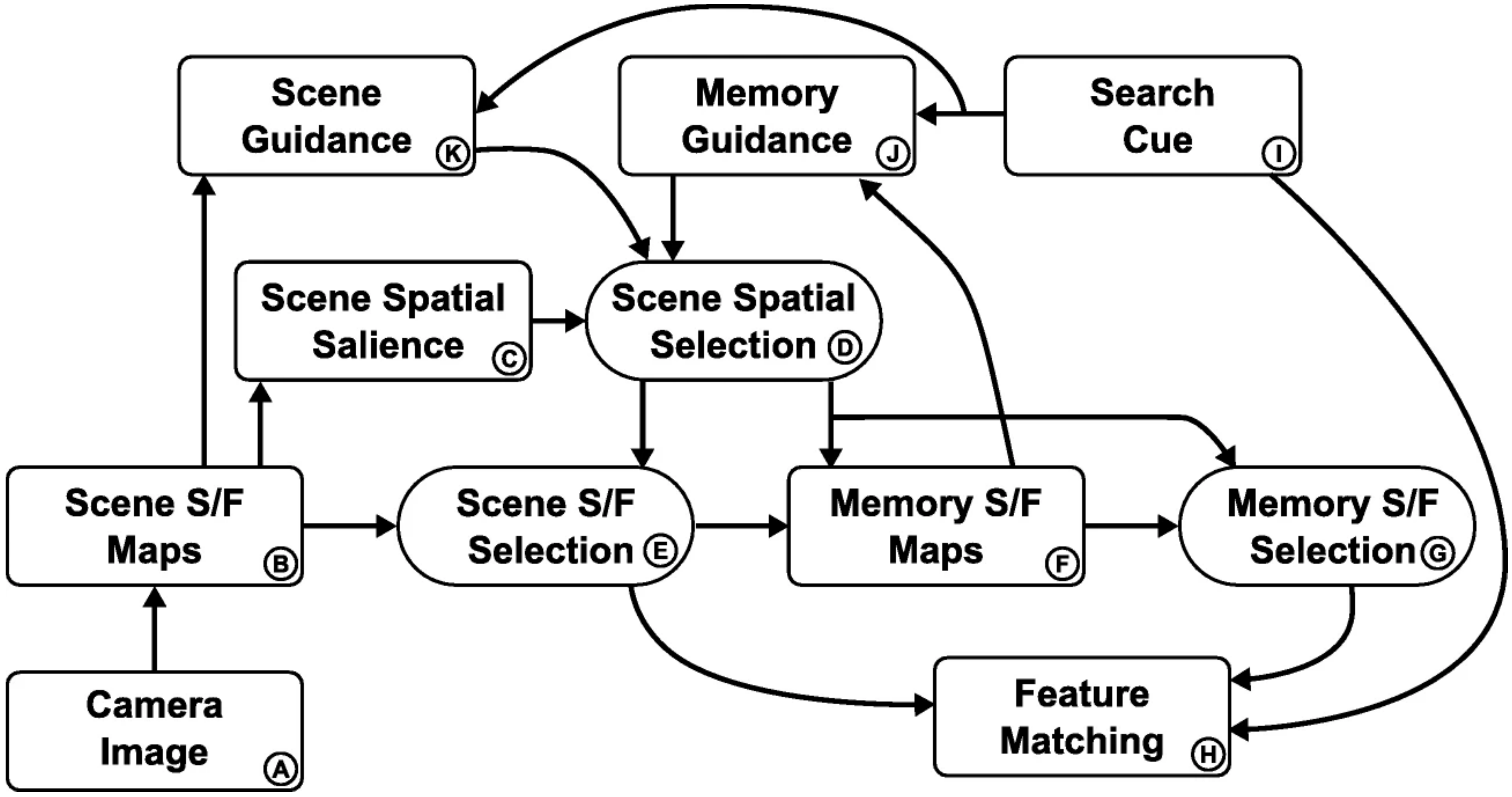
Model

- The model captures three fundamental processes of visual cognition:
  - Exploring the visual array through sequences of attentional selection decisions, and at each attended location, committing the perceived feature values to scene memory.
  - **Shifting attention** to **locations** at which **visual transients** are **detected** and **committing feature** information from those locations to a working **memory** of the feature **cue** of visual search.

Model

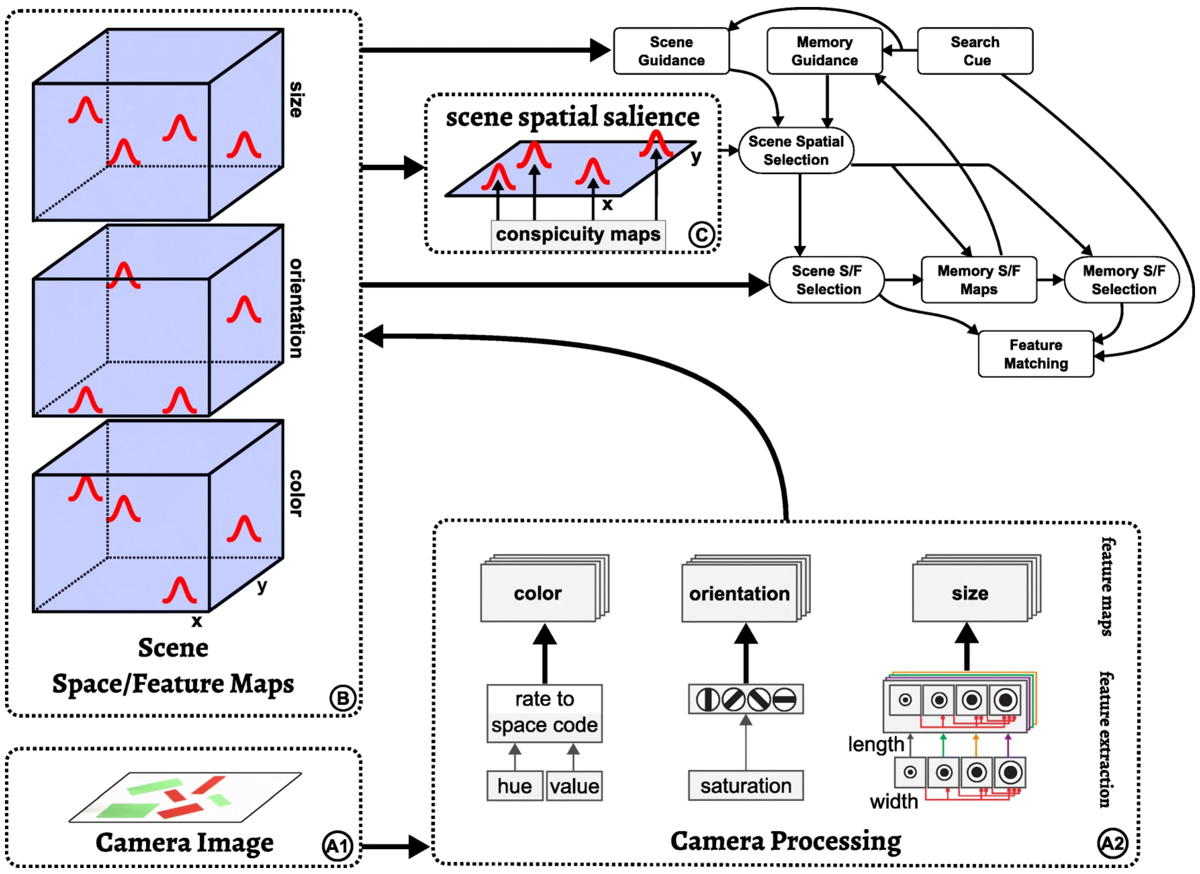
- The model captures three fundamental processes of visual cognition:
  - Exploring the visual array through sequences of attentional selection decisions, and at each attended location, committing the perceived feature values to scene memory.
  - Shifting attention to locations at which visual transients are detected and committing feature information from those locations to a working memory of the feature cue of visual search.
  - **Visually searching** for **locations** in the visual **array** at which the **cued** feature conjunctions are seen.

Model



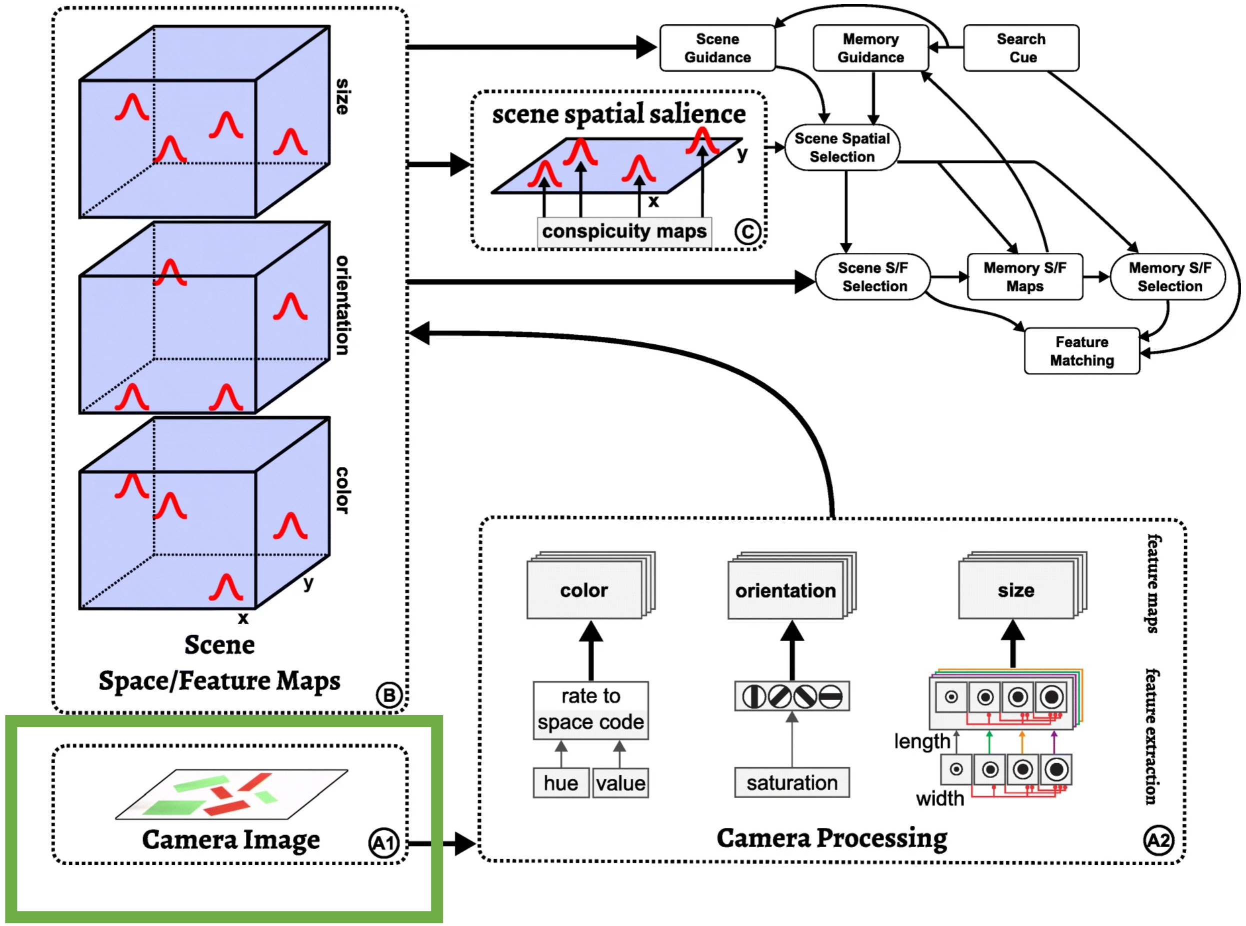
Grieben et al. Scene memory and spatial inhibition in visual search. Atten Percept Psychophys (2020)

Subsystem 1: Feed-forward feature and salience maps



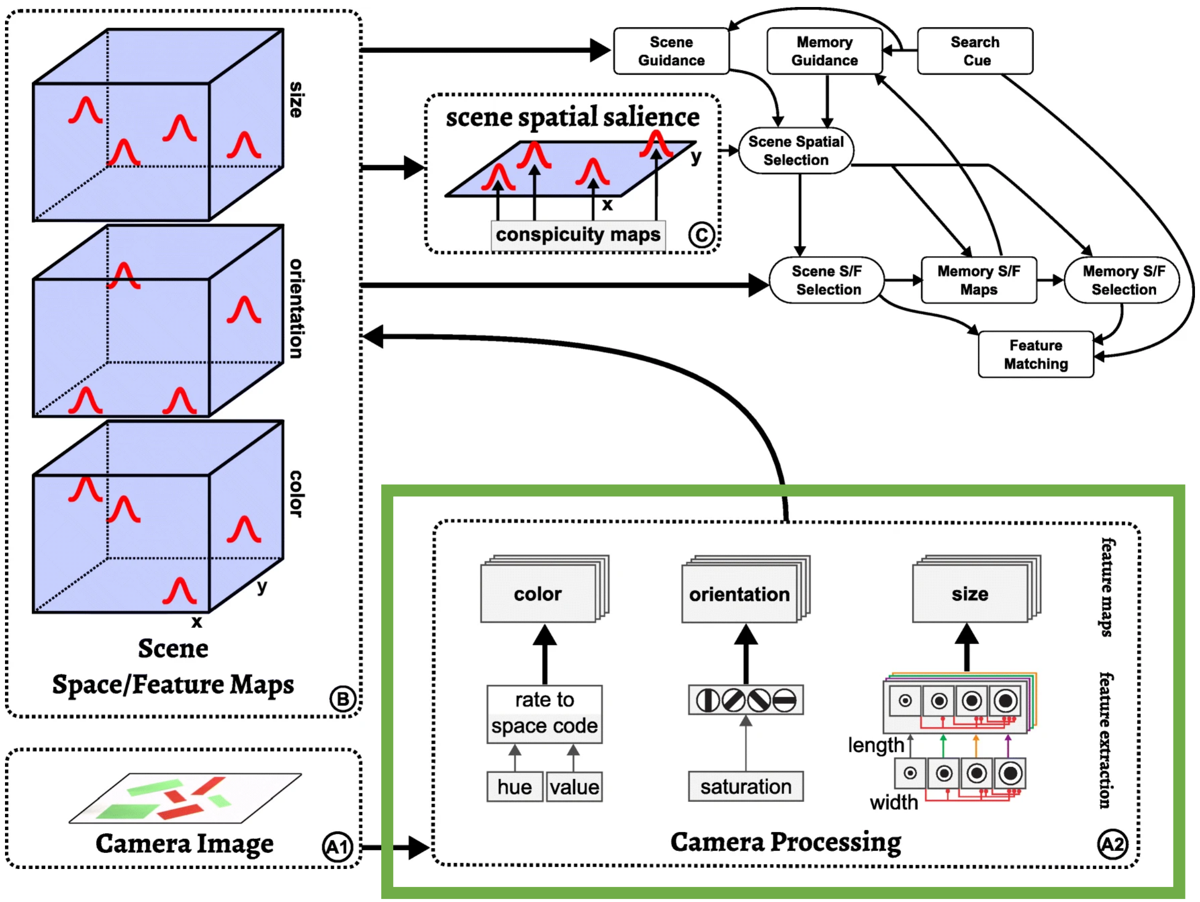
- **Visual cognition** builds on visual **input** from which **features** are **extracted**.

Subsystem 1: Feed-forward feature and salience maps



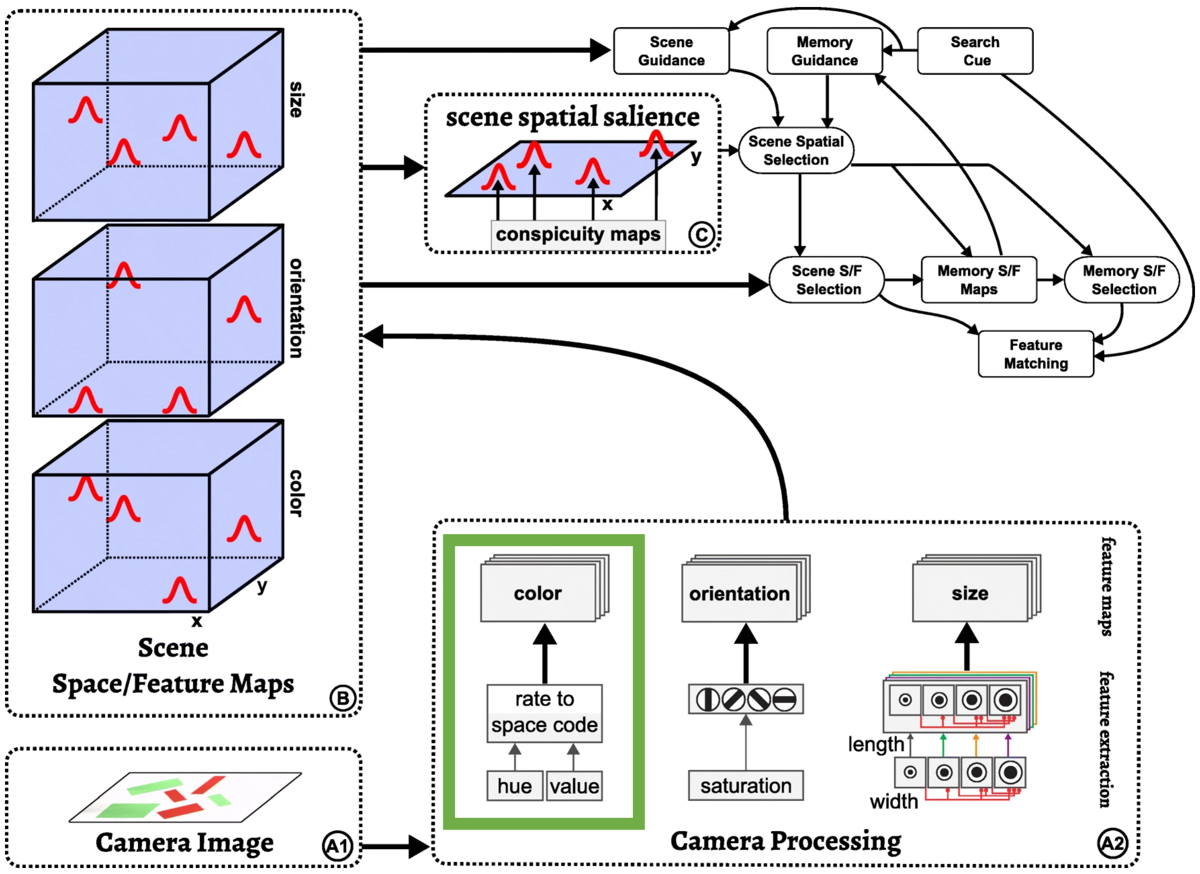
- Visual cognition builds on visual input from which features are extracted.
- **Visual input** may take the form of a video stream from live **camera** input or from **sequences** of synthetic **images**.

Subsystem 1: Feed-forward feature and salience maps



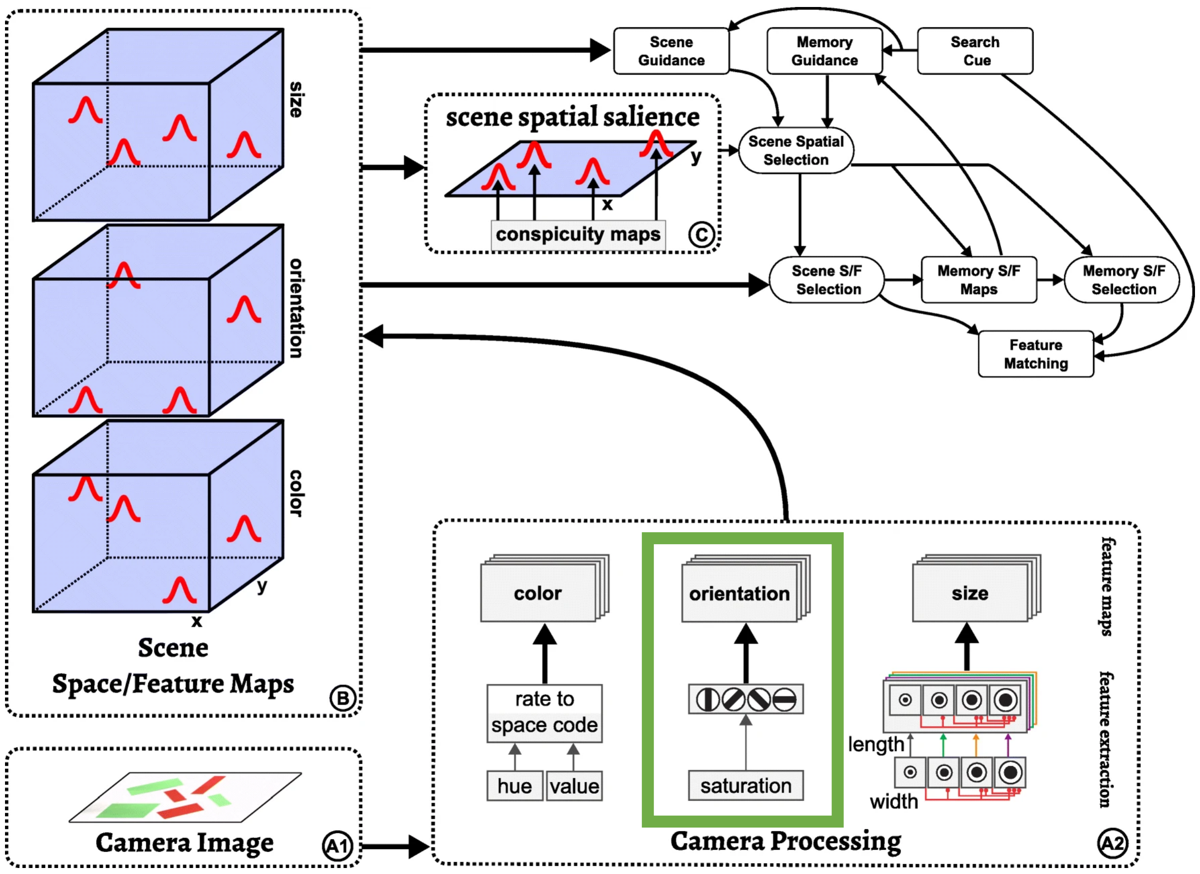
- Visual cognition builds on visual input from which features are extracted.
- Visual input may take the form of a video stream from live camera input or from sequences of synthetic images.
- Three simple **features** are used in the model: **color**, **orientation**, and **size**.

Subsystem 1: Feed-forward feature and salience maps



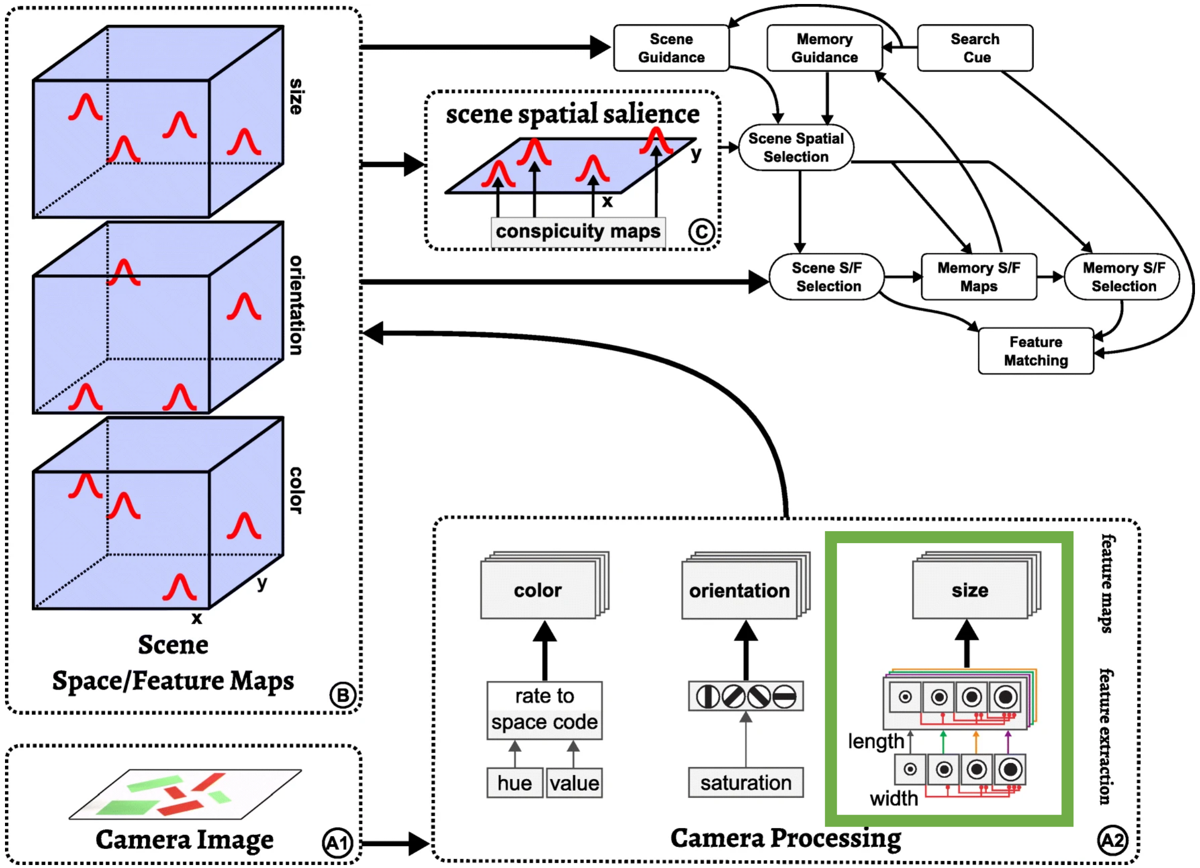
- **Color** is extracted by transforming RGB values into **hue-space**.

Subsystem 1: Feed-forward feature and salience maps



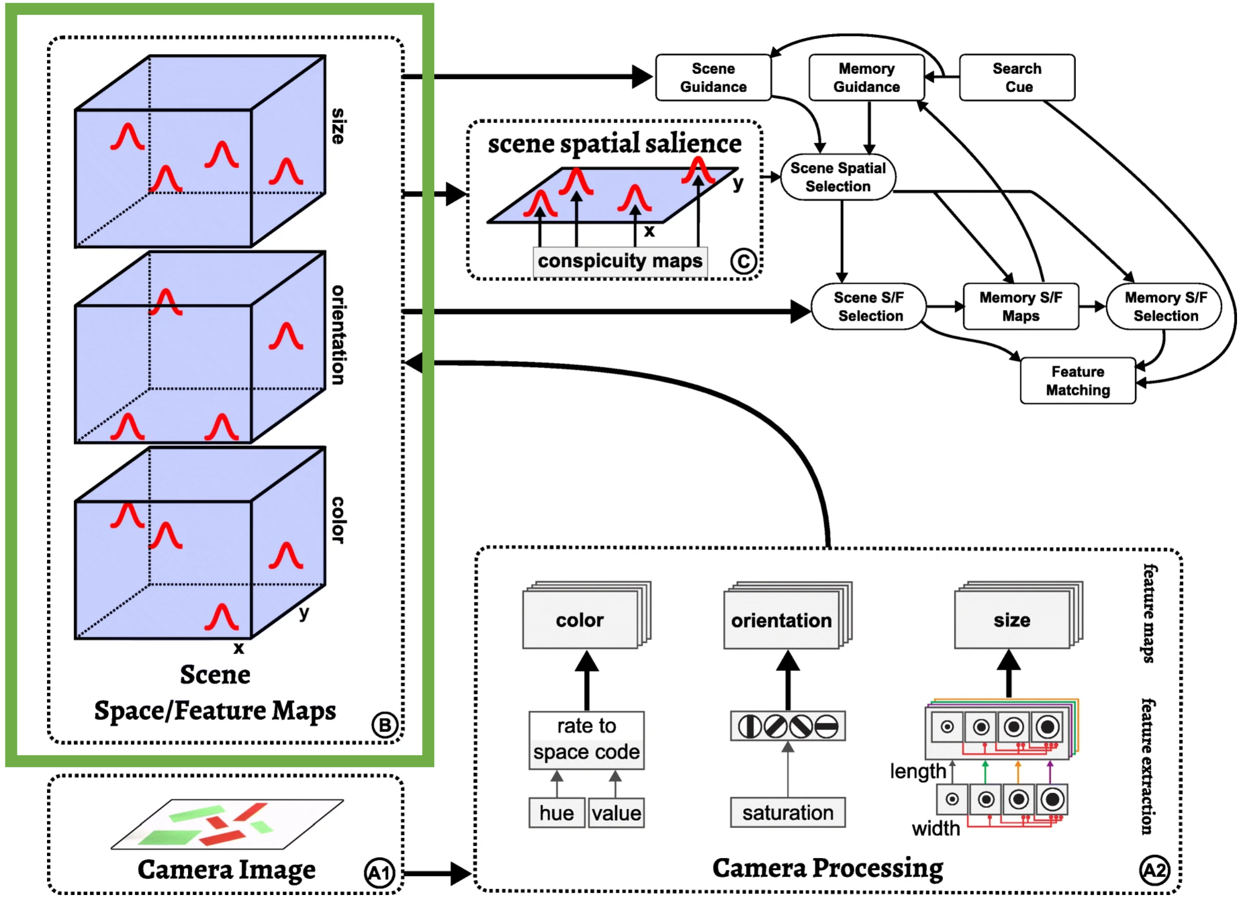
- Color is extracted by transforming RGB values into hue-space.
- **Saturation** is passed through a threshold function and four **elongated center-surround filters** to extract **orientation**.

Subsystem 1: Feed-forward feature and salience maps



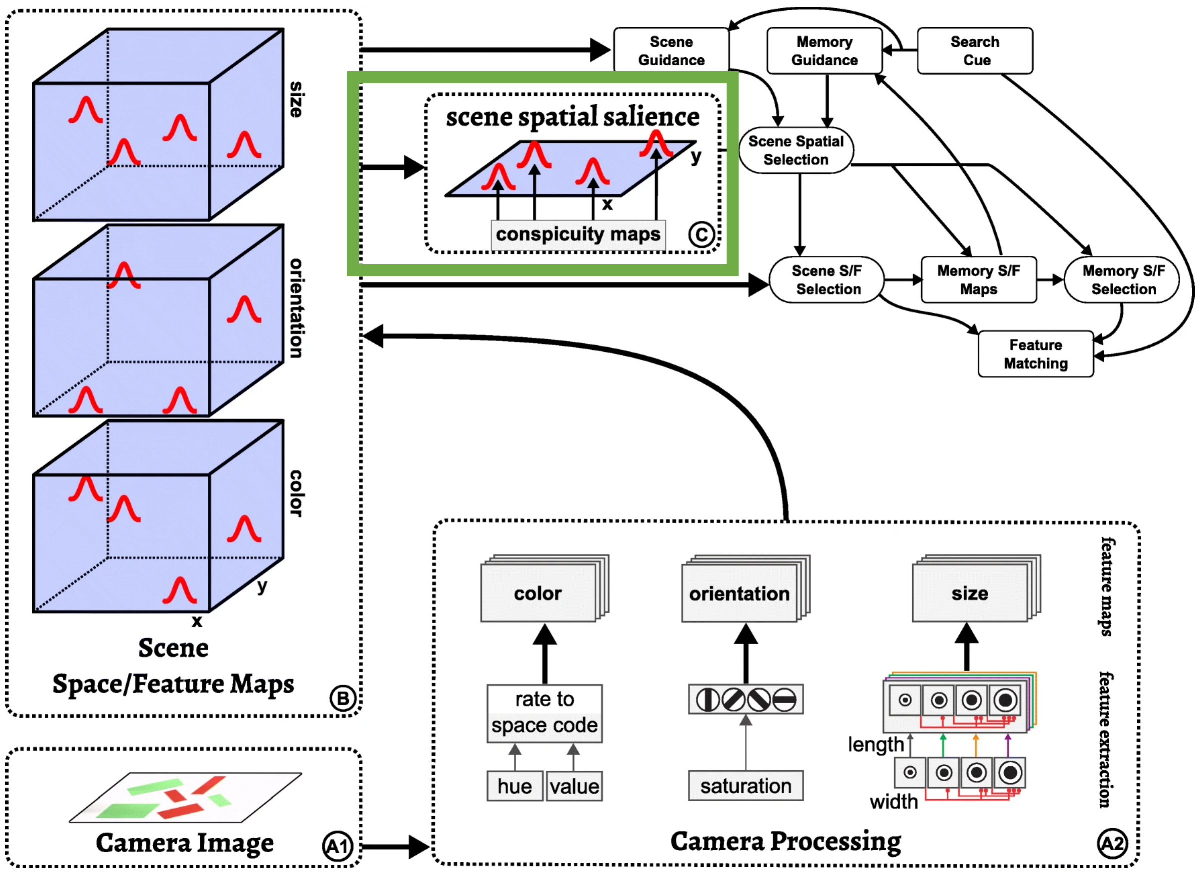
- Color is extracted by transforming RGB values into hue-space.
- Saturation is passed through a threshold function and four elongated center-surround filters to extract orientation.
- **Size** is extracted using a **pyramid of center-surround filters** of increasing size with a **one-way inhibition** along the scale dimension.

Subsystem 1: Feed-forward feature and salience maps



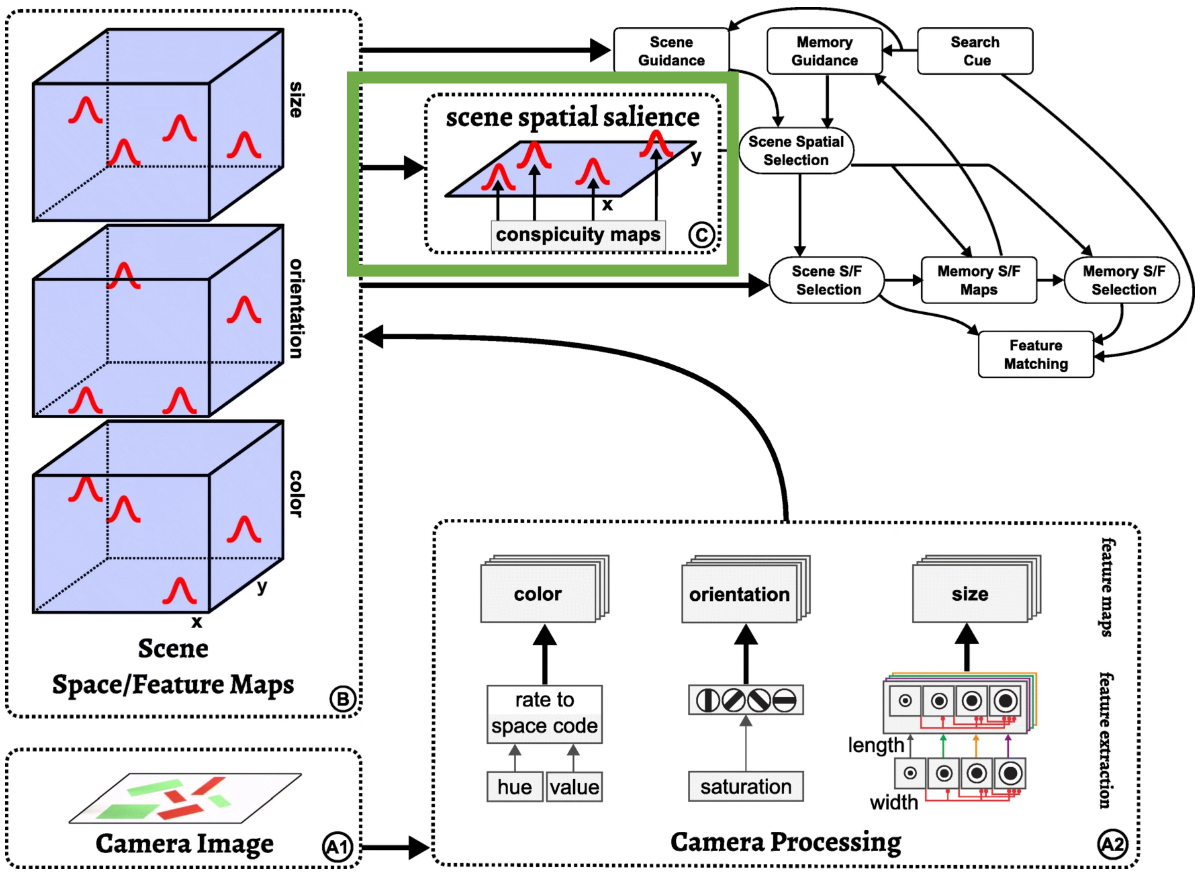
- The normalized **output** of the **feature** extraction **pathway** provides **input** into three **space/feature fields**, which each combine **two** dimensions of visual **space** with **one feature dimension**.

Subsystem 1: Feed-forward feature and salience maps



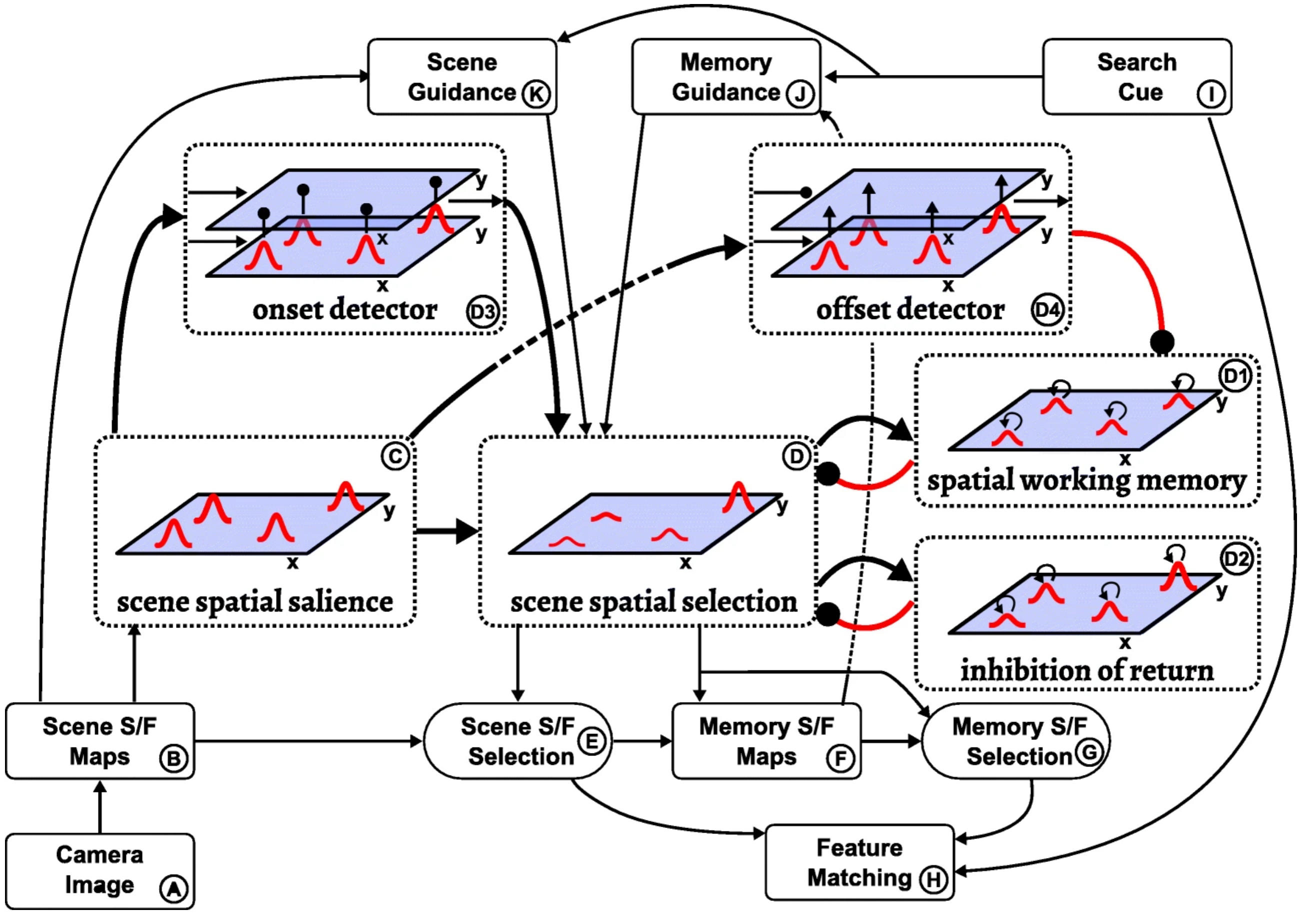
- **Each** of the three scene space/feature maps **projects** to the **scene spatial salience field**.

Subsystem 1: Feed-forward feature and salience maps



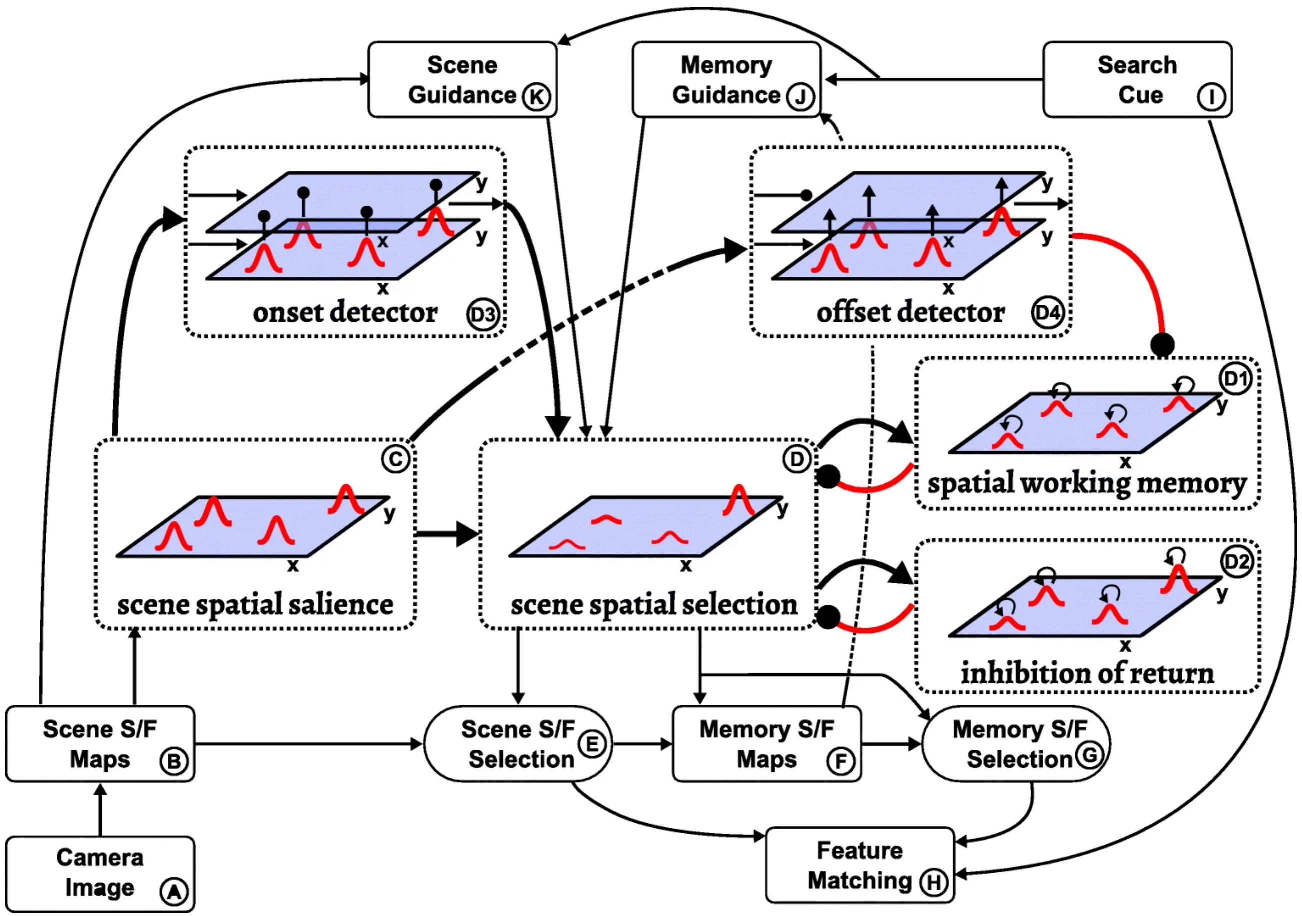
- Each of the three scene space/feature maps projects to the scene spatial salience field.
- **These projections** marginalize the feature dimension, so they **are purely spatial**.

Subsystem 2: Attentional selection



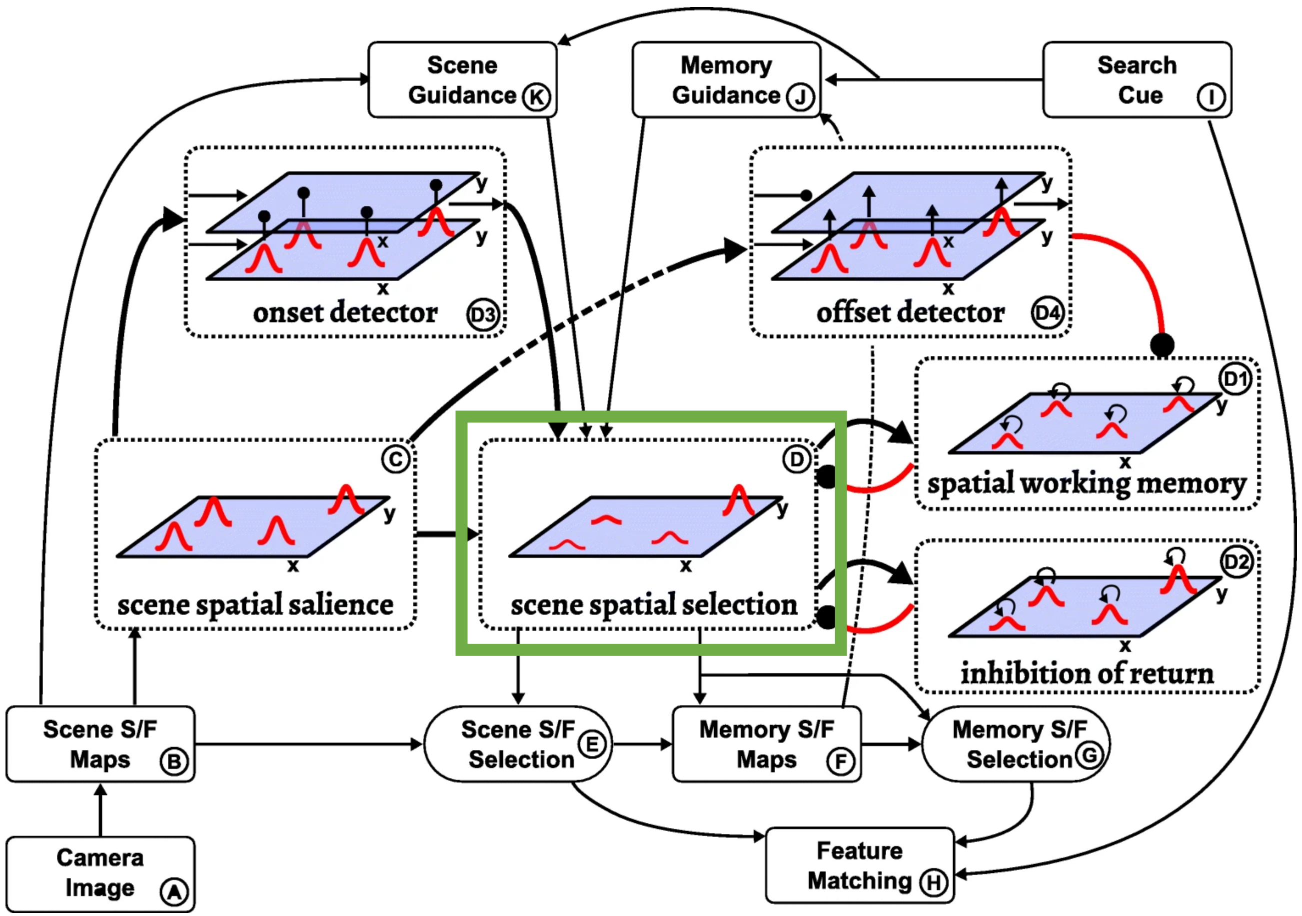
- **Visual cognition** always **entails attentional selection decisions**.

Subsystem 2: Attentional selection



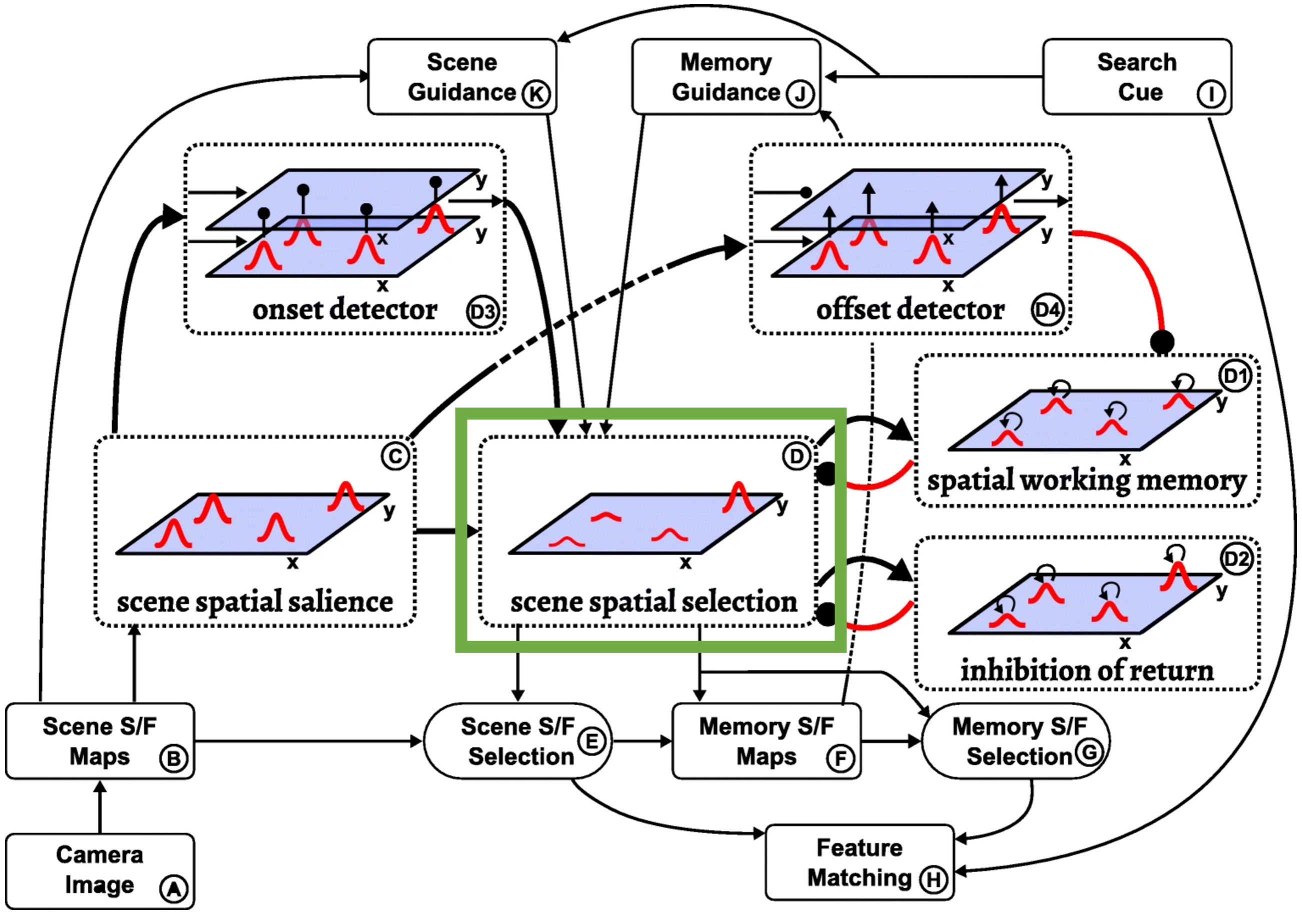
- Visual cognition always entails attentional selection decisions.
- This is the **sub-system** of the neural dynamic architecture **that generates such selection decisions**.

Subsystem 2: Attentional selection



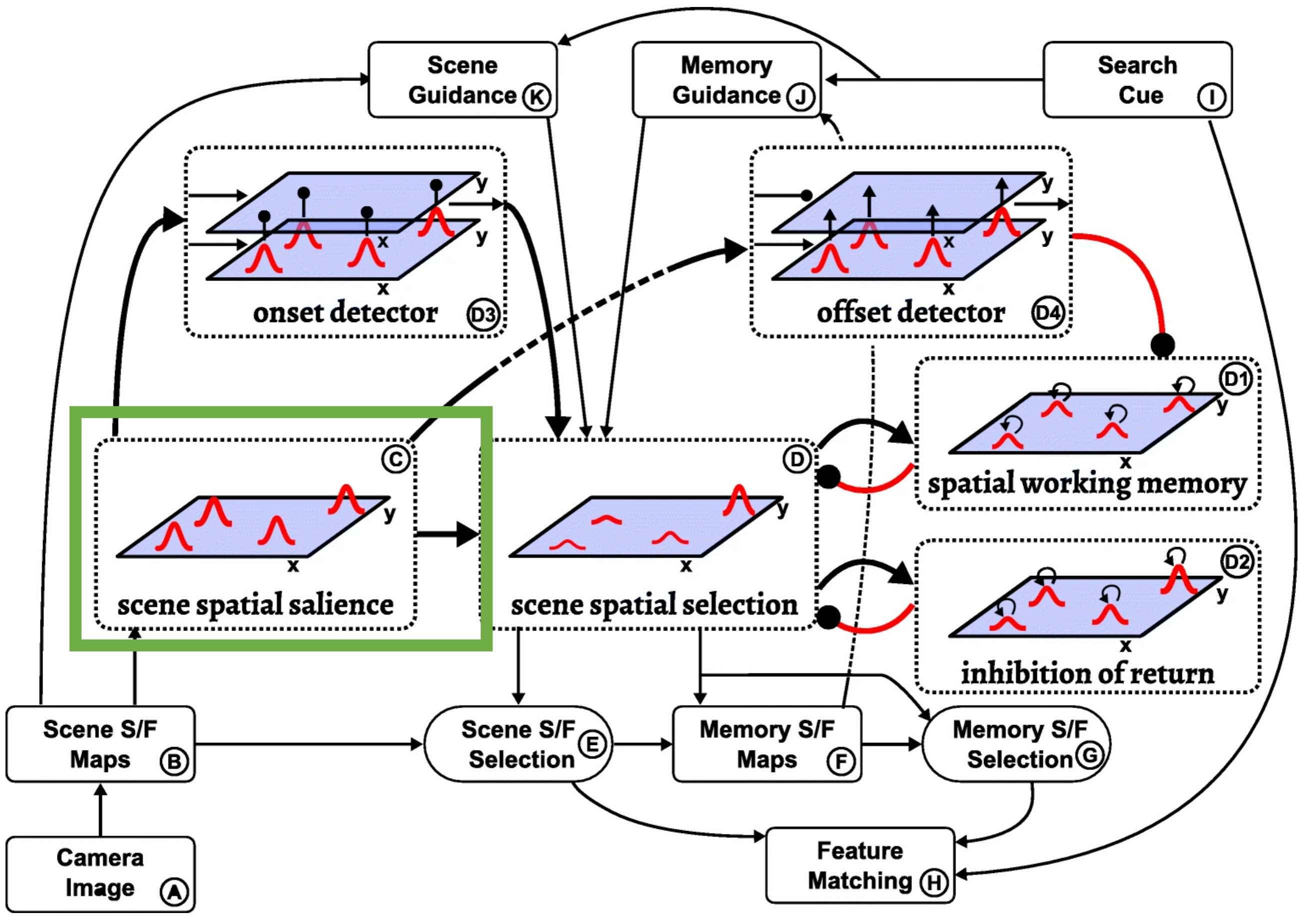
- Visual cognition always entails attentional selection decisions.
- This is the sub-system of the neural dynamic architecture that generates such selection decisions.
- **Central** is the **scene spatial selection field**, which **represents** the **current** location of spatial **attention**.

Subsystem 2: Attentional selection



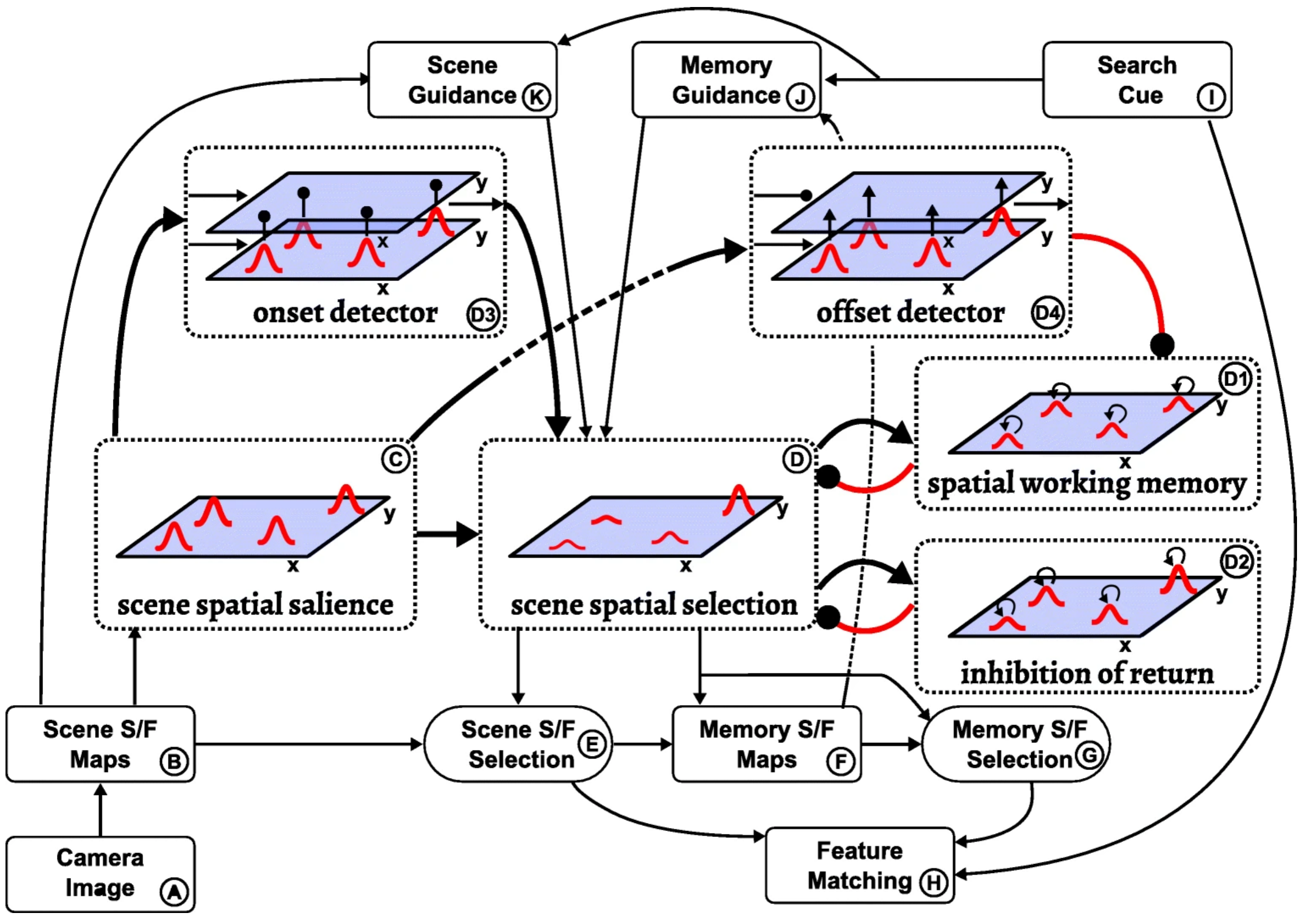
- **This field** is in the **dynamic regime** of **selection** so that it can support only a **single** supra-threshold **peak** at any point in time.

Subsystem 2: Attentional selection



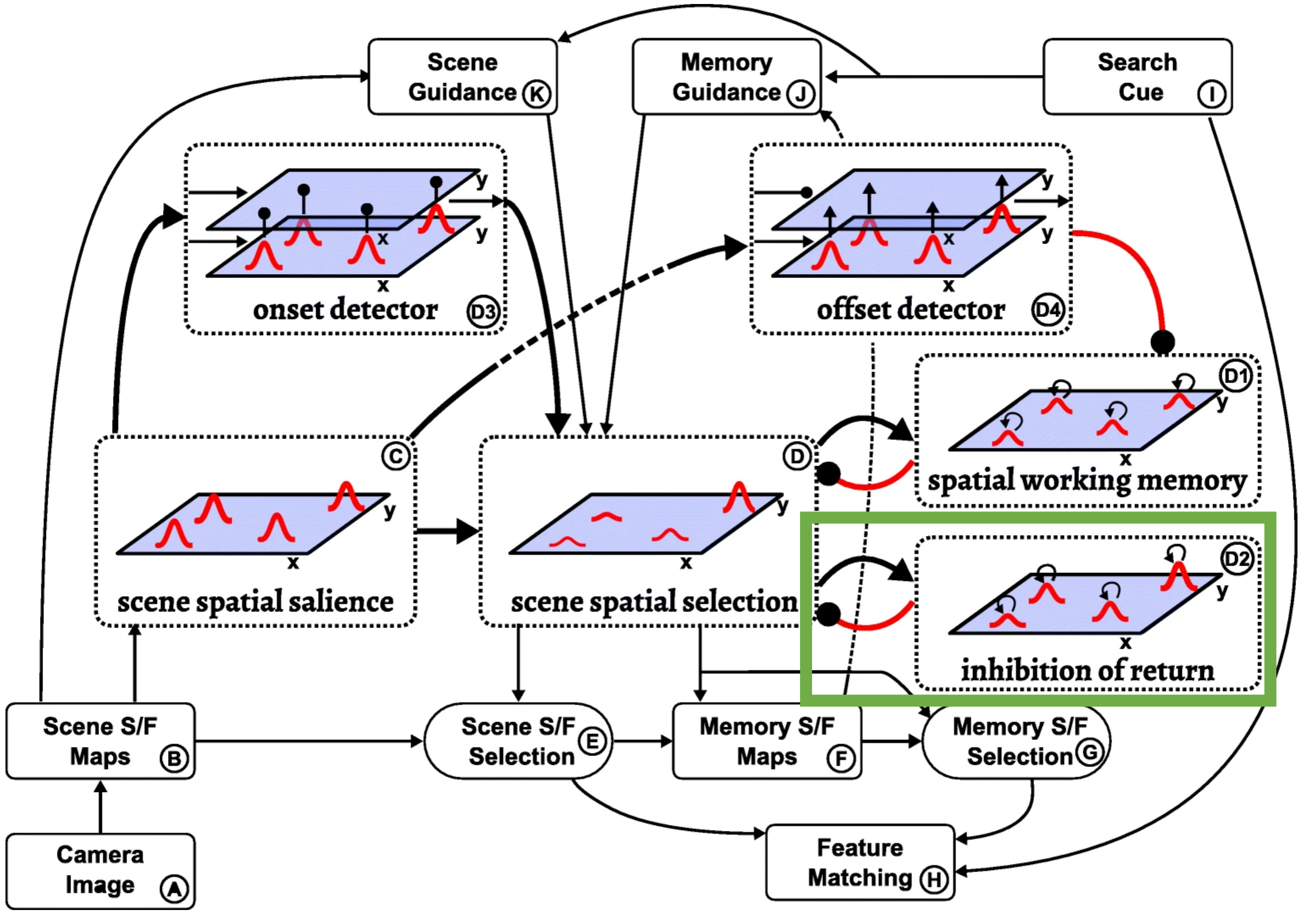
- This field is in the dynamic regime of selection so that it can support only a single supra-threshold peak at any point in time.
- It **receives** multi-modal **input** from the **salience field** and **selects** the **most salient location**.

Subsystem 2: Attentional selection



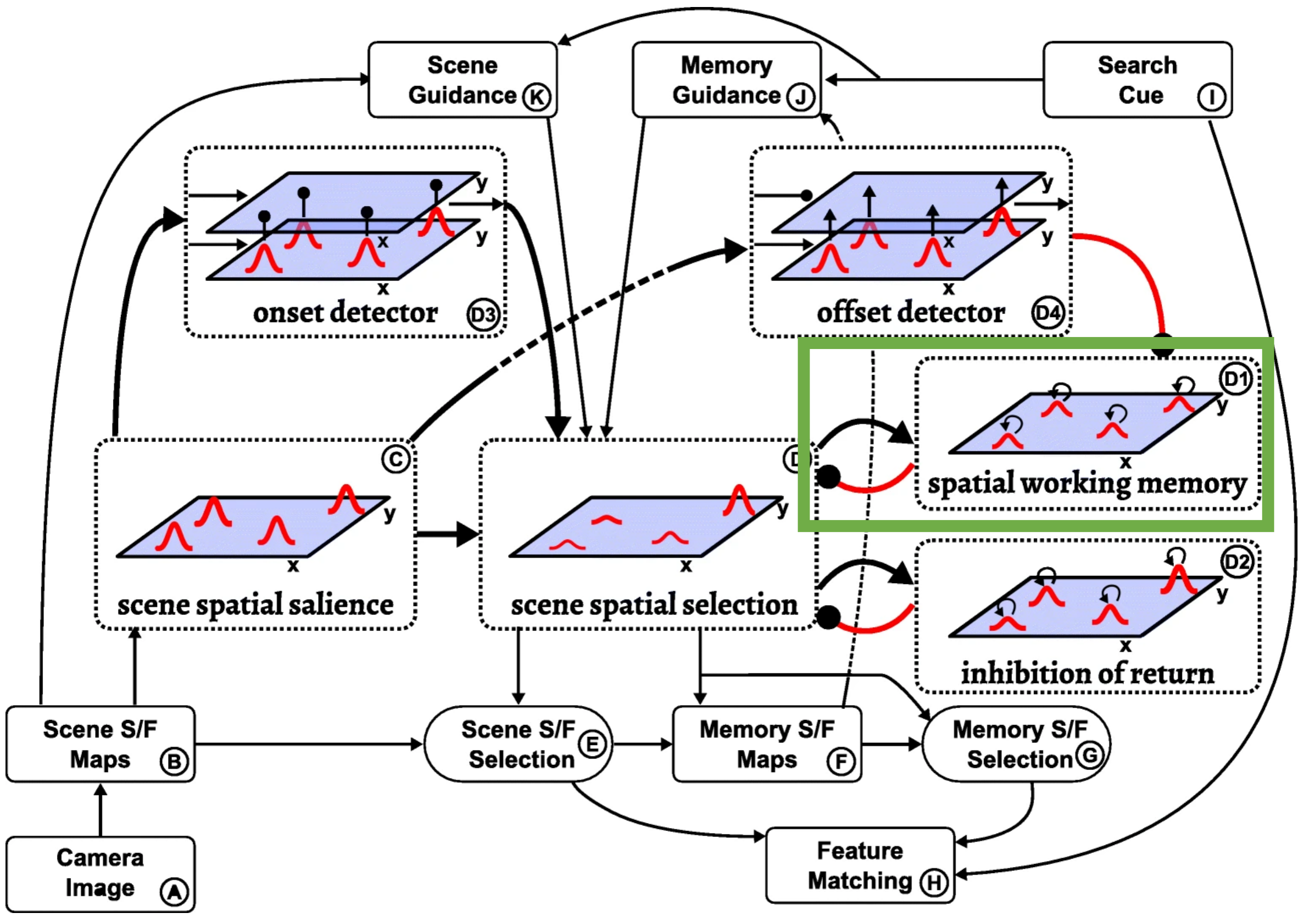
- This field is in the dynamic regime of selection so that it can support only a single supra-threshold peak at any point in time.
- It receives multi-modal input from the salience field and selects the most salient location.
- That **selection** is **biased** by **three additional** sources of **input**.

Subsystem 2: Attentional selection



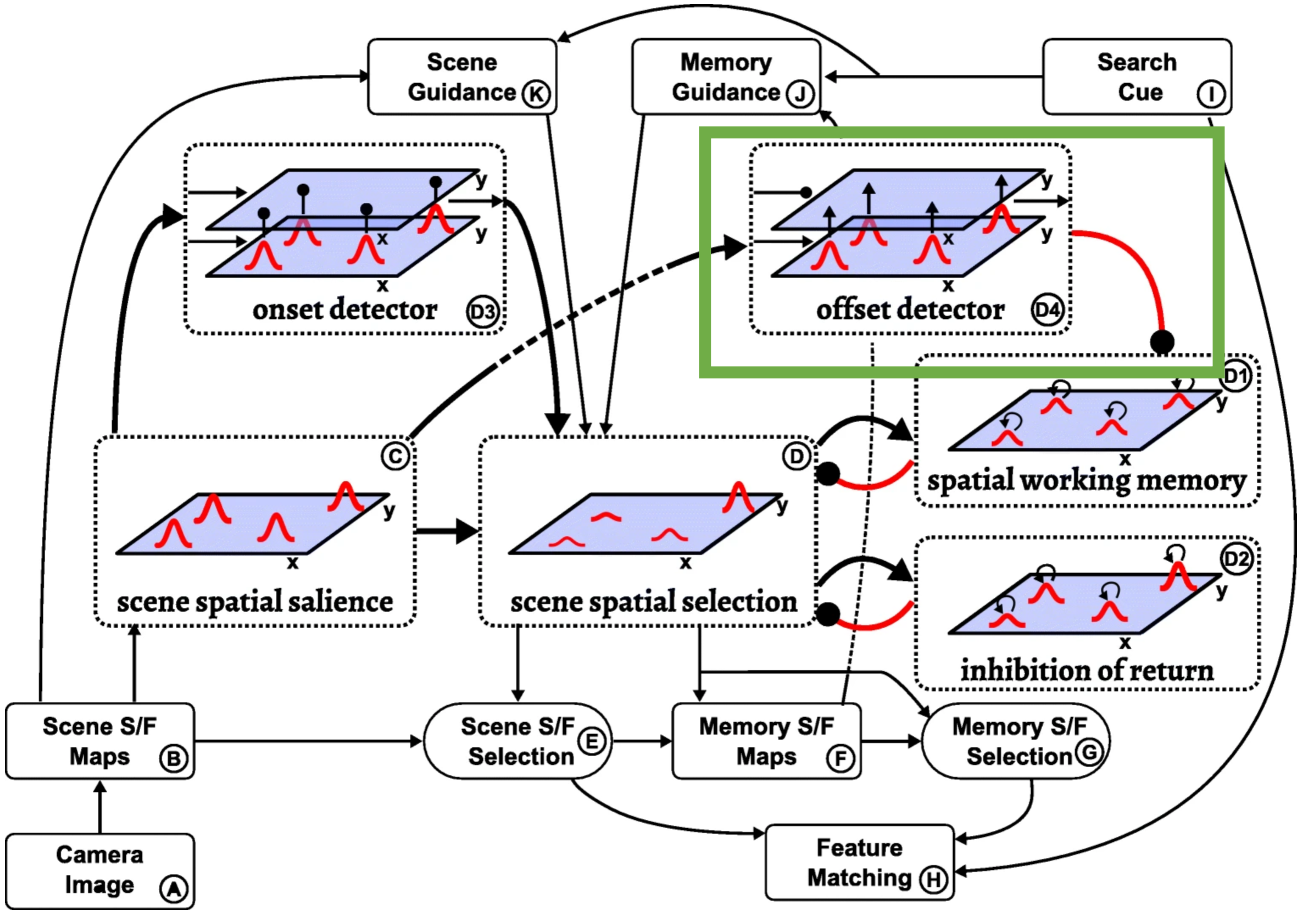
- **First**, it is **biased away** from **previously attended locations** by inhibitory input from the **inhibition of return memory trace** that reflects the recent history of activation of the scene spatial selection field.

Subsystem 2: Attentional selection



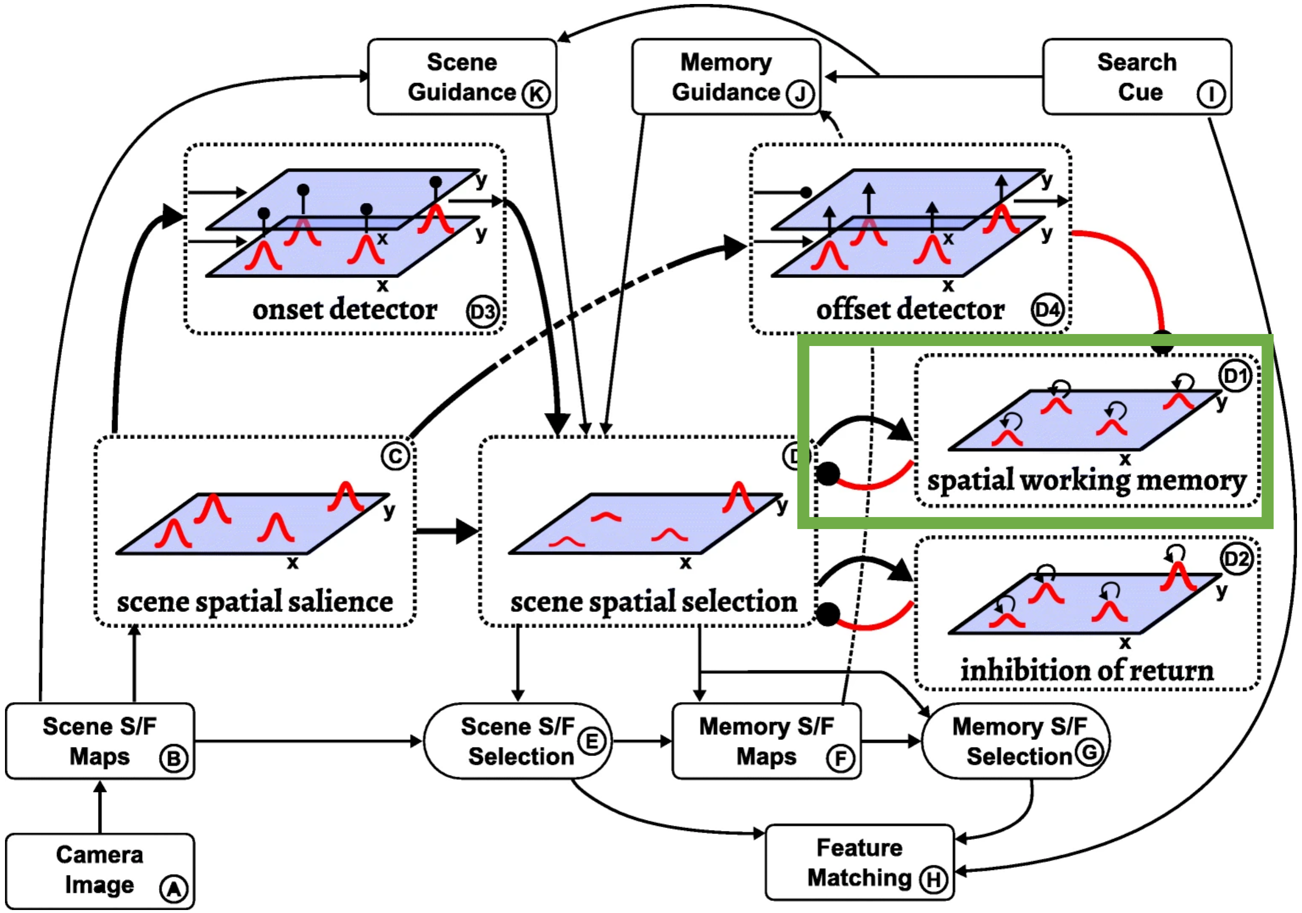
- **Second**, it is **biased away** from locations that receive **inhibitory input** from **the spatial working memory** field.

Subsystem 2: Attentional selection



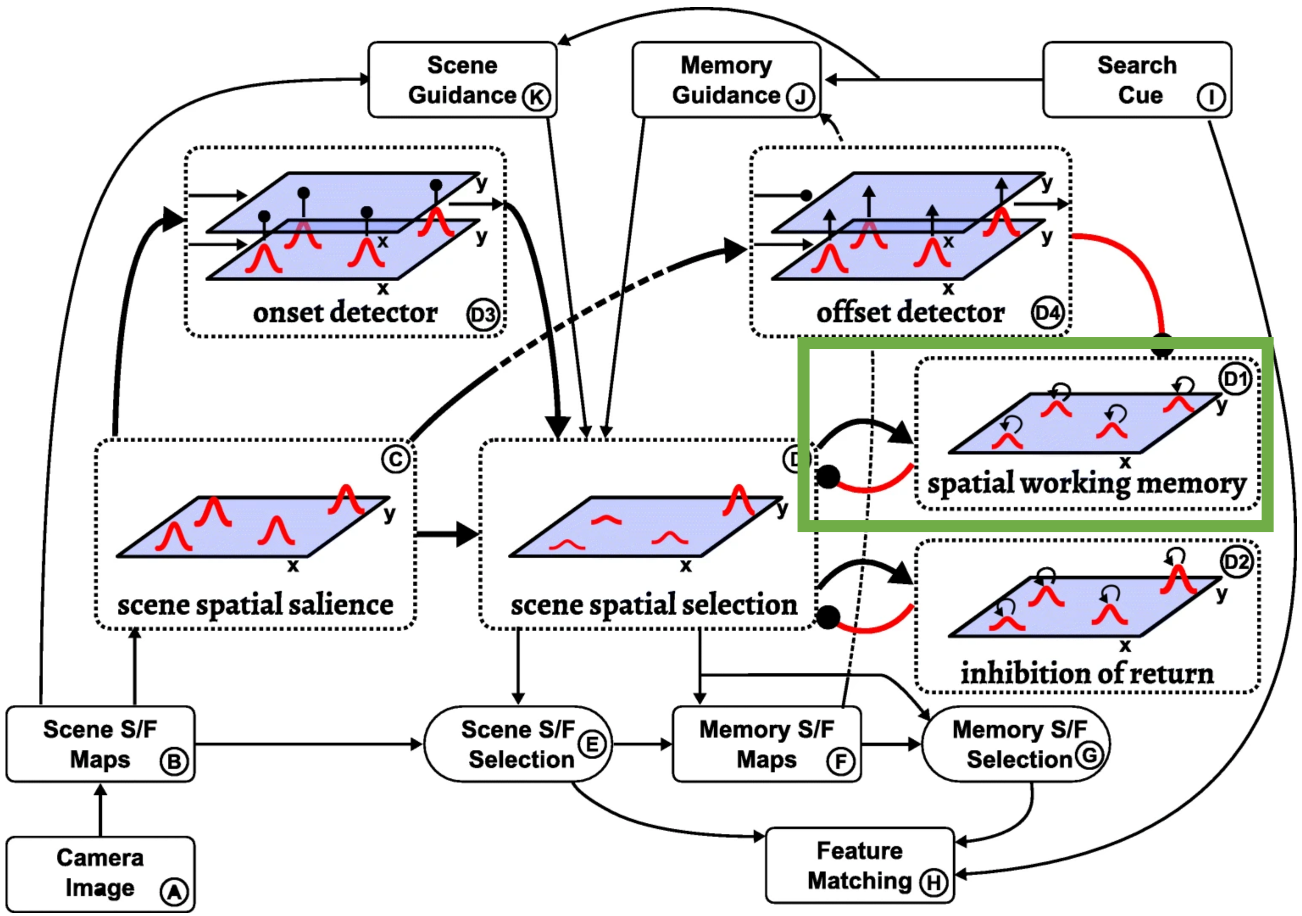
- Sustained **peaks** in that **field** are **destabilized**, however, whenever **movement** is **detected** in the **scene**. This happens **through** a **two-layer offset detector** that generates a transient activation peak whenever salience input moves or vanishes.

Subsystem 2: Attentional selection



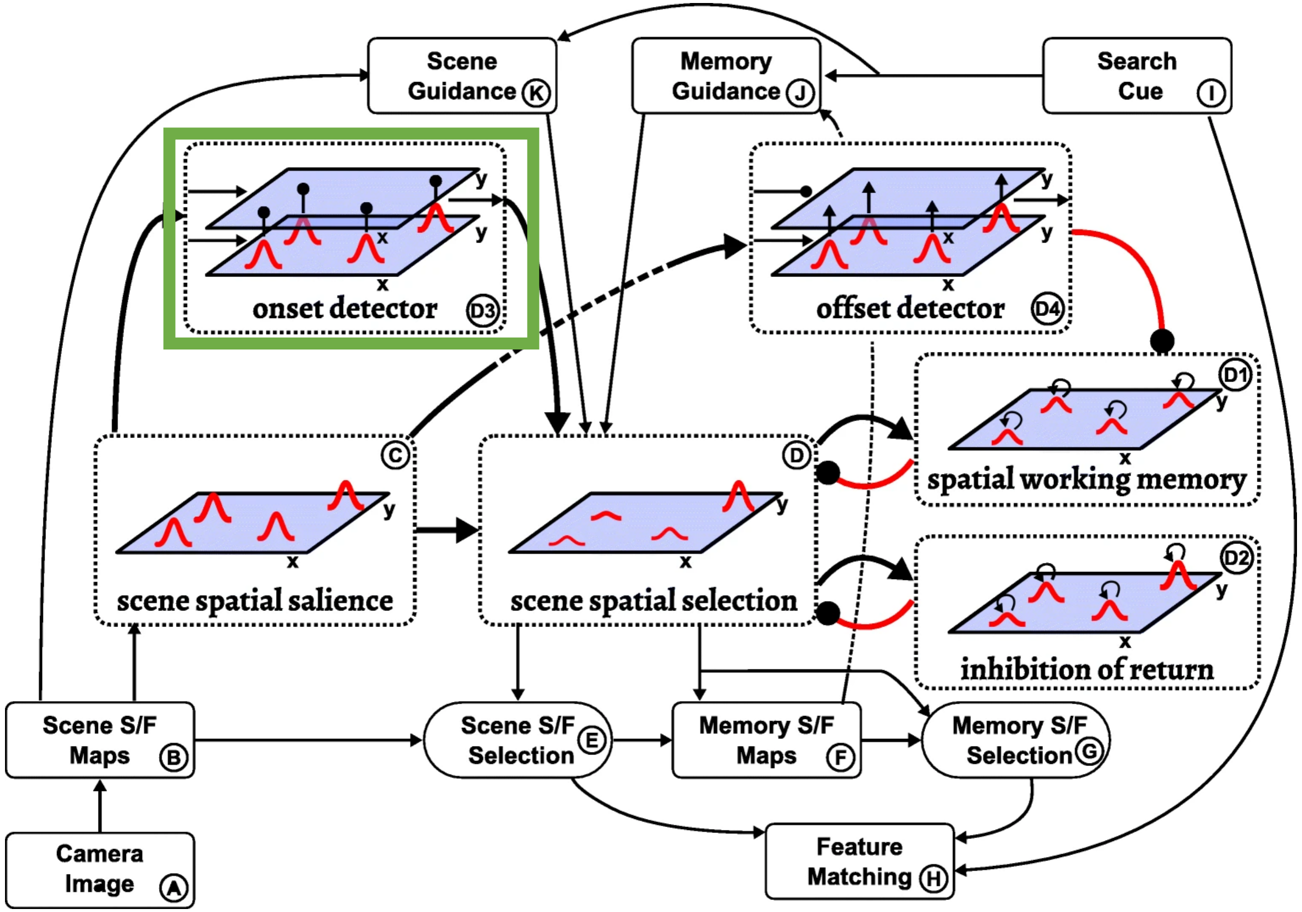
- The **number of peaks** that can be **simultaneously sustained** in the spatial **working memory** field is **limited** by accumulating inhibition from these peaks.

Subsystem 2: Attentional selection



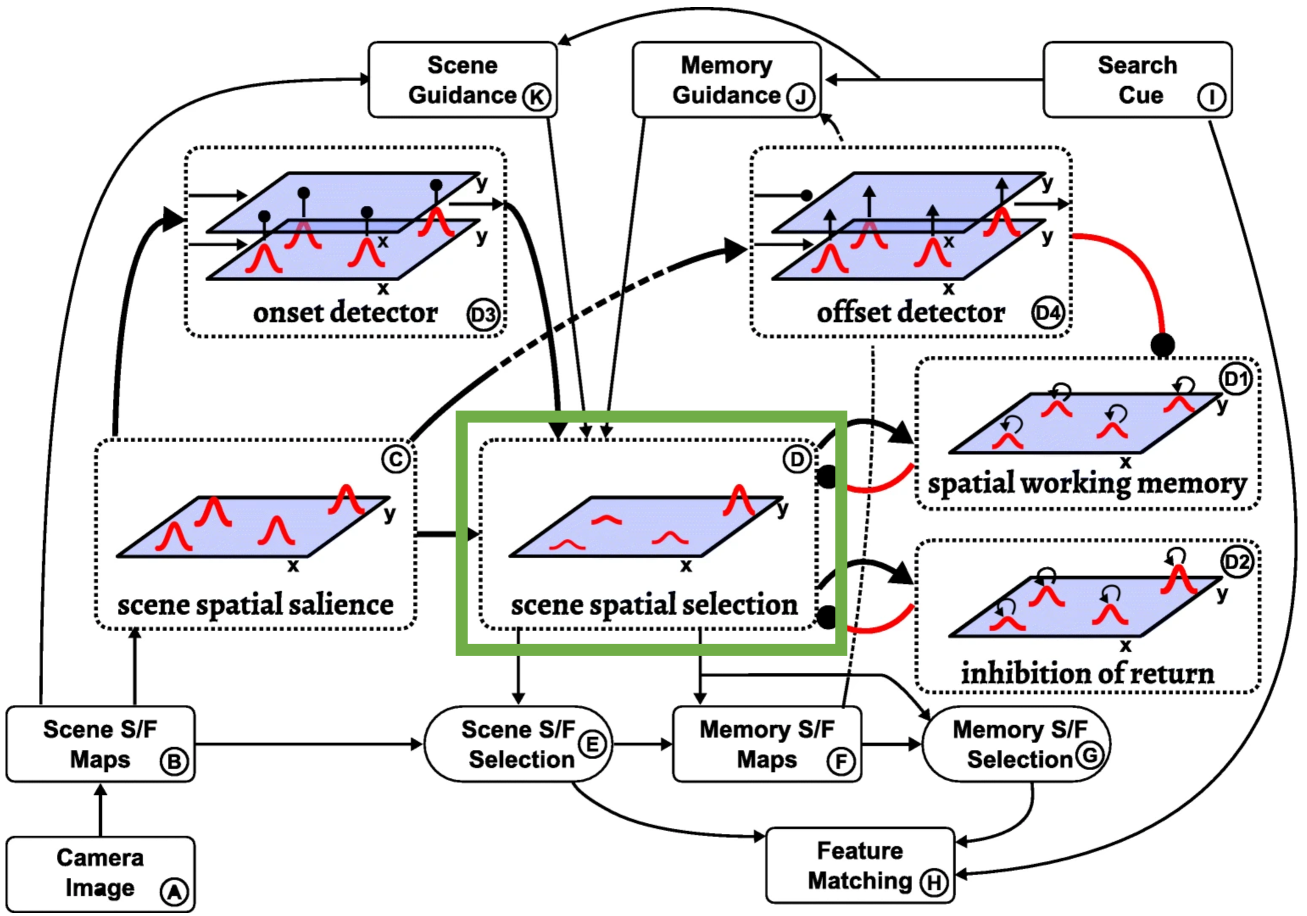
- The exact number, that reflects the **capacity of working memory**, **depends** on the **balance** of neural **inhibition** and **excitation** in this field and provides an important **constraint** for **fitting** the **experimental** results.

Subsystem 2: Attentional selection



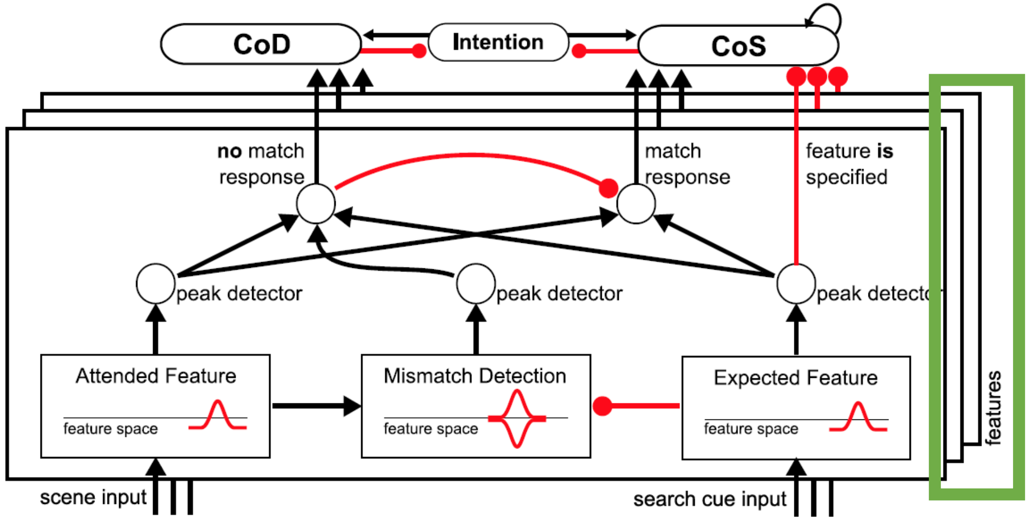
- **Third**, **attention** is **attracted** to **locations** at which **rapid changes** of spatial salience **occur**. This bias arises due to **input** from **an onset detector**, a two-layer neural dynamic field that generates a transient activation peak in response to shifts of input.

Subsystem 2: Attentional selection



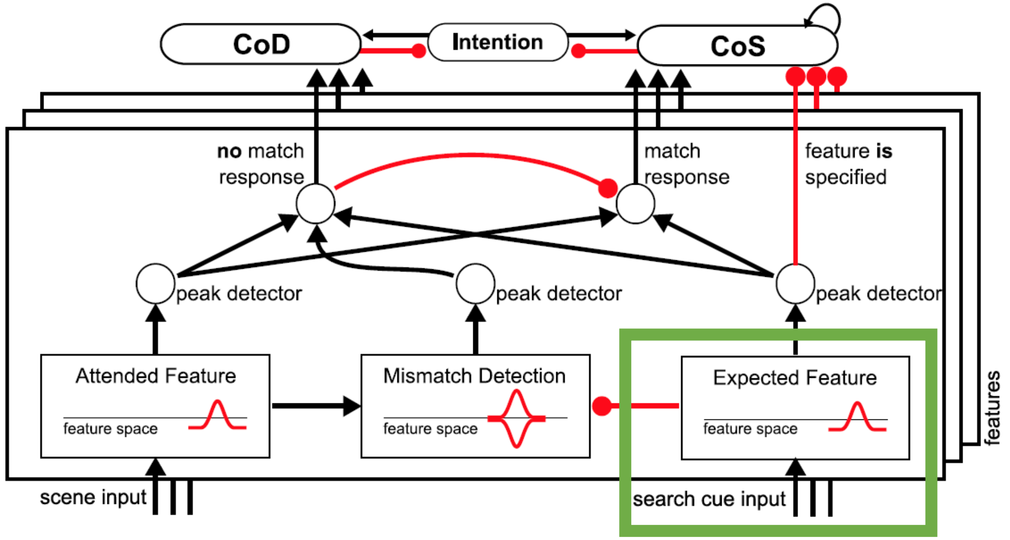
- **Spatial attention**, **represented** by a **self-stabilized peak** in the **scene spatial selection field**, plays a critical roll in **feature binding**. Feature binding occurs in the model in a manner that could be viewed as a **neural implementation** of **Treisman's feature integration theory**.

Subsystem 3: Feature matching



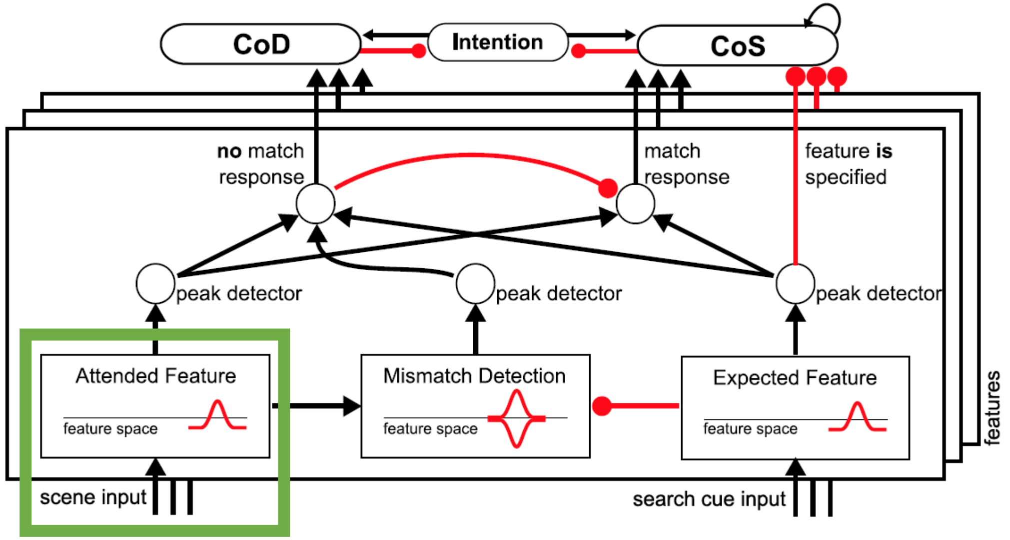
The **feature matching sub-network compares** (in **parallel** across feature dimensions)

Subsystem 3: Feature matching



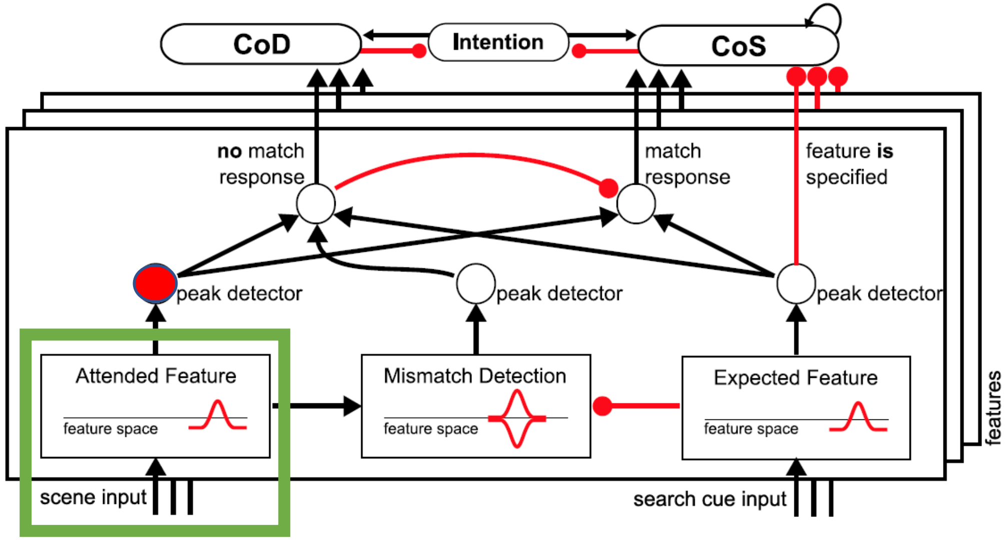
The feature matching sub-network compares (in parallel across feature dimensions) the **expected feature** (search cue)

Subsystem 3: Feature matching



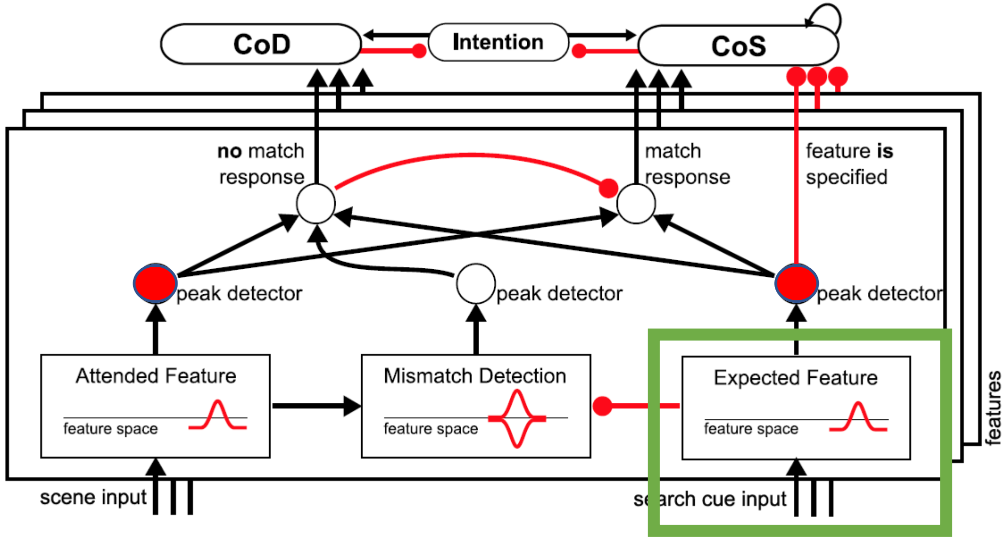
The feature matching sub-network compares (in parallel across feature dimensions) the expected feature (search cue) and **attended feature** at the **attended location**

Subsystem 3: Feature matching



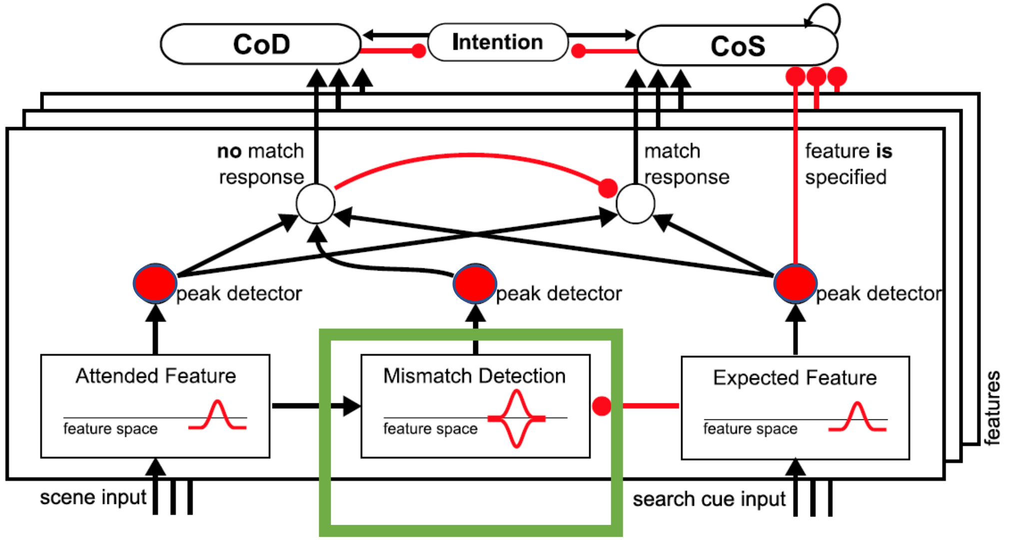
A **peak** in **all three** fields (**attended** feature

Subsystem 3: Feature matching



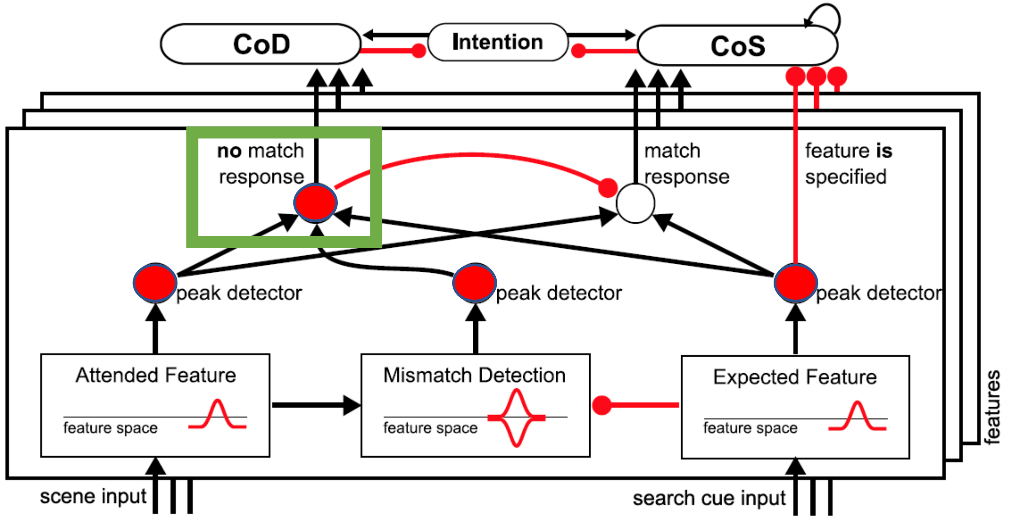
A peak in all three fields (attended feature, **expected** feature

Subsystem 3: Feature matching



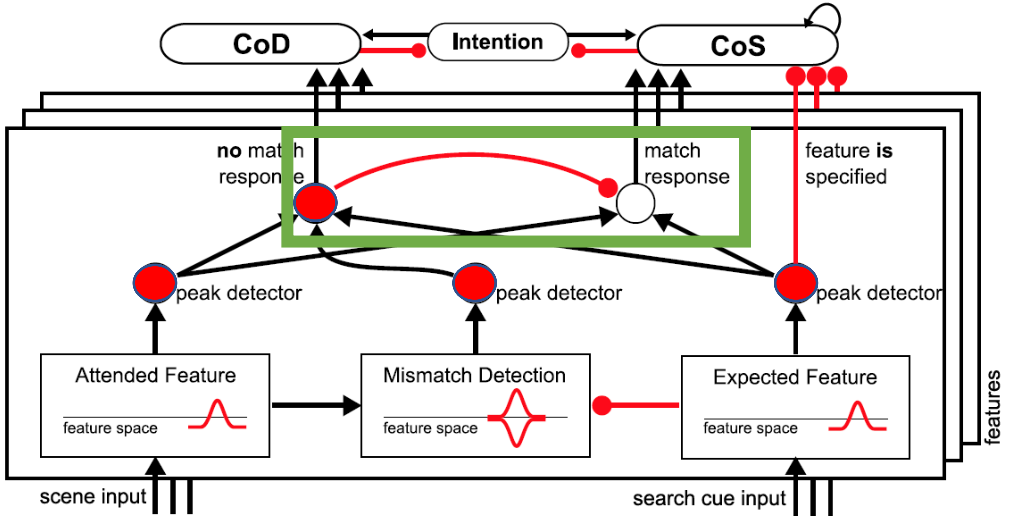
A peak in all three fields (attended feature, expected feature, and **mismatch detection**)

Subsystem 3: Feature matching



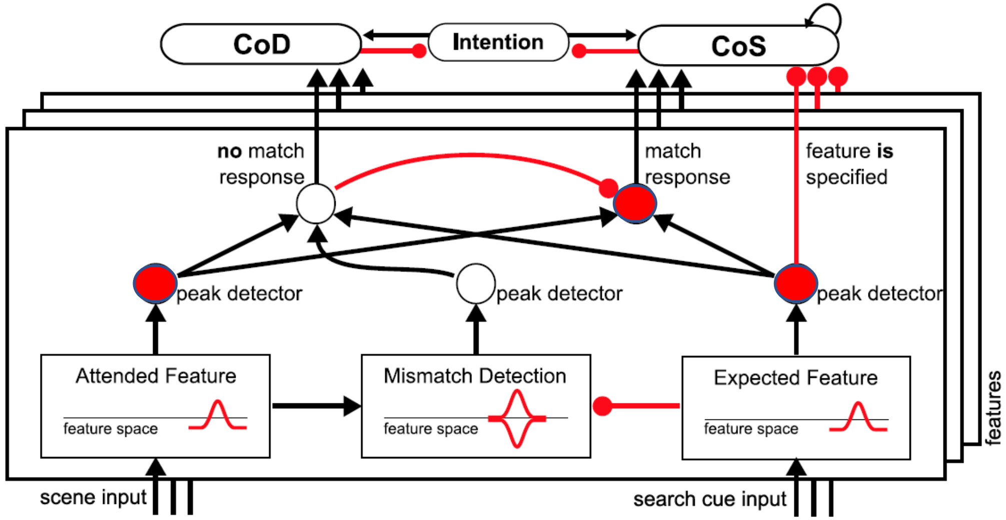
A peak in all three fields (attended feature, expected feature, and mismatch detection) **signals a no match**, activating the no-match response node

Subsystem 3: Feature matching



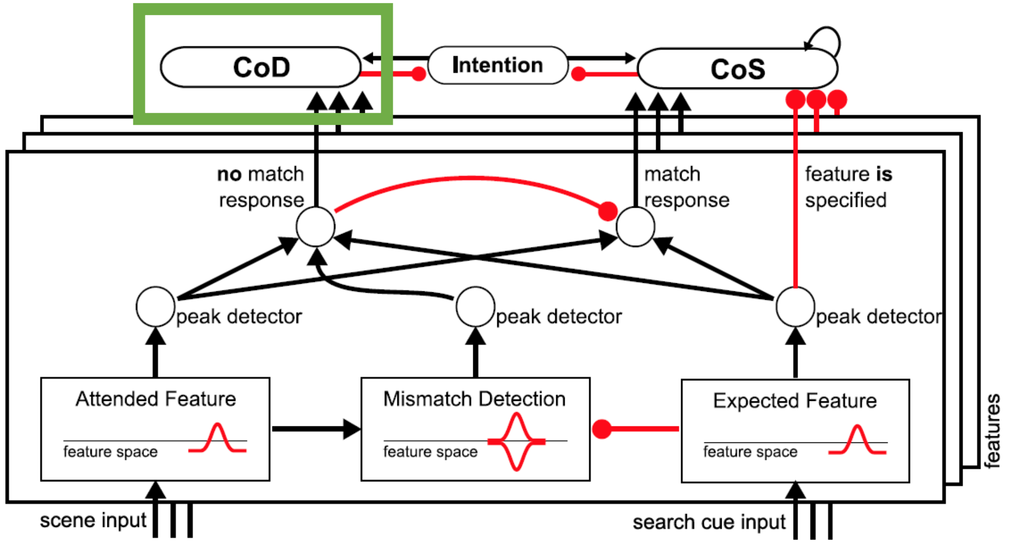
A peak in all three fields (attended feature, expected feature, and mismatch detection) signals a no match, activating the no-match response node and **inhibiting** the **match** response **node**

Subsystem 3: Feature matching



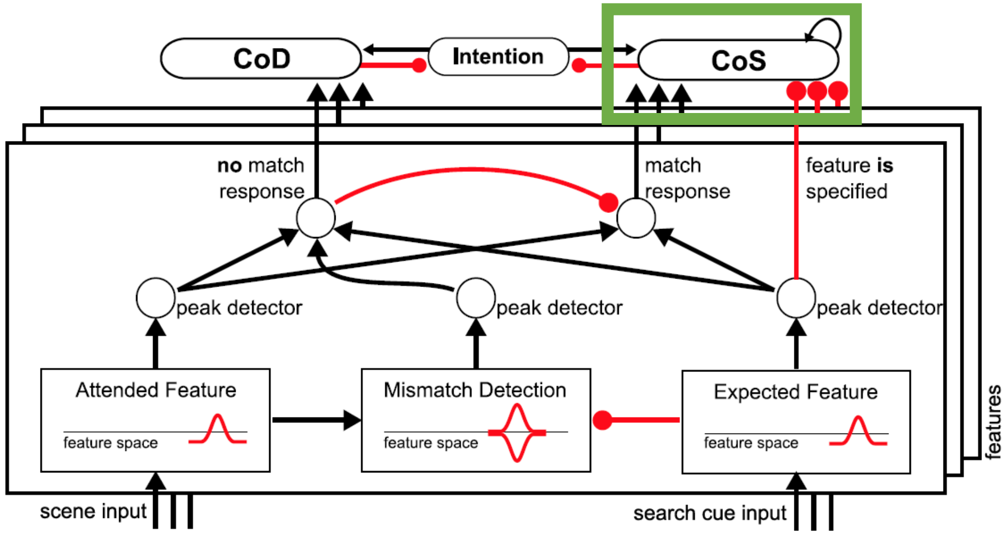
Absence of a **peak** in the **mismatch detection** field, with **peaks** in the **two other fields**, **signals** a **match** and **activates** the **match** response **node**

Subsystem 3: Feature matching



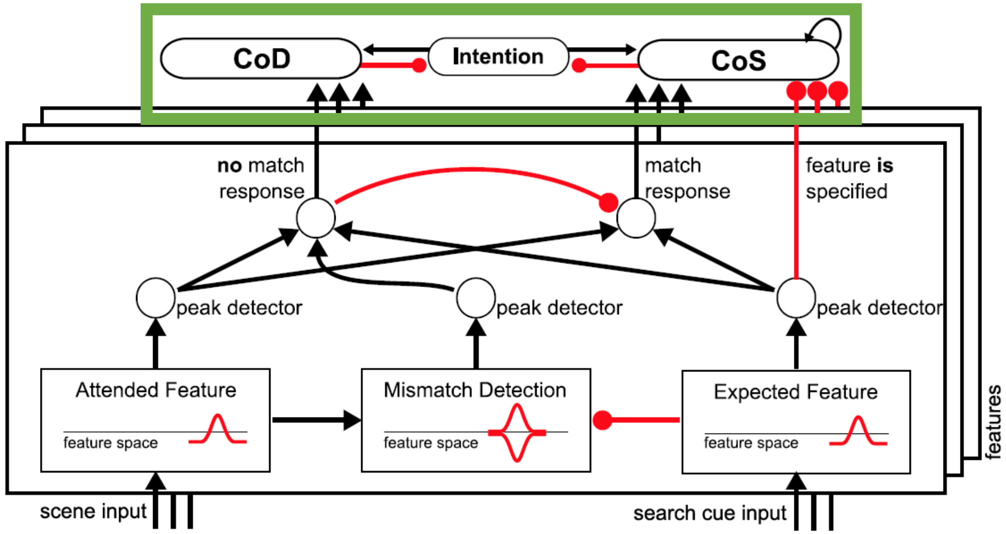
Mismatch within a **single** feature **dimension** is **sufficient** to **activate** the condition of dissatisfaction (**CoD**)

Subsystem 3: Feature matching



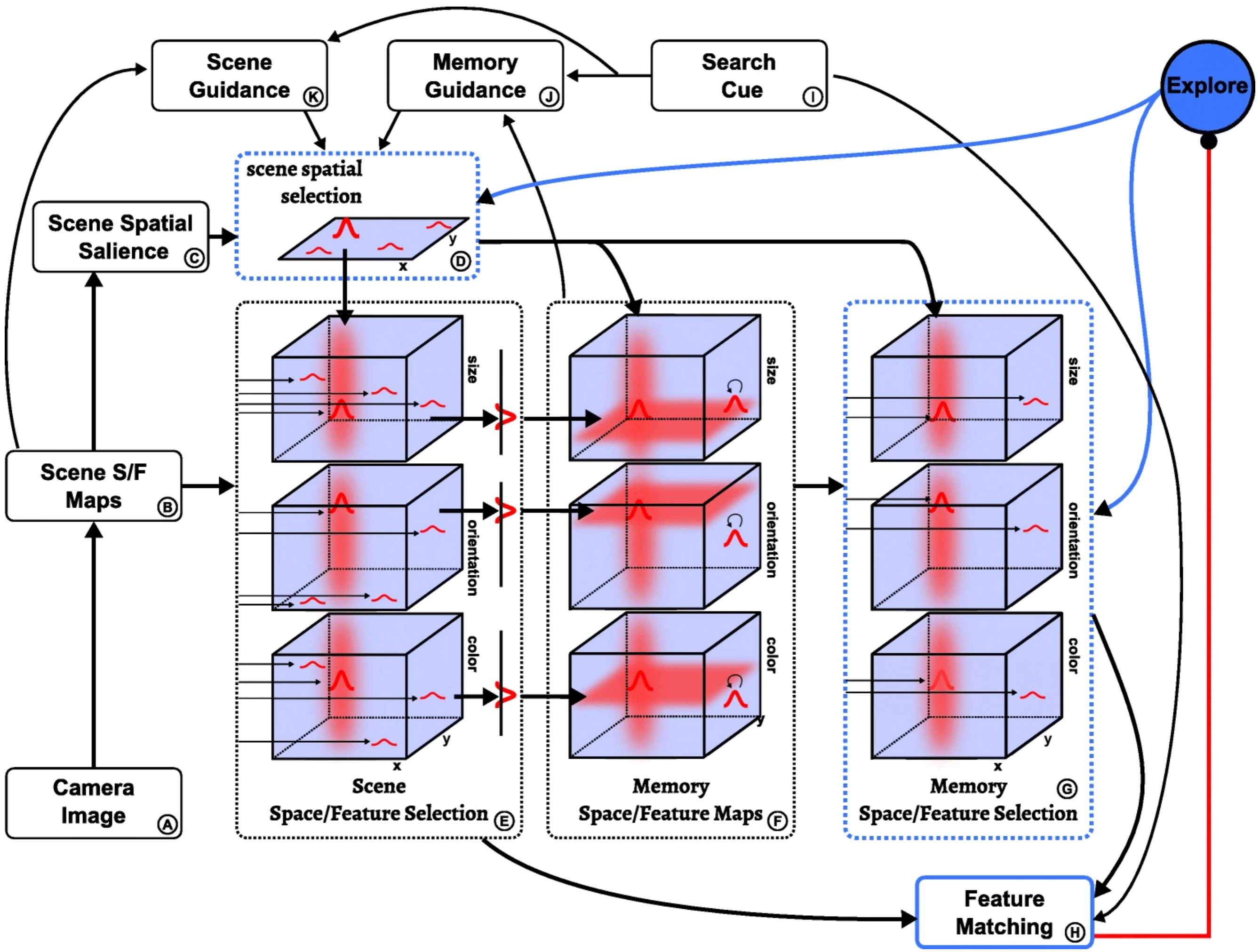
In **contrast**, the condition of satisfaction (**CoS**) node is **only activated** if **all** attended **features match** the search **cue**

Subsystem 3: Feature matching



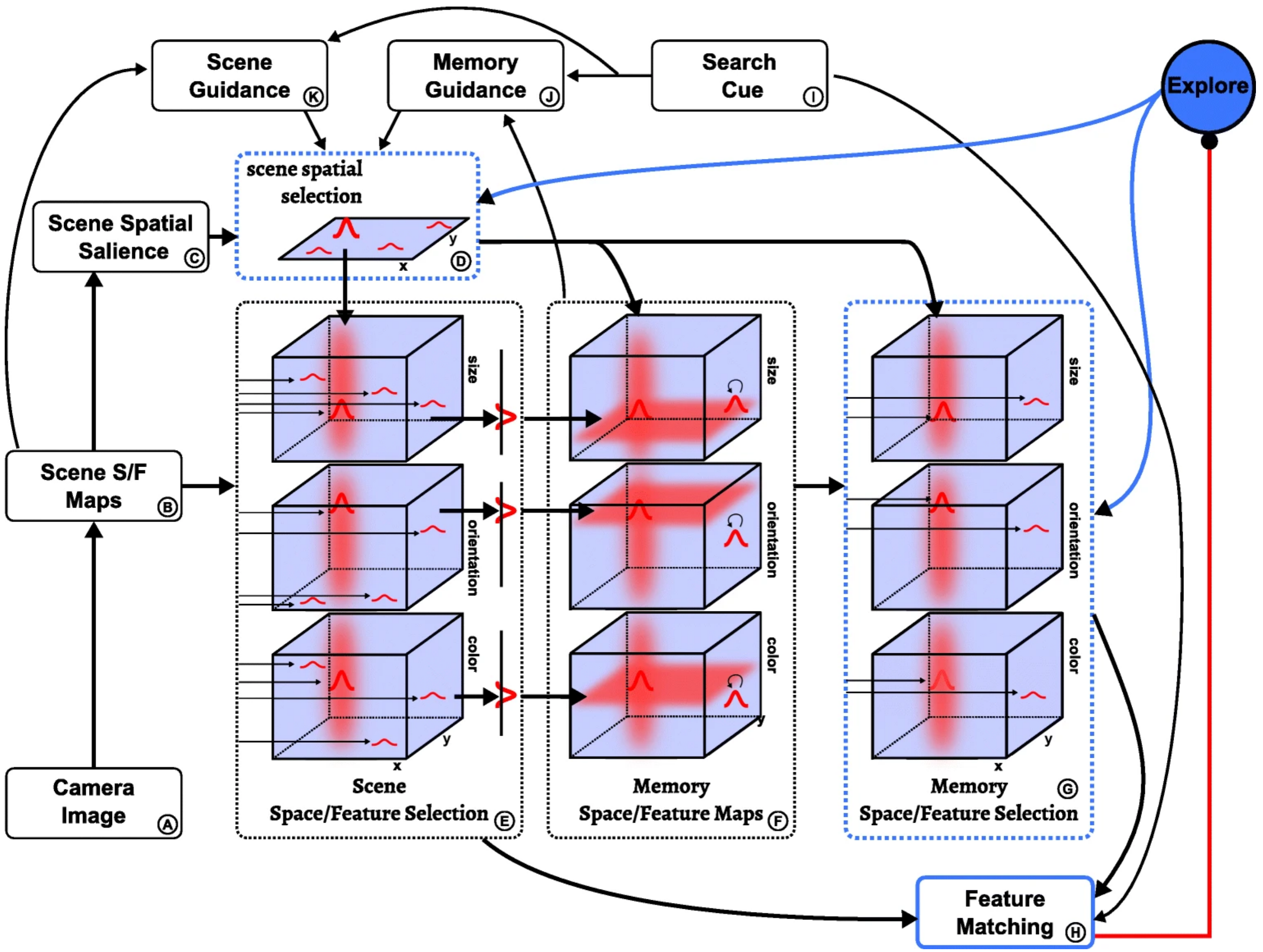
Together with the **intention node**, these two nodes are used to **autonomously generate sequences** of neural processing steps

Task 1: Visual exploration



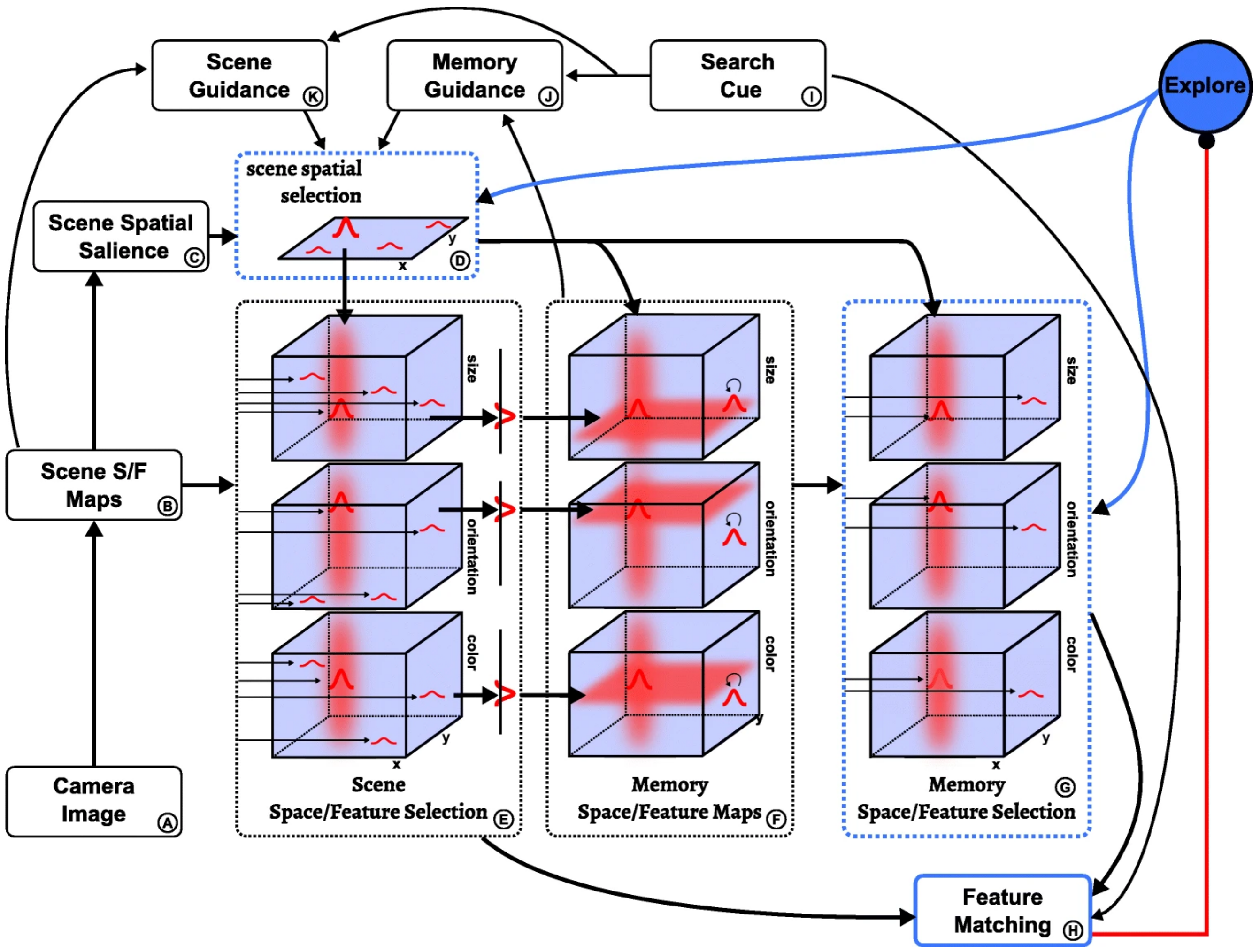
- The **default behavior** of the architecture is the **autonomous** visual **exploration** of the **scene**.

Task 1: Visual exploration



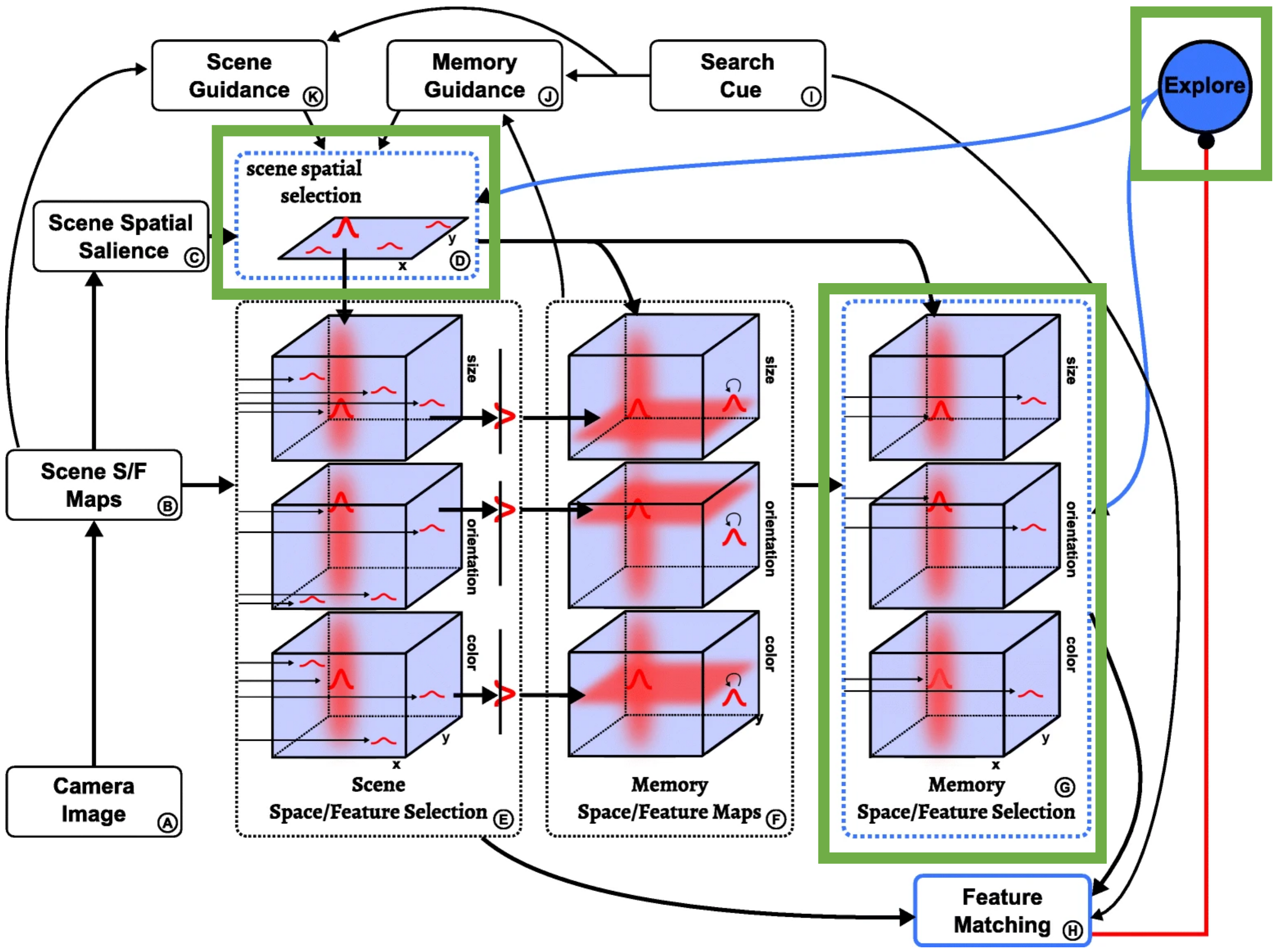
- The default behavior of the architecture is the autonomous visual exploration of the scene.
- In visual **exploration**, salient **locations** in the visual array are **sequentially selected** into the attentional foreground and **features** at these locations are **transferred** to **working memory**.

Task 1: Visual exploration



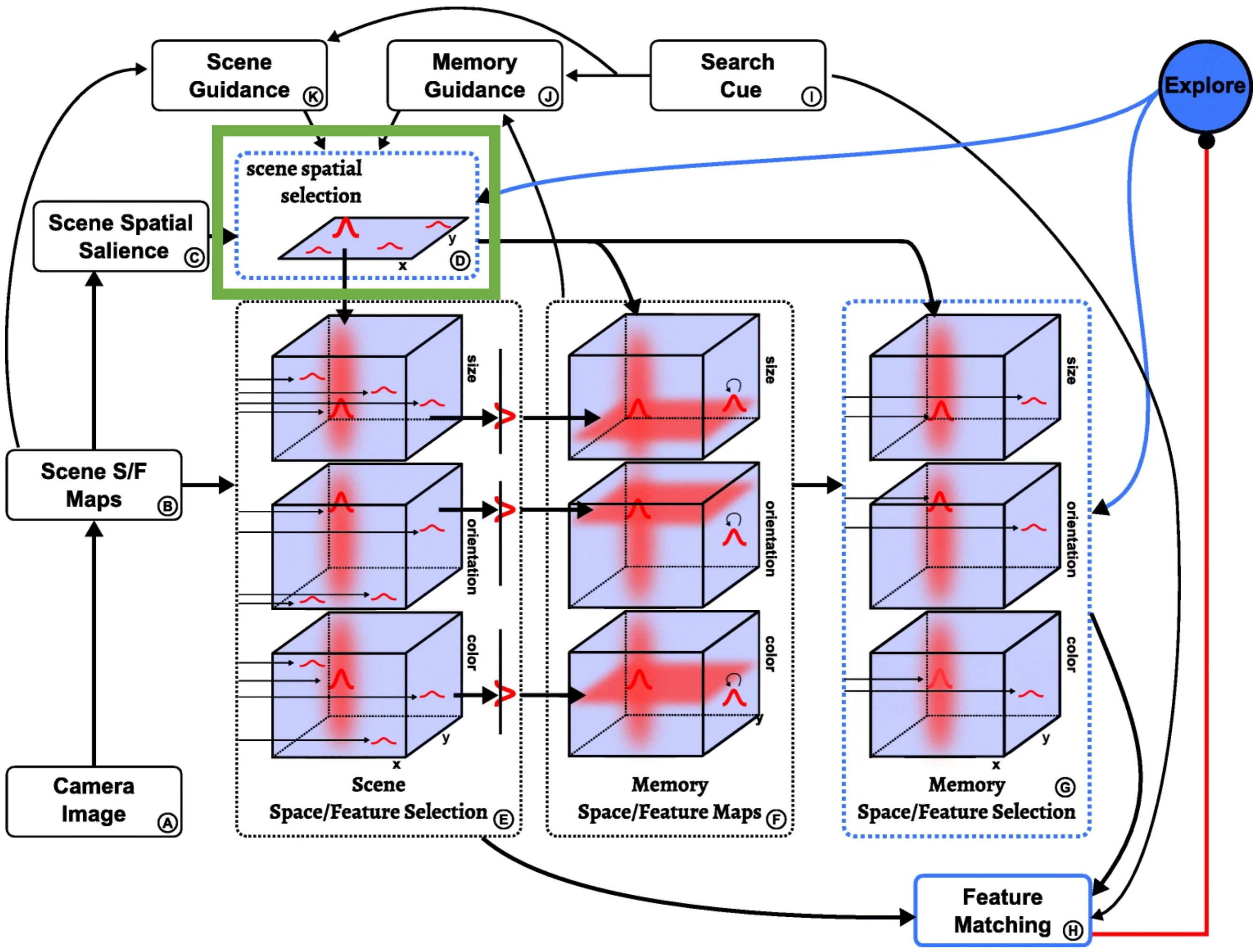
- This is the **sub-network responsible** for visual **exploration** and **memory formation.**

Task 1: Visual exploration



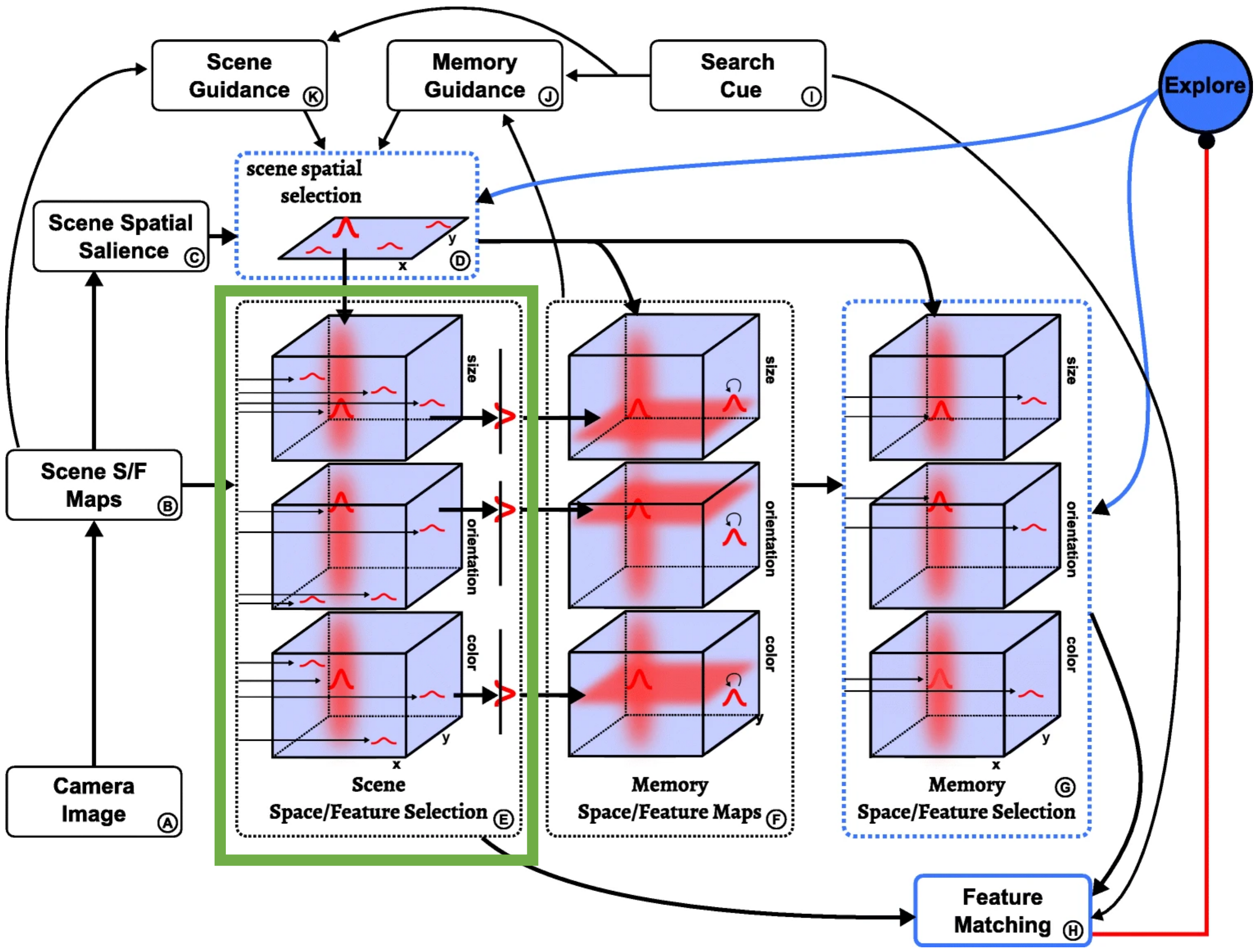
- It becomes **active** when the **explore** task **node boosts** the **scene spatial selection field** and the **memory space/feature selection fields**, enabling these to generate peaks.

Task 1: Visual exploration



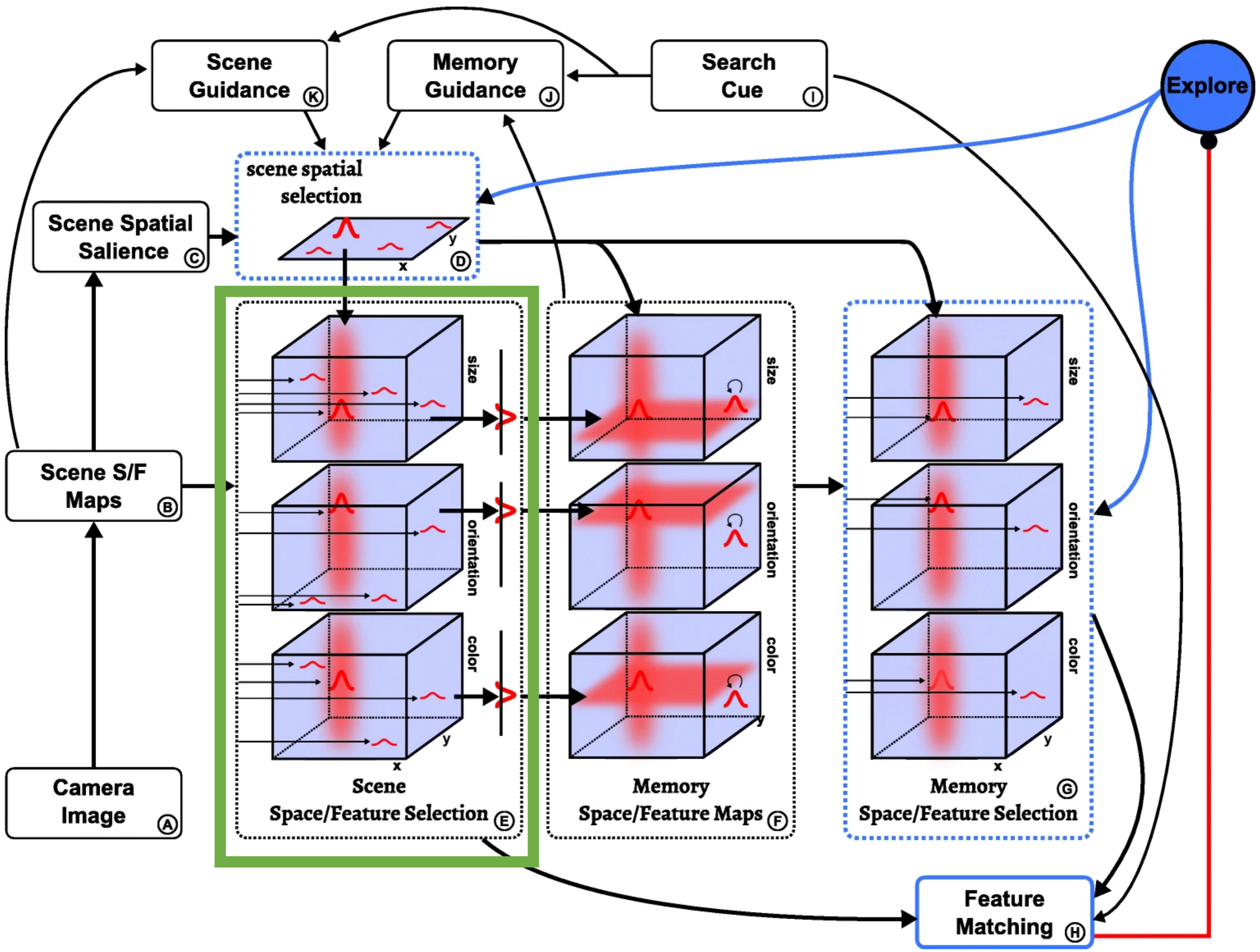
- The **scene spatial selection field** forms a **peak** at a single **location** that is favored by its inputs.

Task 1: Visual exploration



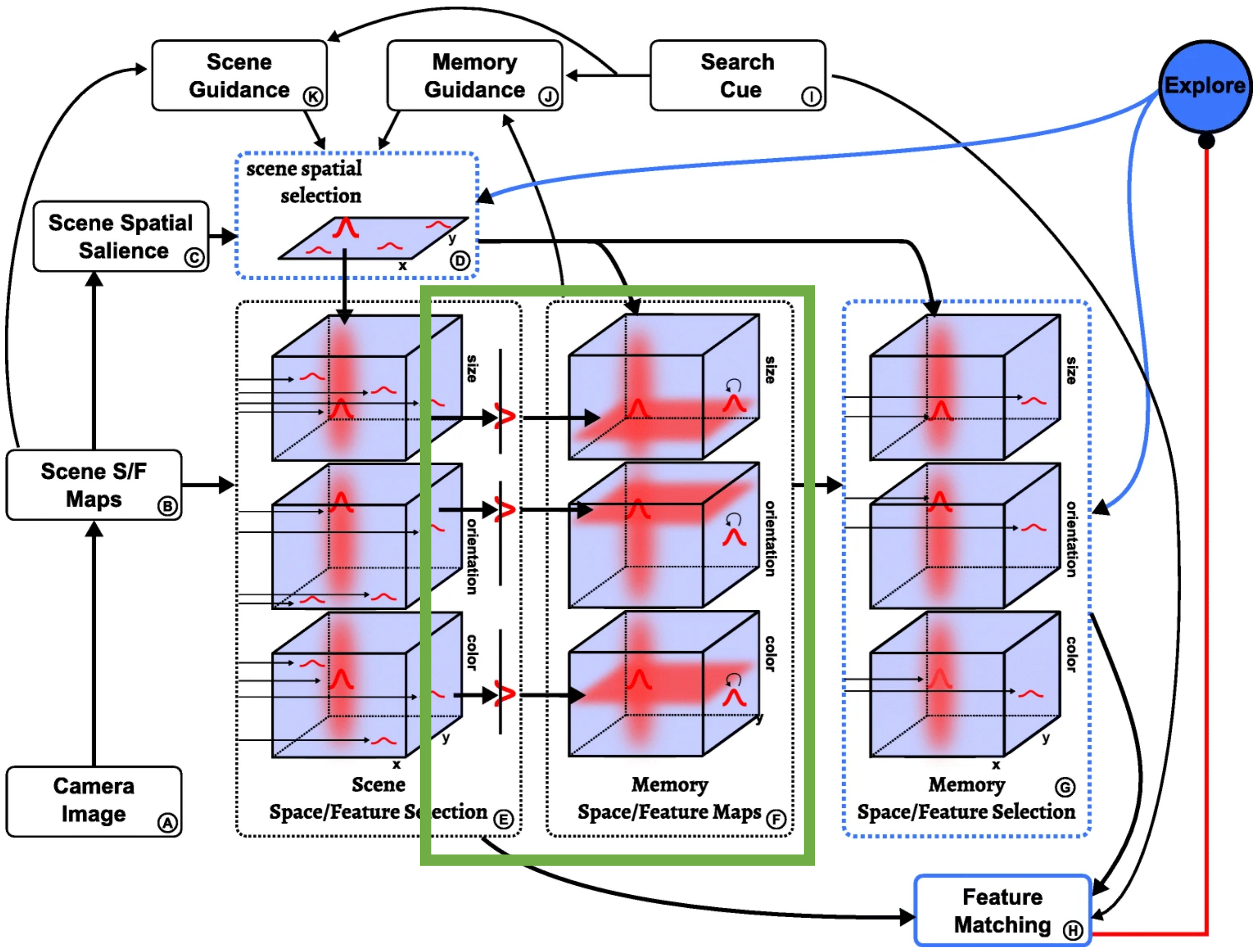
- The **attended location** provides a **cylinder-shaped input** to a set of three-dimensional **scene space/feature selection fields**, which have the same structure as the scene space/feature maps described earlier.

Task 1: Visual exploration



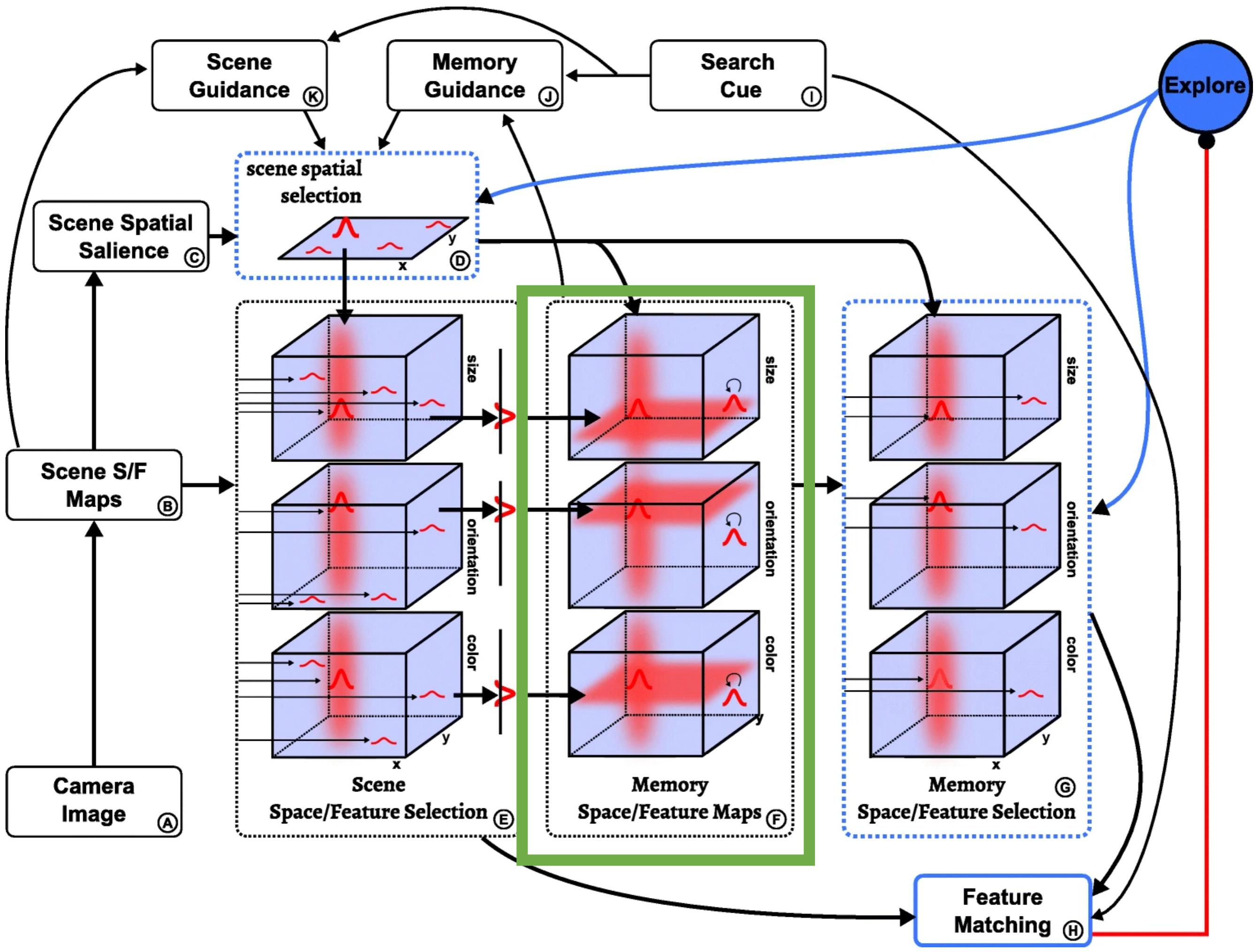
- **Peaks form** where **input** from the scene space/**feature maps overlaps** with the spatially localized **cylinders**, **representing** the **space/feature values** of the **attended object**.

Task 1: Visual exploration



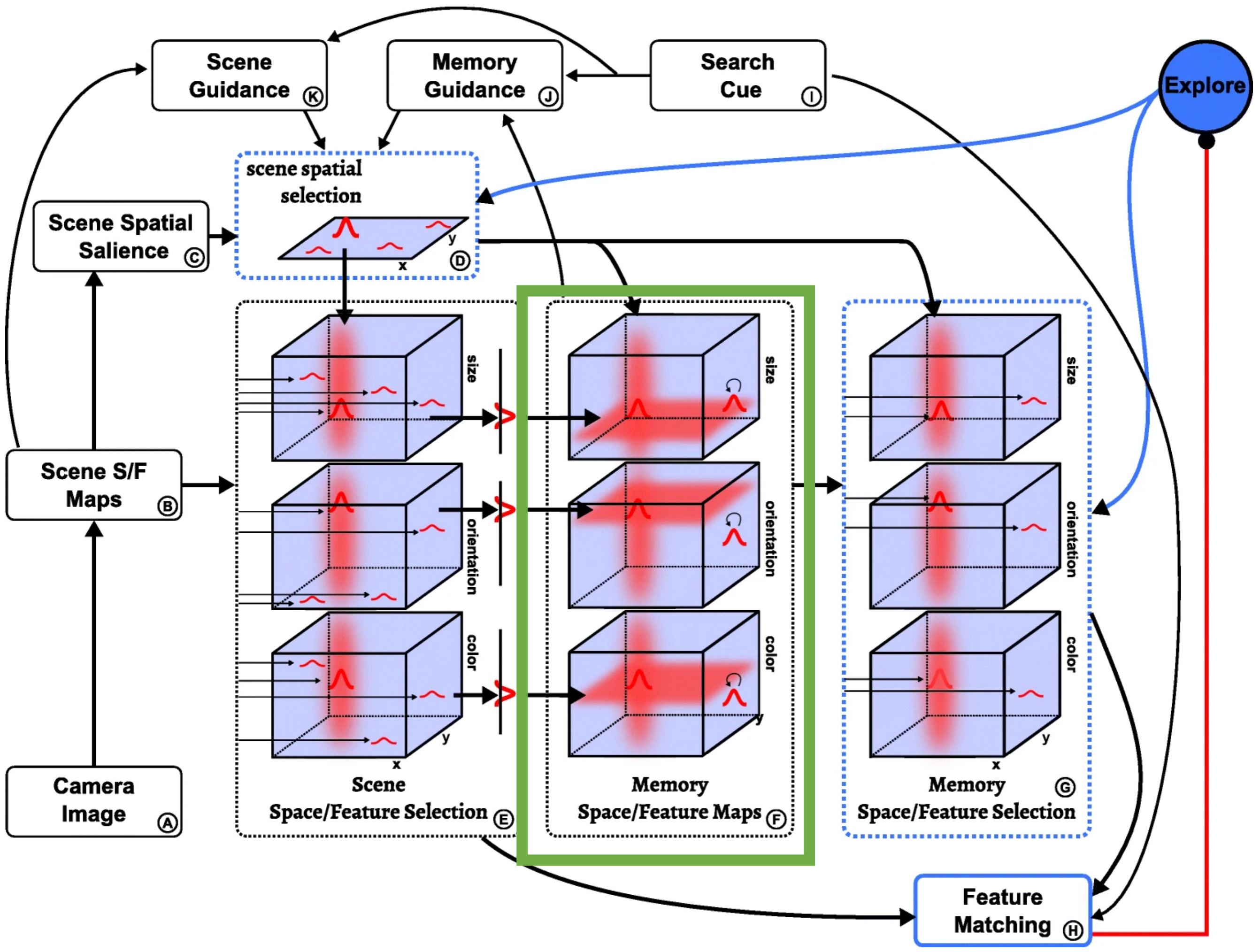
- **Feature information** is **extracted** by **integrating across space** and feeding that sum **as slice input** into the **corresponding** space/feature map in another set of such three-dimensional fields, the **scene memory**, which is operated in the **dynamic regime of sustained activation**.

Task 1: Visual exploration



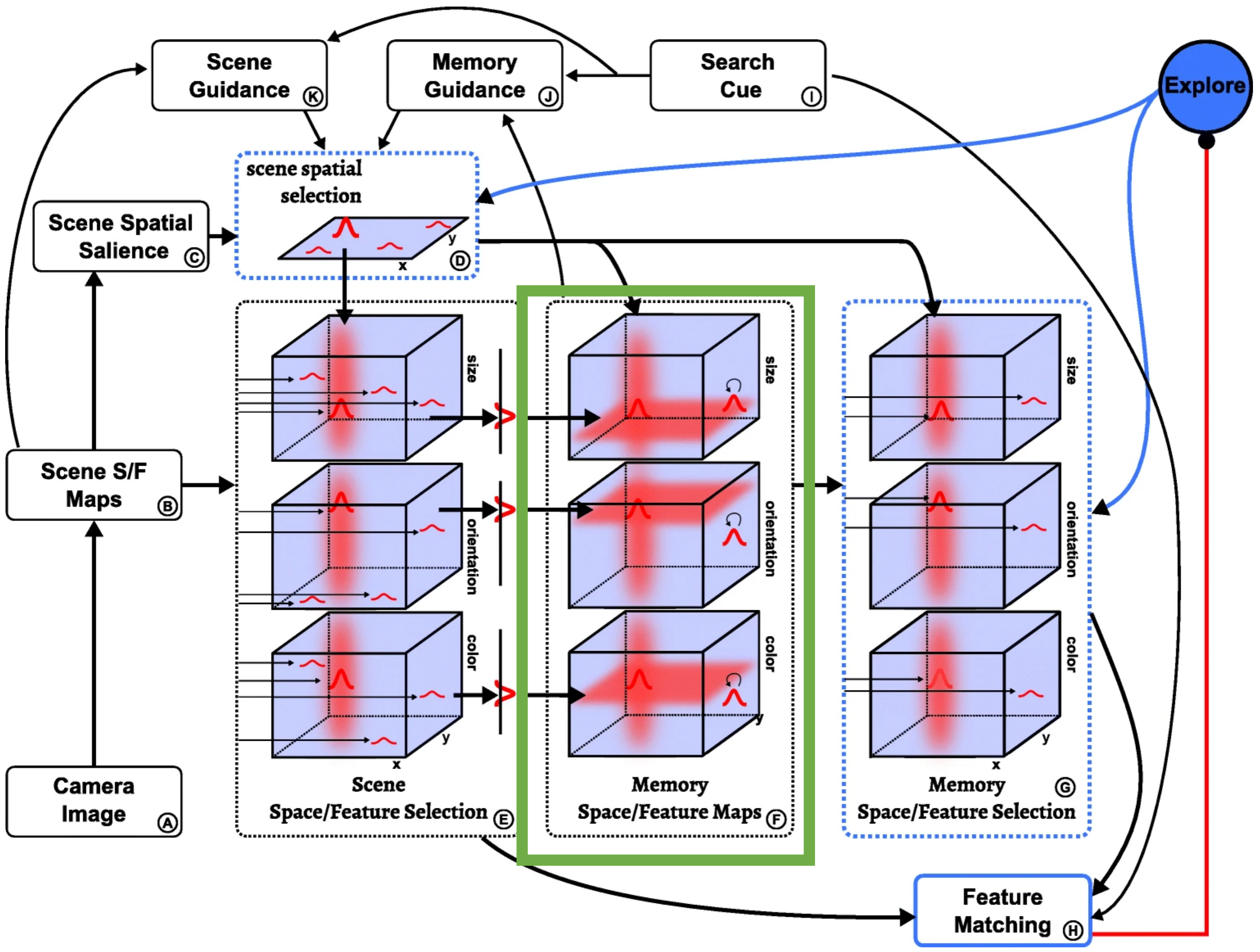
- In **these memory fields**, **peaks form** again where these **slices overlap** with **cylinder** input from the scene spatial selection field. These peaks are added to the **scene working memory**.

Task 1: Visual exploration



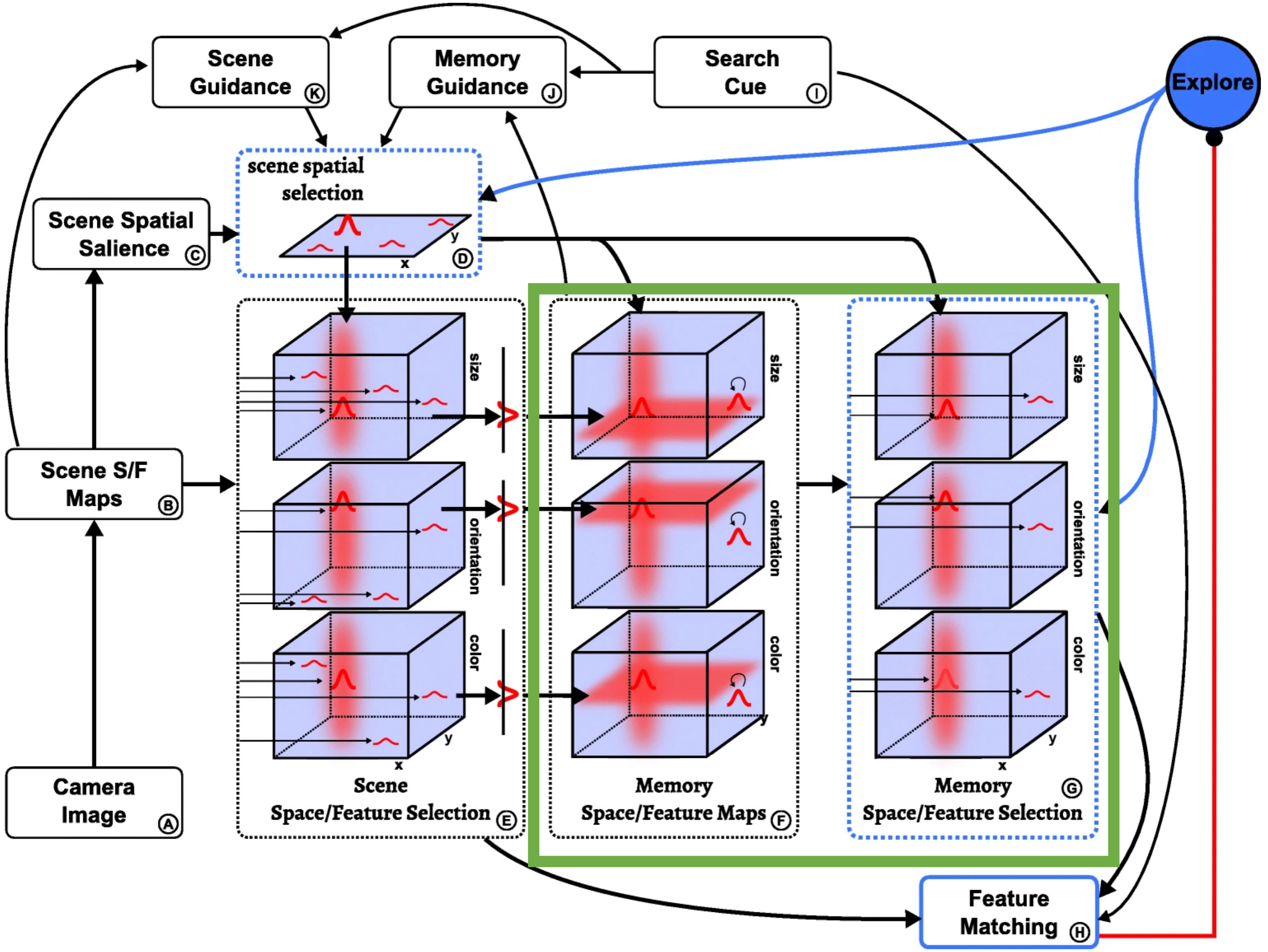
- The **number of peaks** that can be simultaneously **sustained** in the memory space/feature maps is **limited** by the **accumulation** of **inhibition** as additional peaks arise.

Task 1: Visual exploration



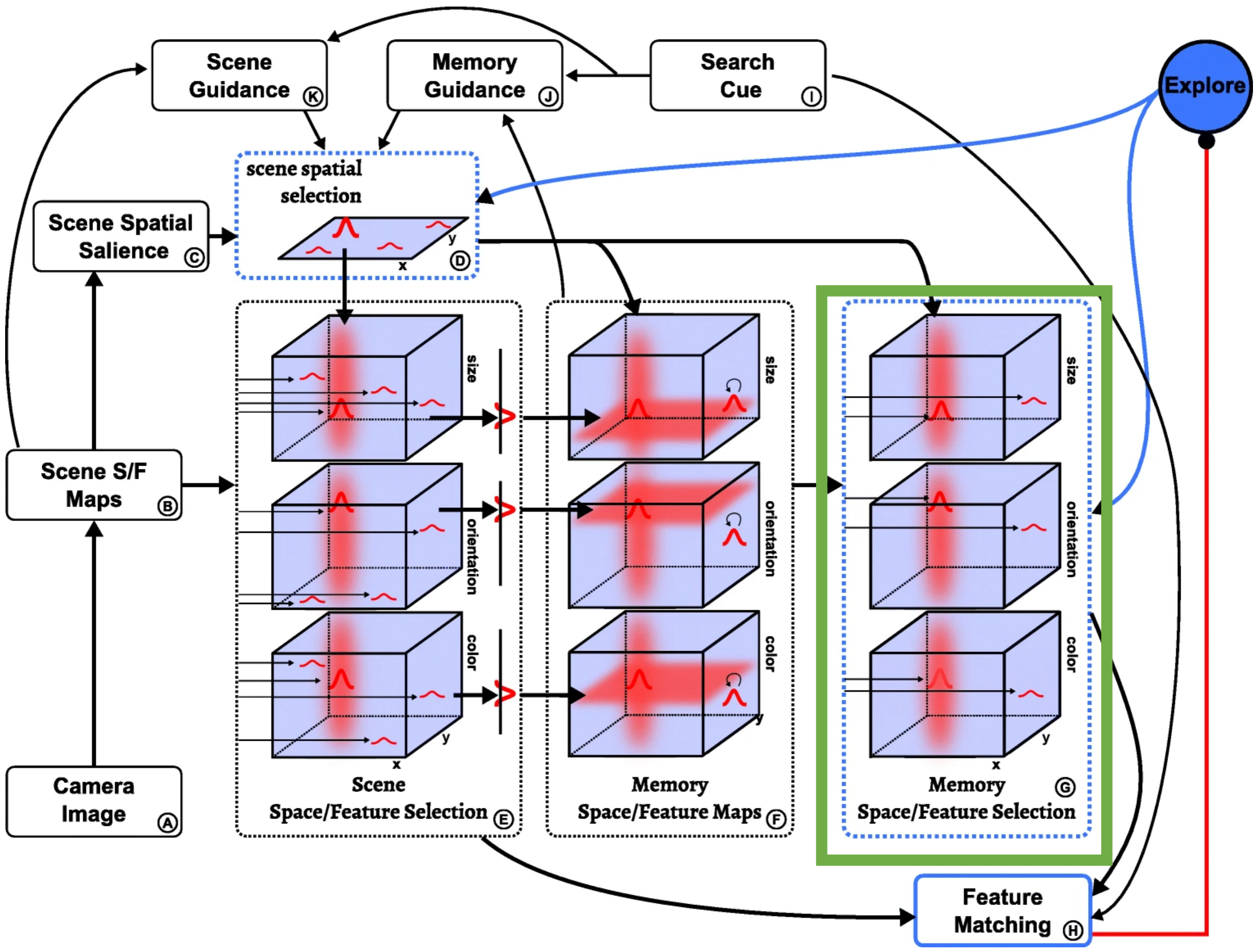
- The **capacity limit depends** on the **balance** of neural **excitation** and **inhibition** in these **fields** and, as was the case for spatial WM, is a **key factor** for **fitting** the **experimental results**.

Task 1: Visual exploration



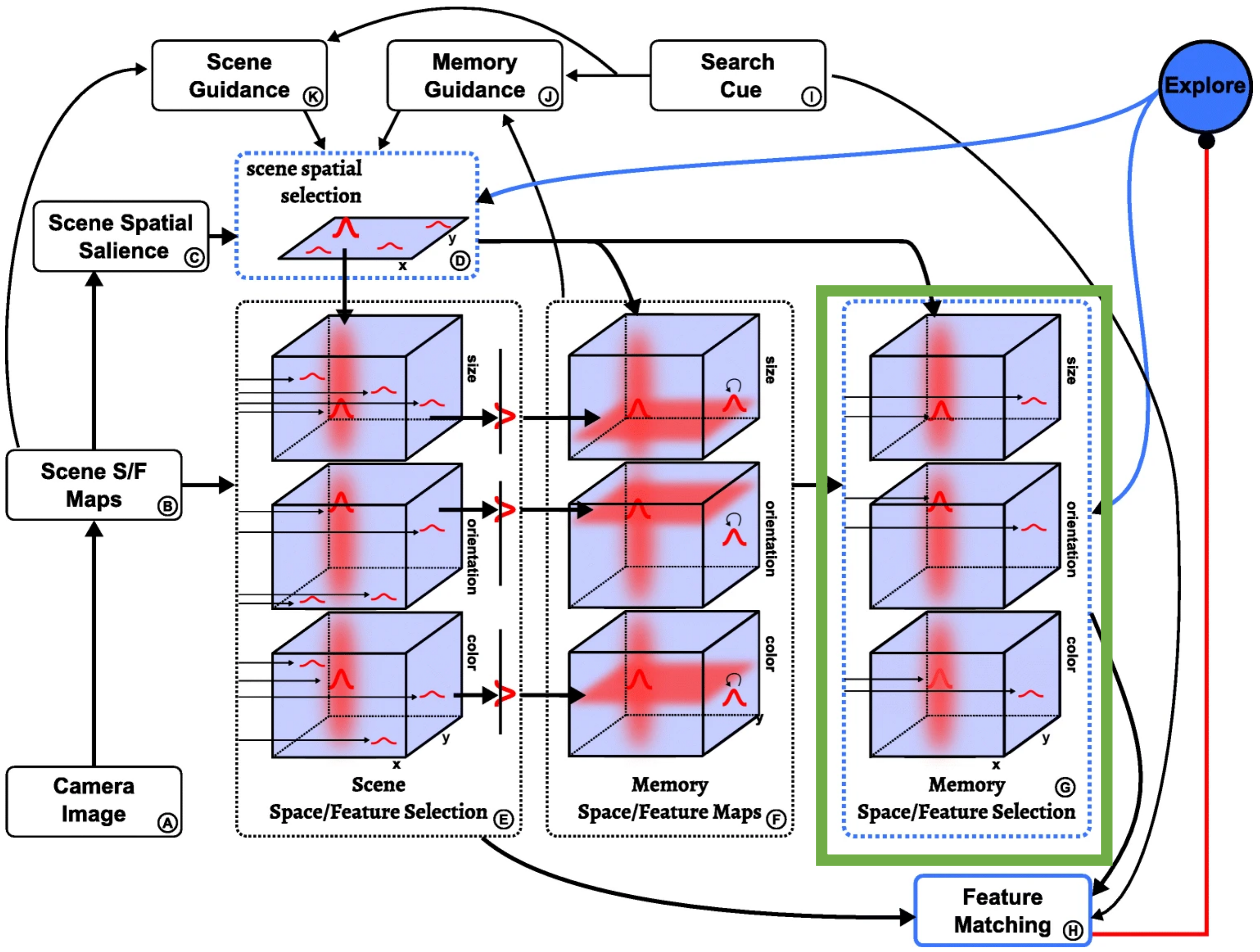
- The **memory** space/feature **maps provide** three-dimensional **input** to an analogous set of three **memory space/feature selection fields**.

Task 1: Visual exploration



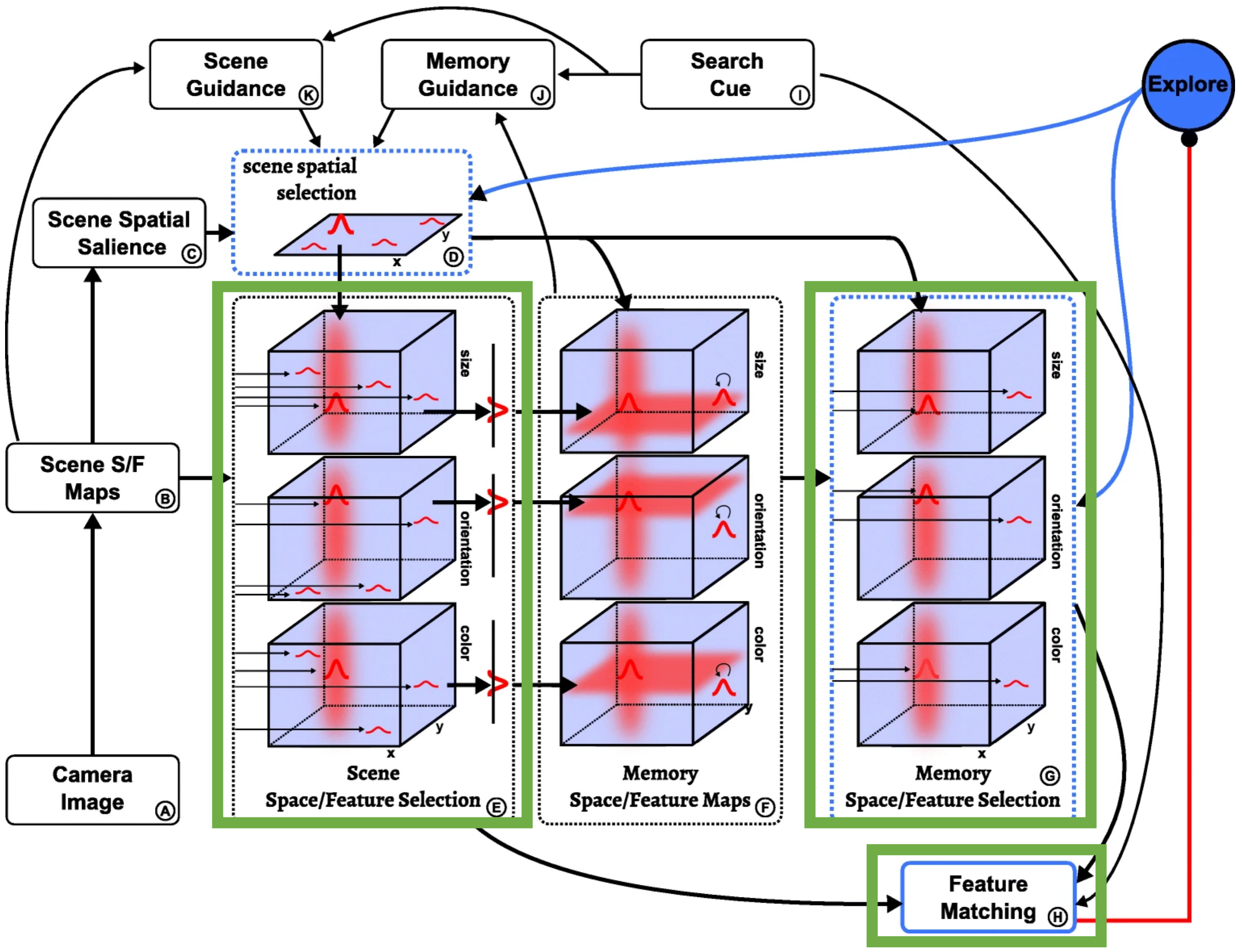
- **In these fields**, one **item from** the **input** is **selected** and **brought above threshold**, again based on **overlap** with **cylinder** input from the **scene spatial selection field**.

Task 1: Visual exploration



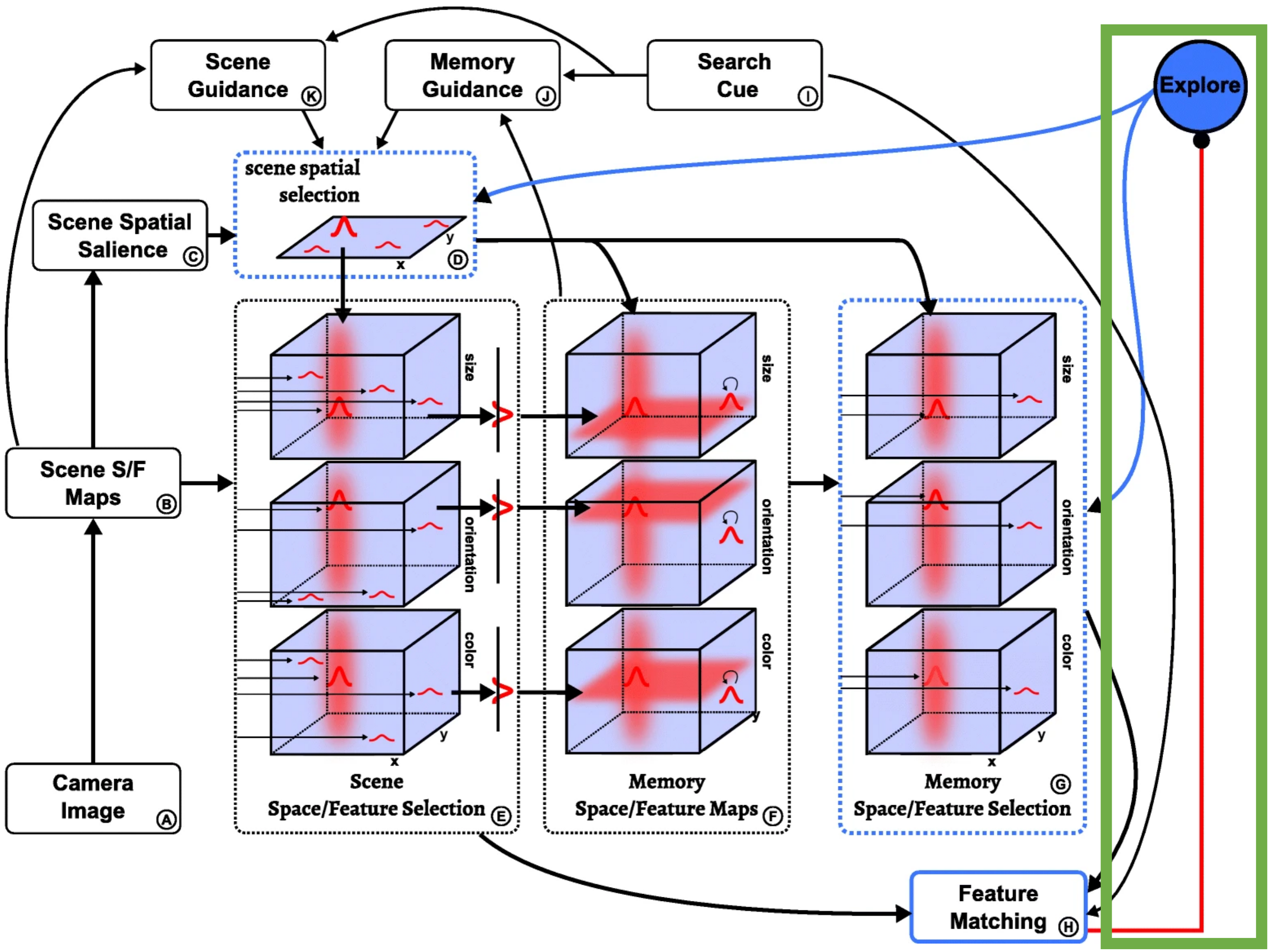
- In these fields, one item from the input is selected and brought above threshold, again based on overlap with cylinder input from the scene spatial selection field.
- The **result** is an **isolated representation** of the **memory item** at the **attended location**.

Task 1: Visual exploration



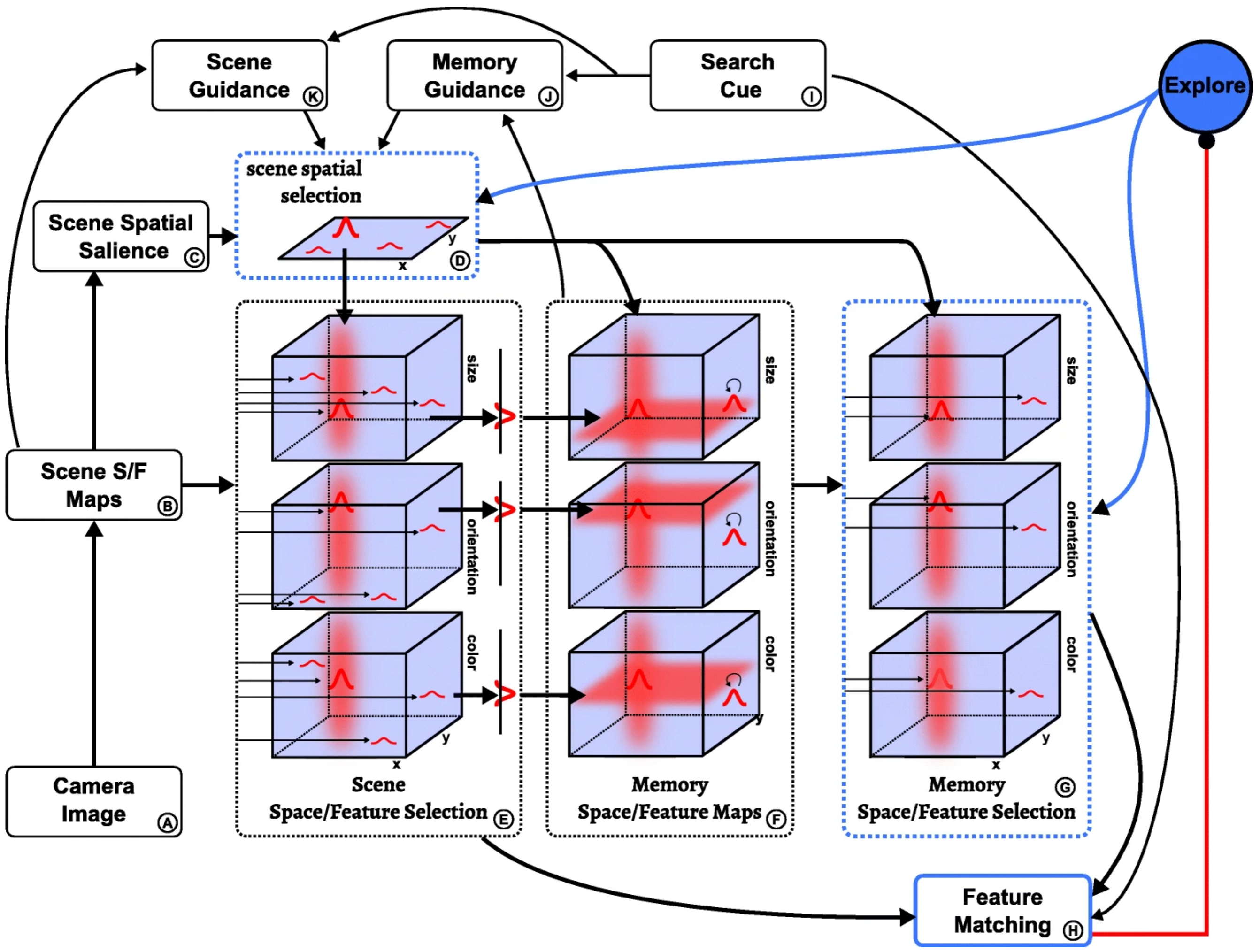
- **Projections** from both **this representation and** the **scene space/feature selection fields converge** onto a **neural feature matching mechanism**, which **detects** whether the attended item's features have been **successfully committed** to scene working **memory**.

Task 1: Visual exploration



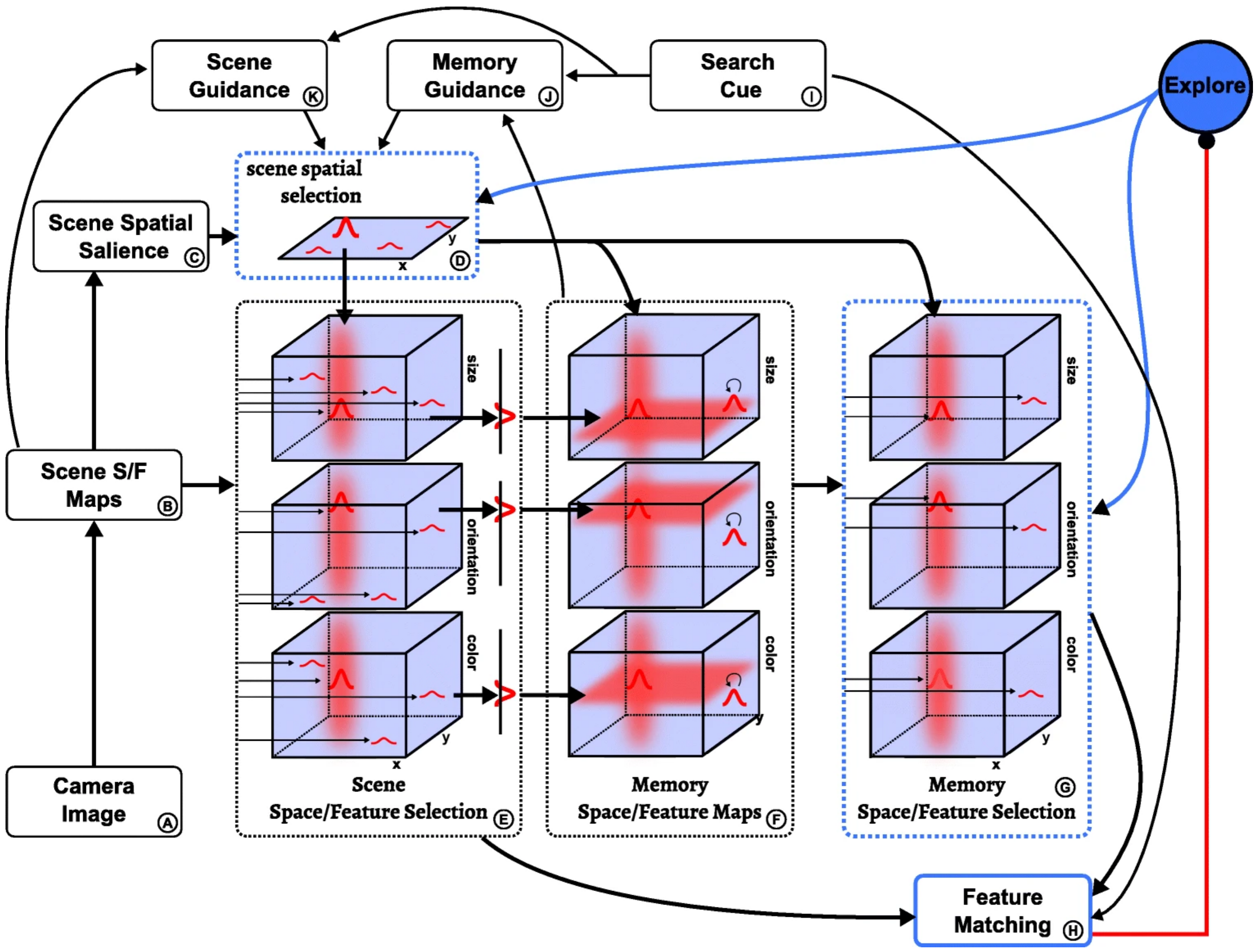
- When this **detection occurs**, the **task node** is **deactivated** through an **inhibitory connection**.

Task 1: Visual exploration



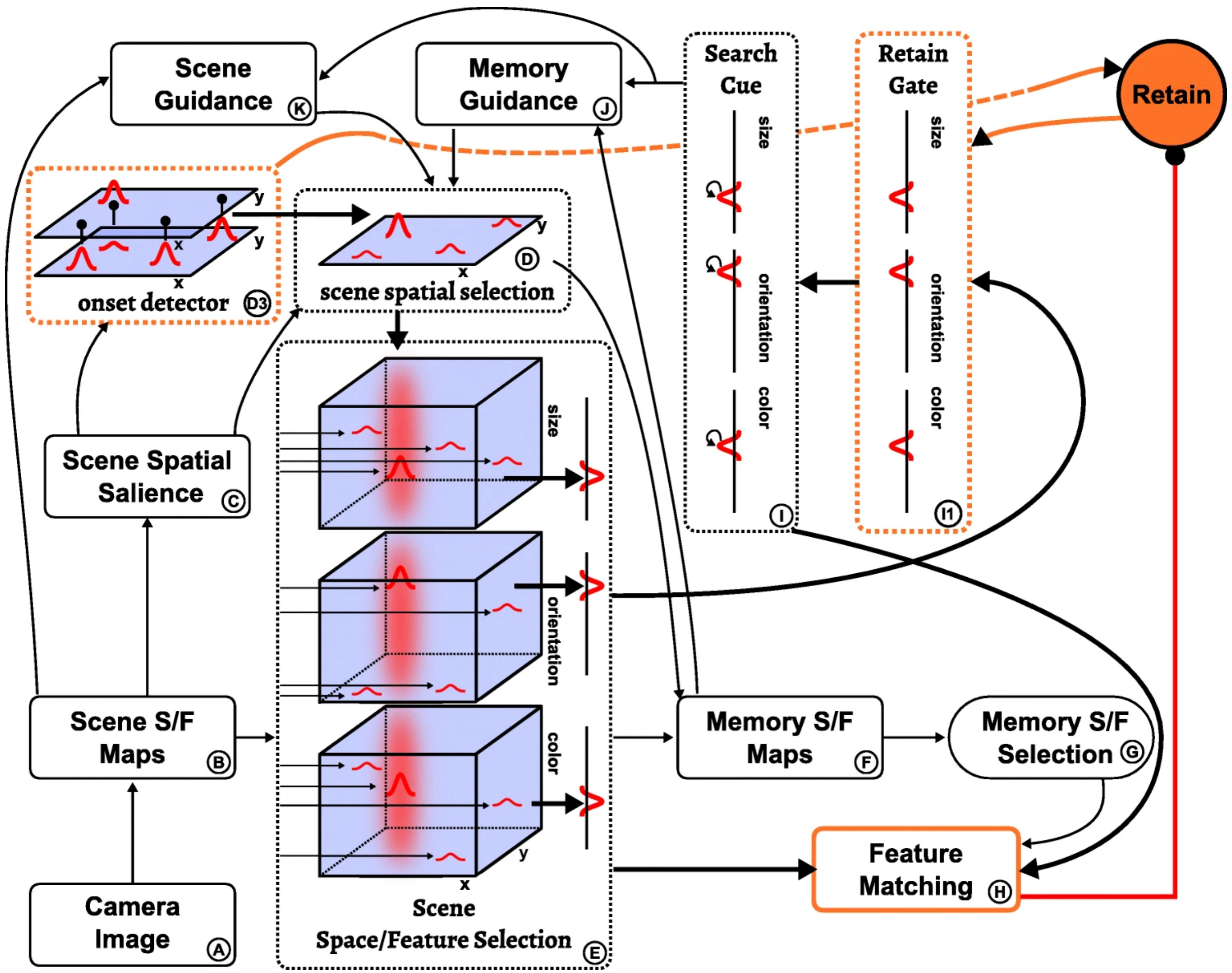
- When this detection occurs, the task node is deactivated through an inhibitory connection.
- **This concludes** one **step** in the **exploration sequence**.

Task 1: Visual exploration



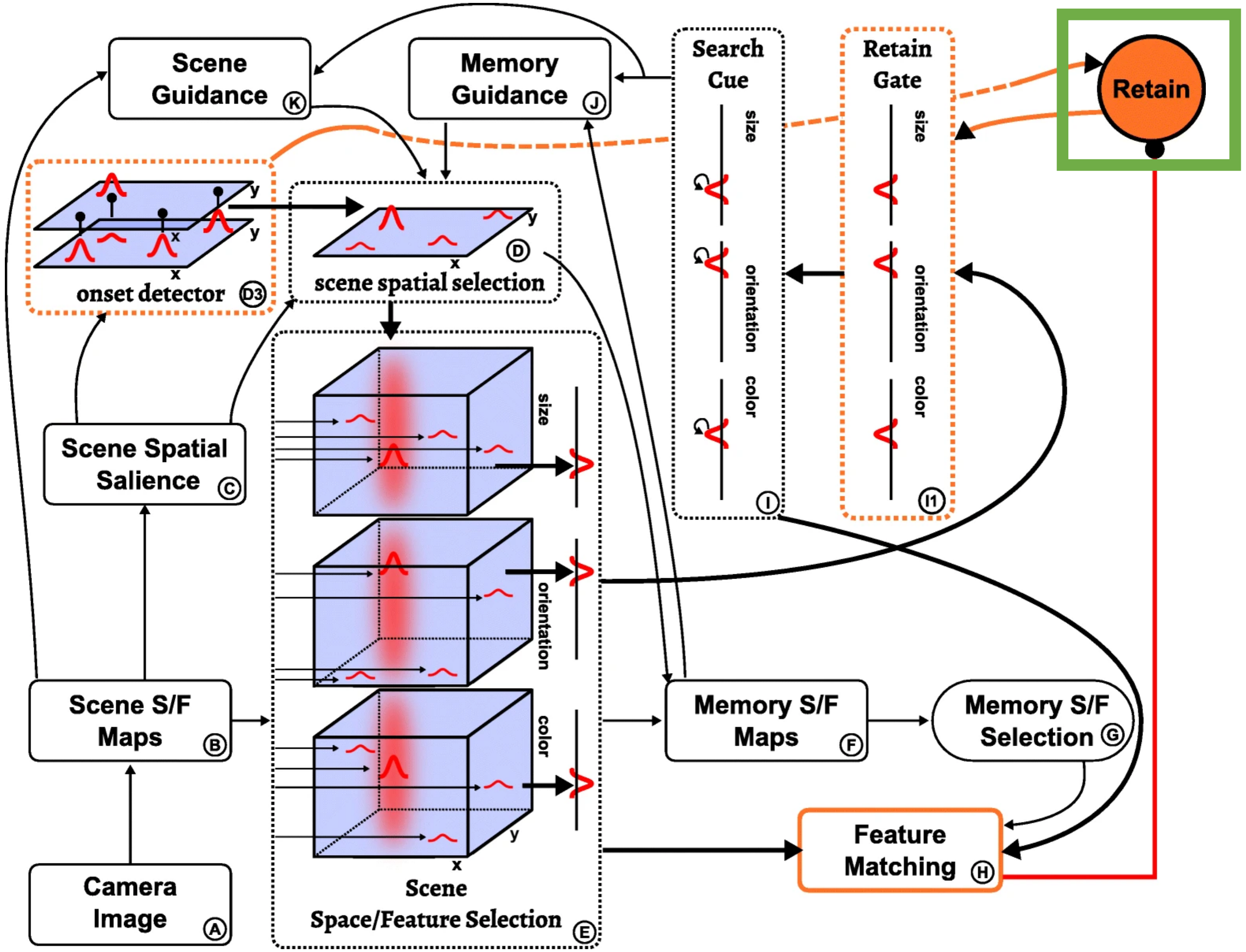
- By **default**, that is, unless another task becomes active, **the task node is then reactivated**, thus **initiating another** cycle of **attentional selection** and commitment to working memory.

Task 2: Retaining feature cue



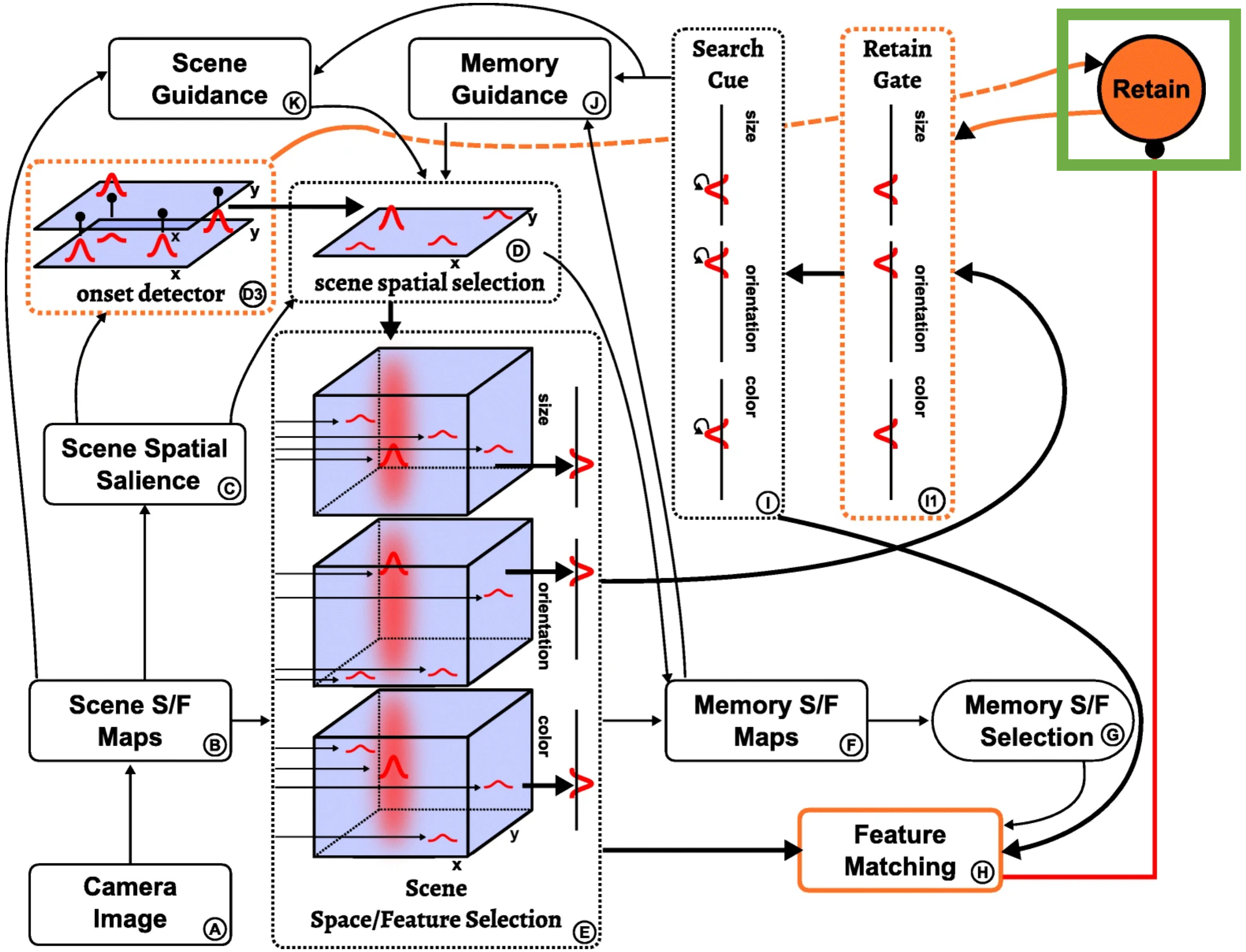
- This is the **sub-network** that is **responsible** for **retaining** a feature **cue** for visual **search**.

Task 2: Retaining feature cue



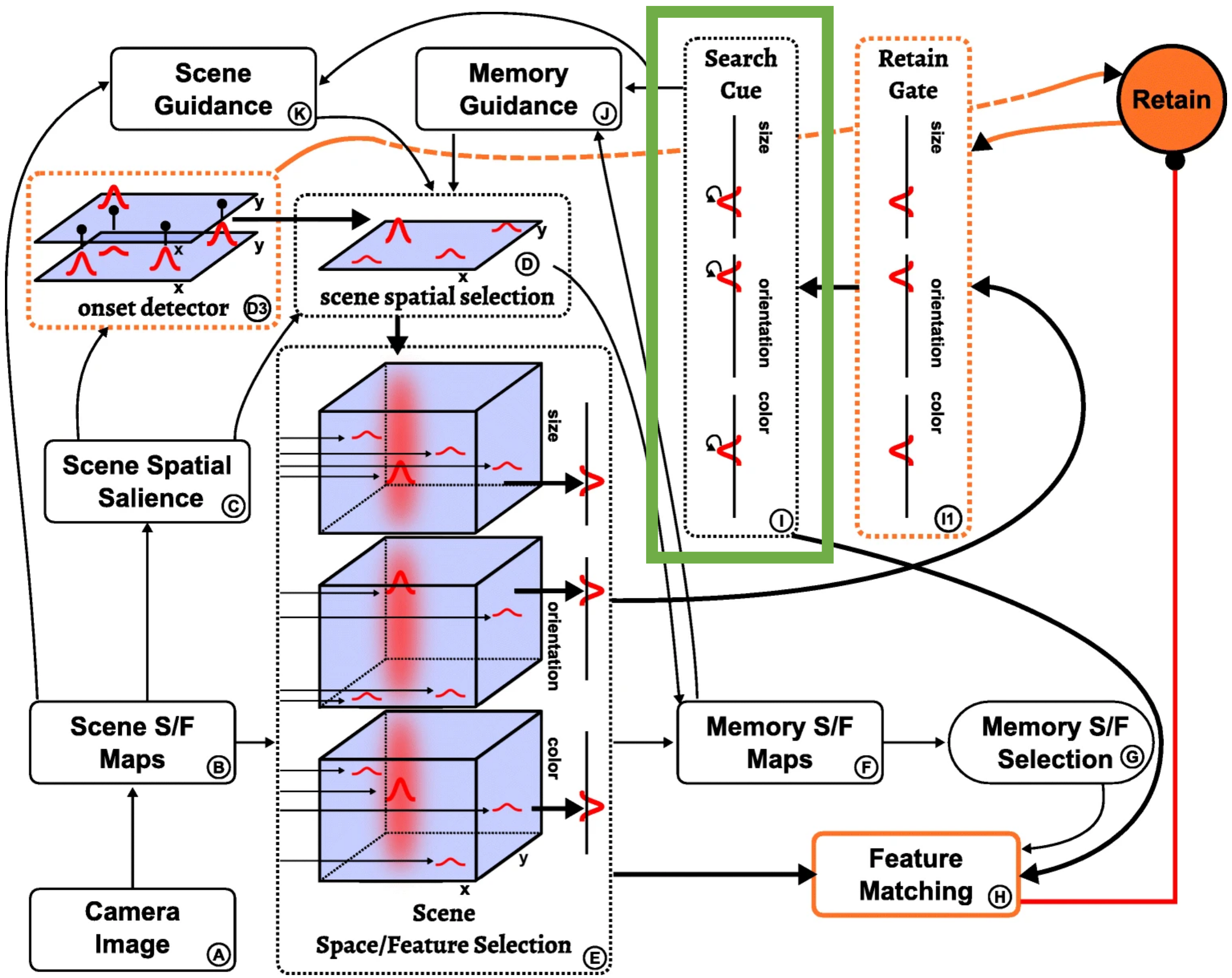
- This is the sub-network that is responsible for retaining a feature cue for visual search.
- It is **activated** by the **retain** task **node**, which may itself be activated from different sources depending on the cognitive task at hand.

Task 2: Retaining feature cue



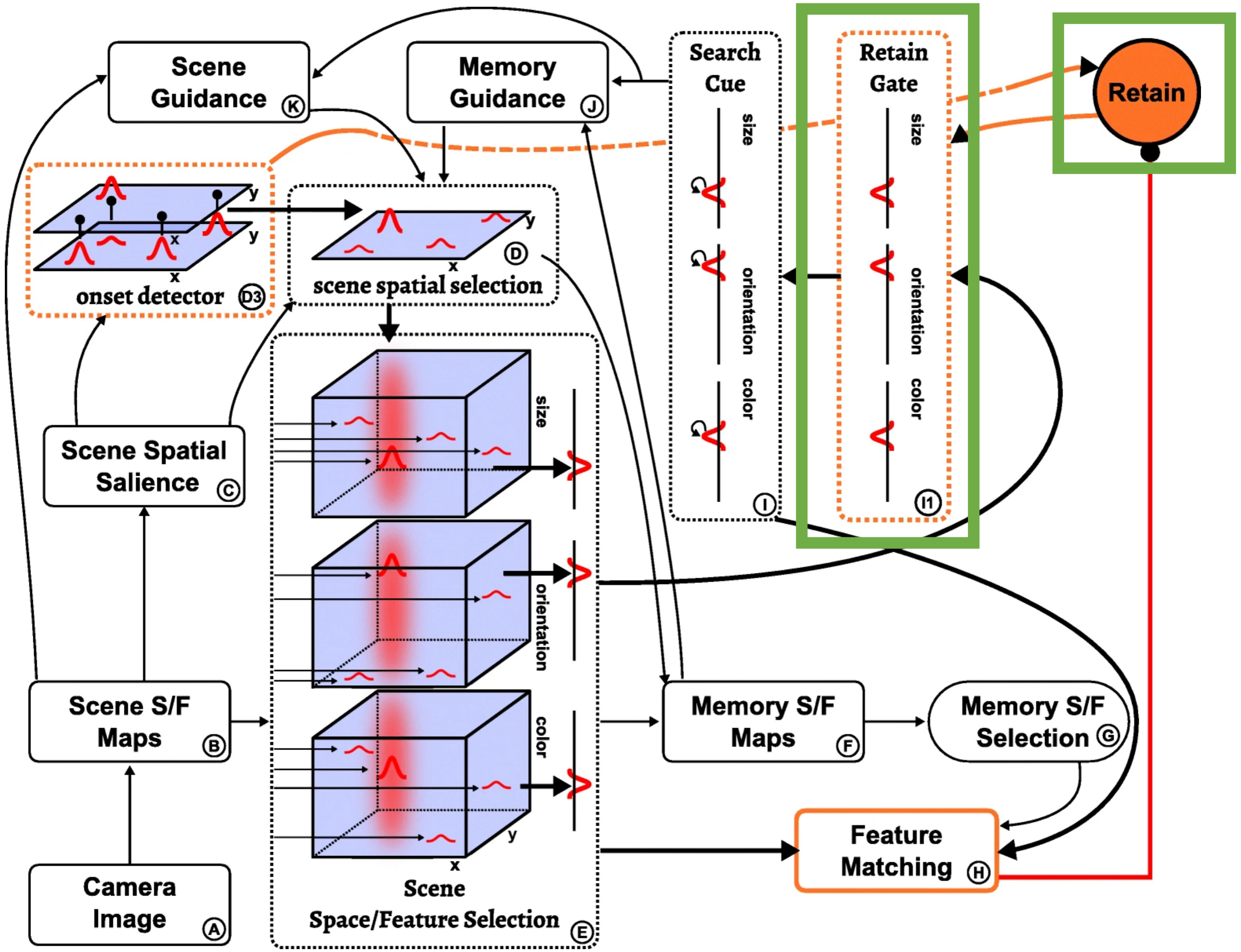
- In the **current context**, the **task node is activated** by the **onset detector** when that system detects a change in the visual scene.

Task 2: Retaining feature cue



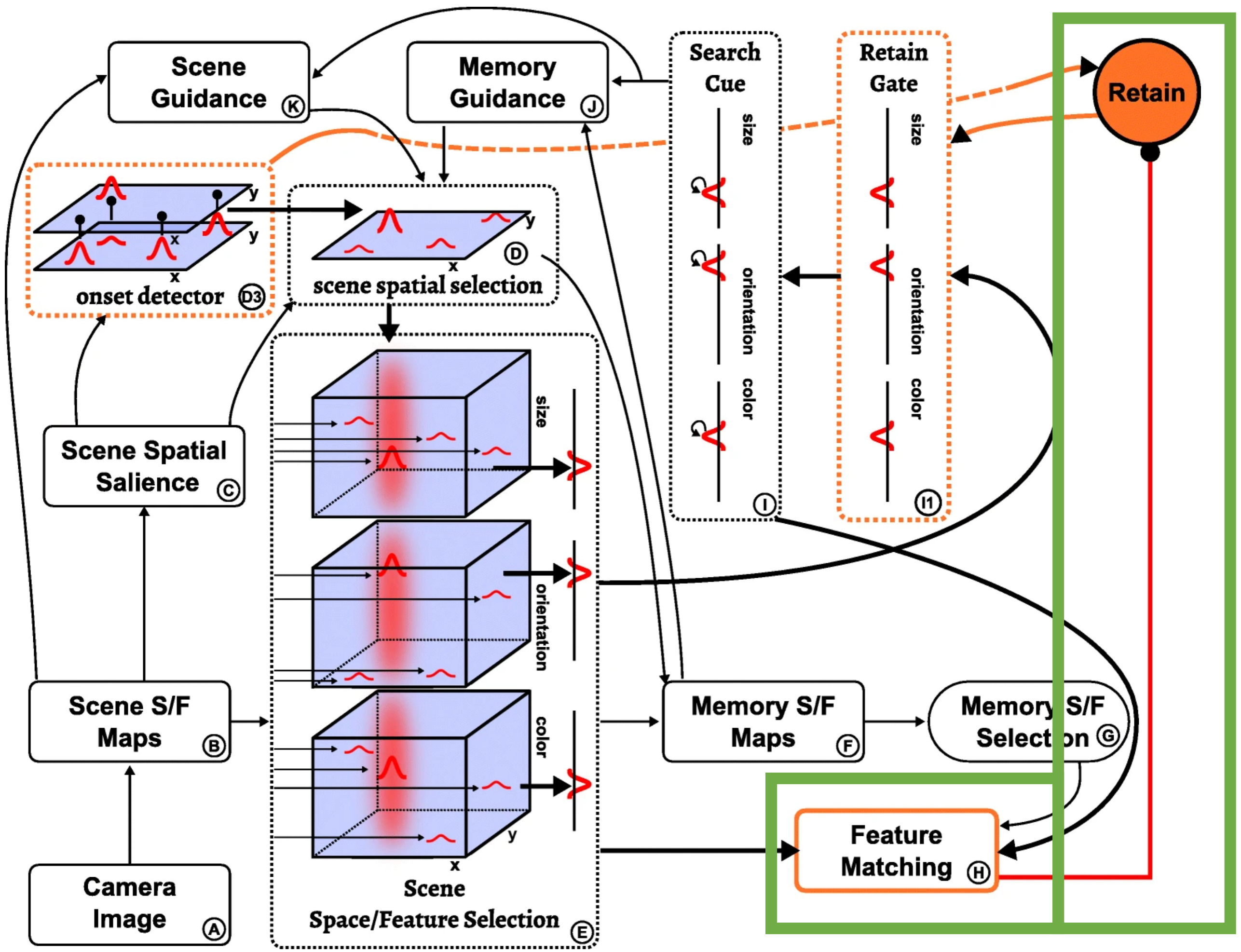
- The **retain process consists** of **storing** currently **attended feature** values as **self-sustained peaks** in the **search cue fields**. These are one-dimensional since only the feature values of the cue, not its location, are relevant.

Task 2: Retaining feature cue



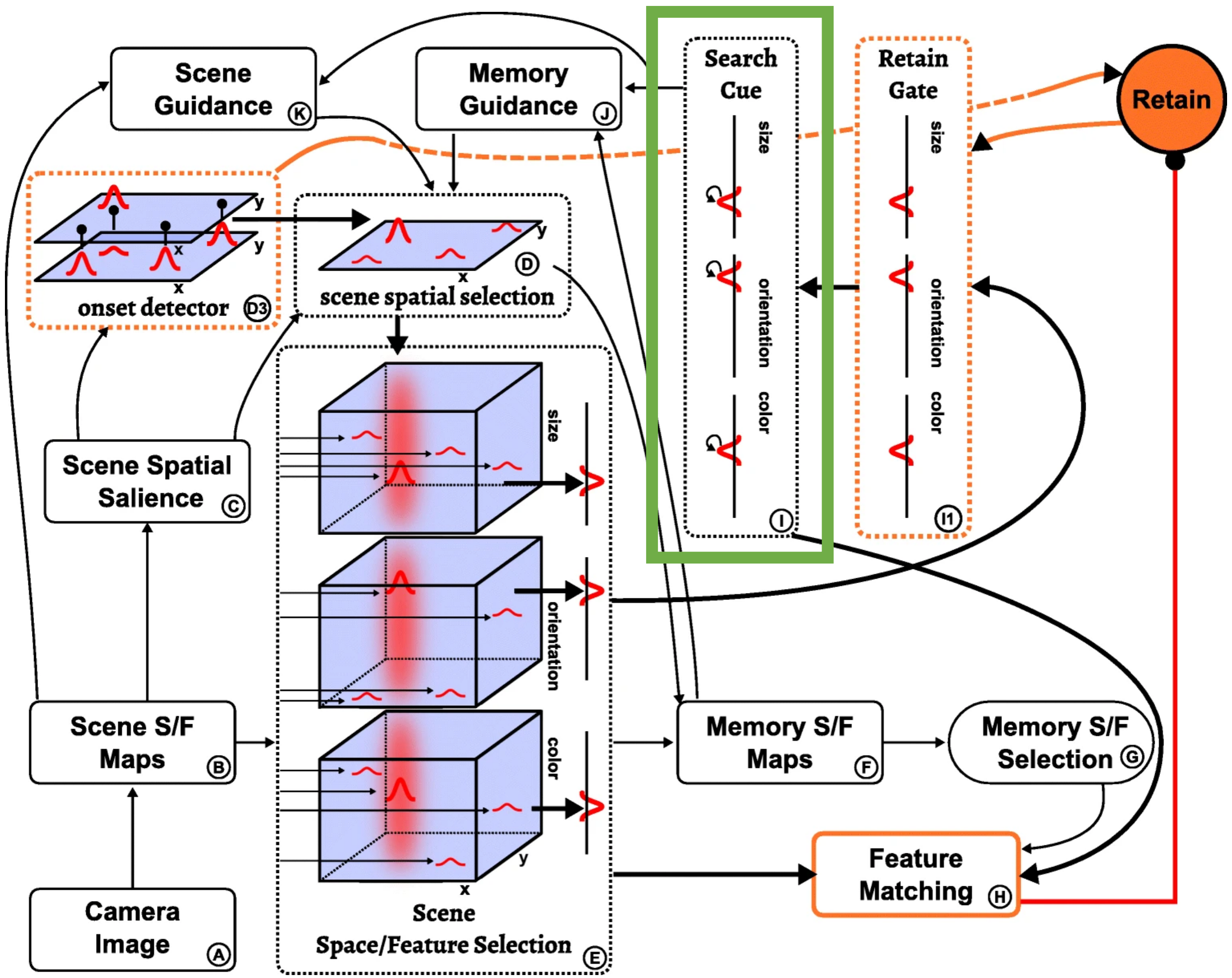
- To **forward feature values** from the **scene space/feature selection fields** to the **search cue fields**, the **retain node** homogeneously **boosts** activation in the **retain gate fields, enabling** them to **build peaks** and thus to **pass** on **activation**.

Task 2: Retaining feature cue



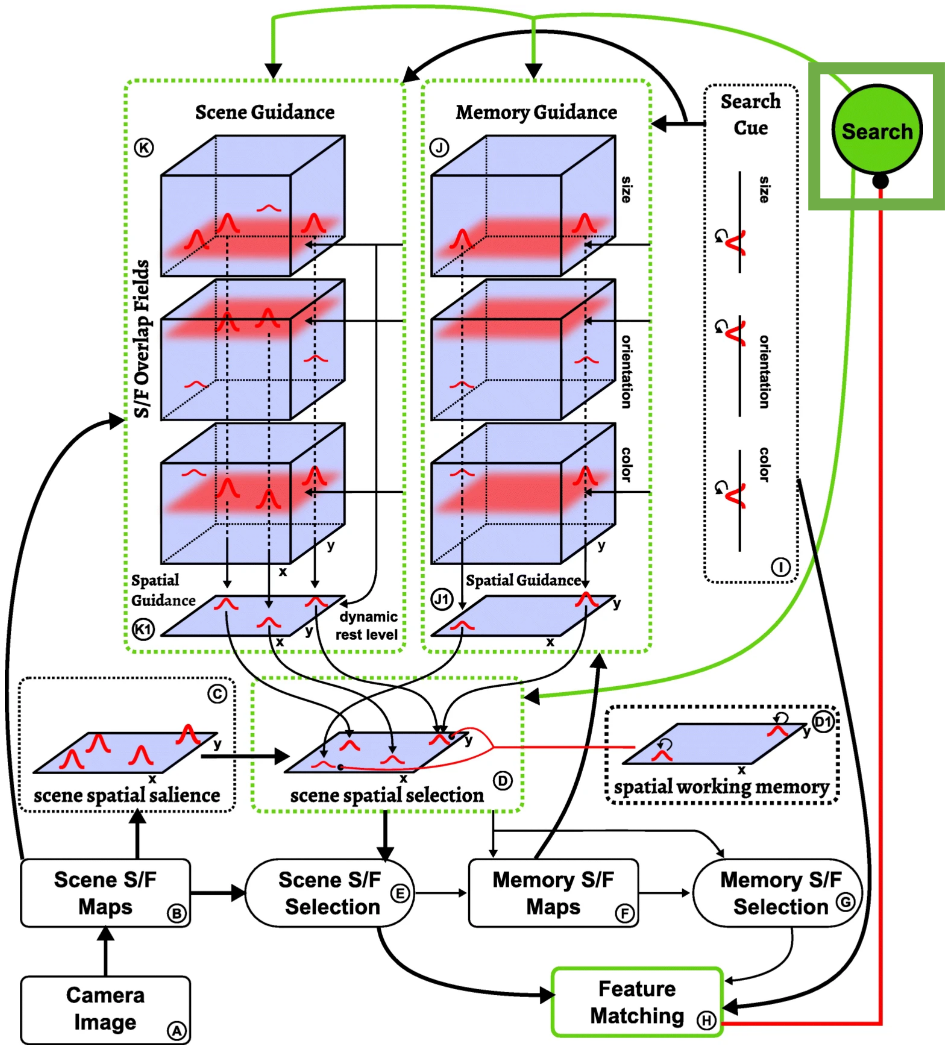
- The **retain sub-task** is **terminated** once the **content** of the **search-cue** fields **matches** the **features** of the **currently attended item**.

Task 2: Retaining feature cue



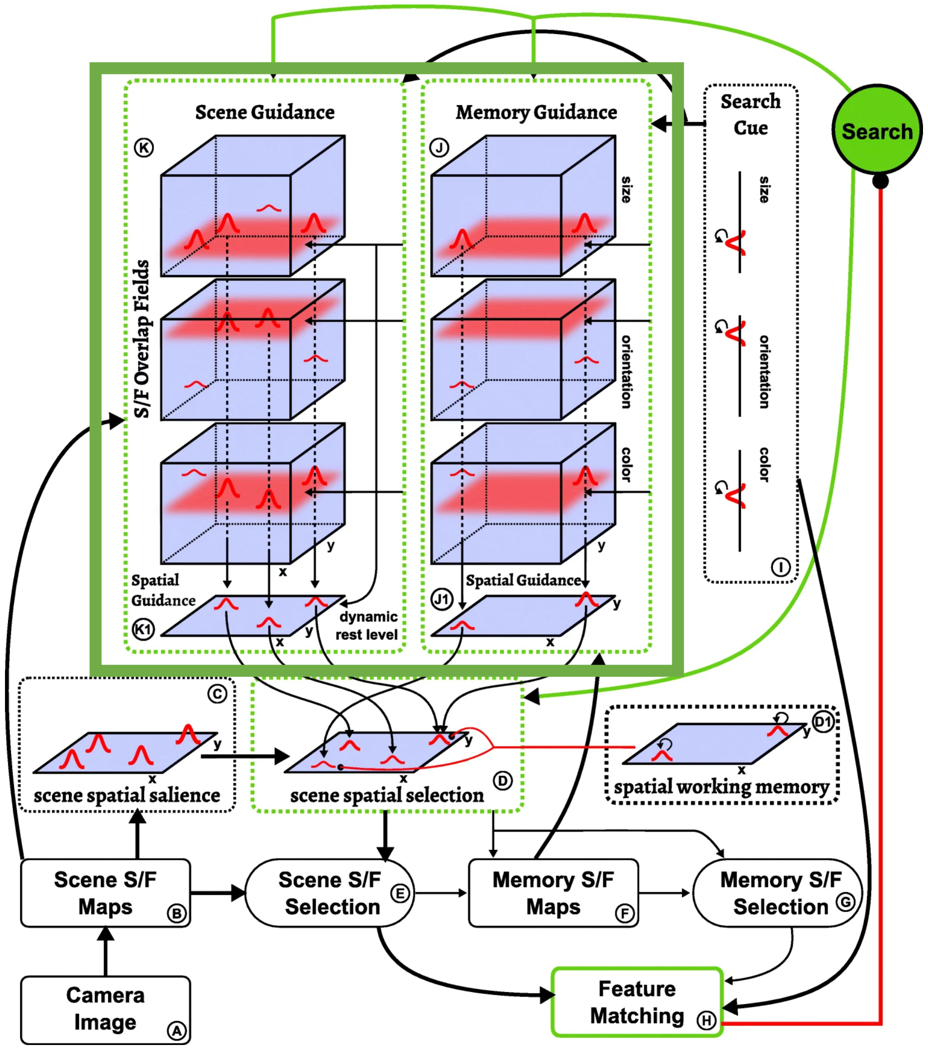
- **Upon deactivation** of the **retain node**, **peaks** in the attention field and the gating fields decay, whereas in the **search cue fields** the cue's feature values **are retained** for later use.

Task 3: Visual search for cued feature conjunctions



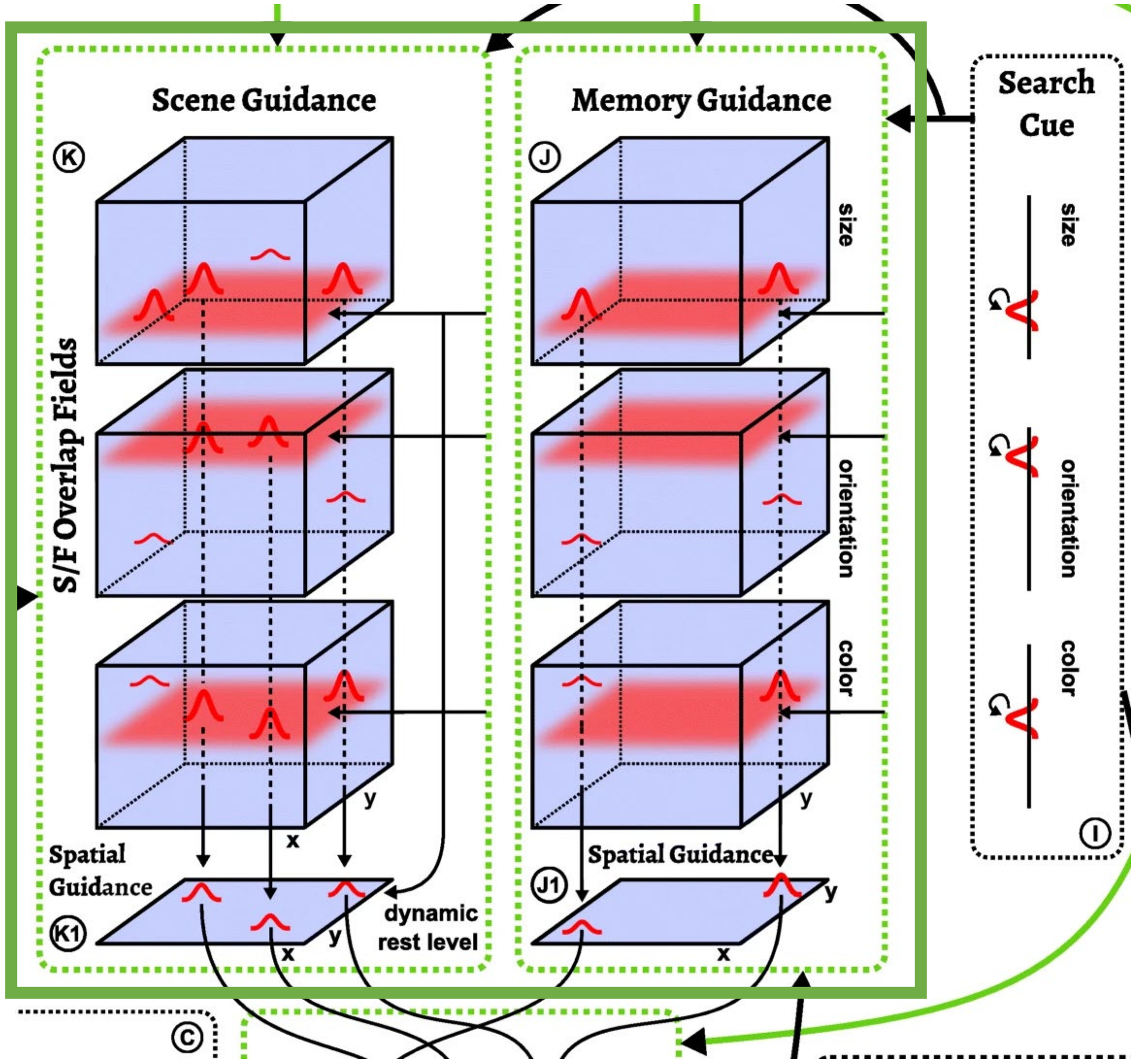
- The **search task node** drives a **sub-network** which **increases** the **likelihood** that **attention** will be **focused** on a location **where** all **features** of the **search cue** are **present**.

Task 3: Visual search for cued feature conjunctions



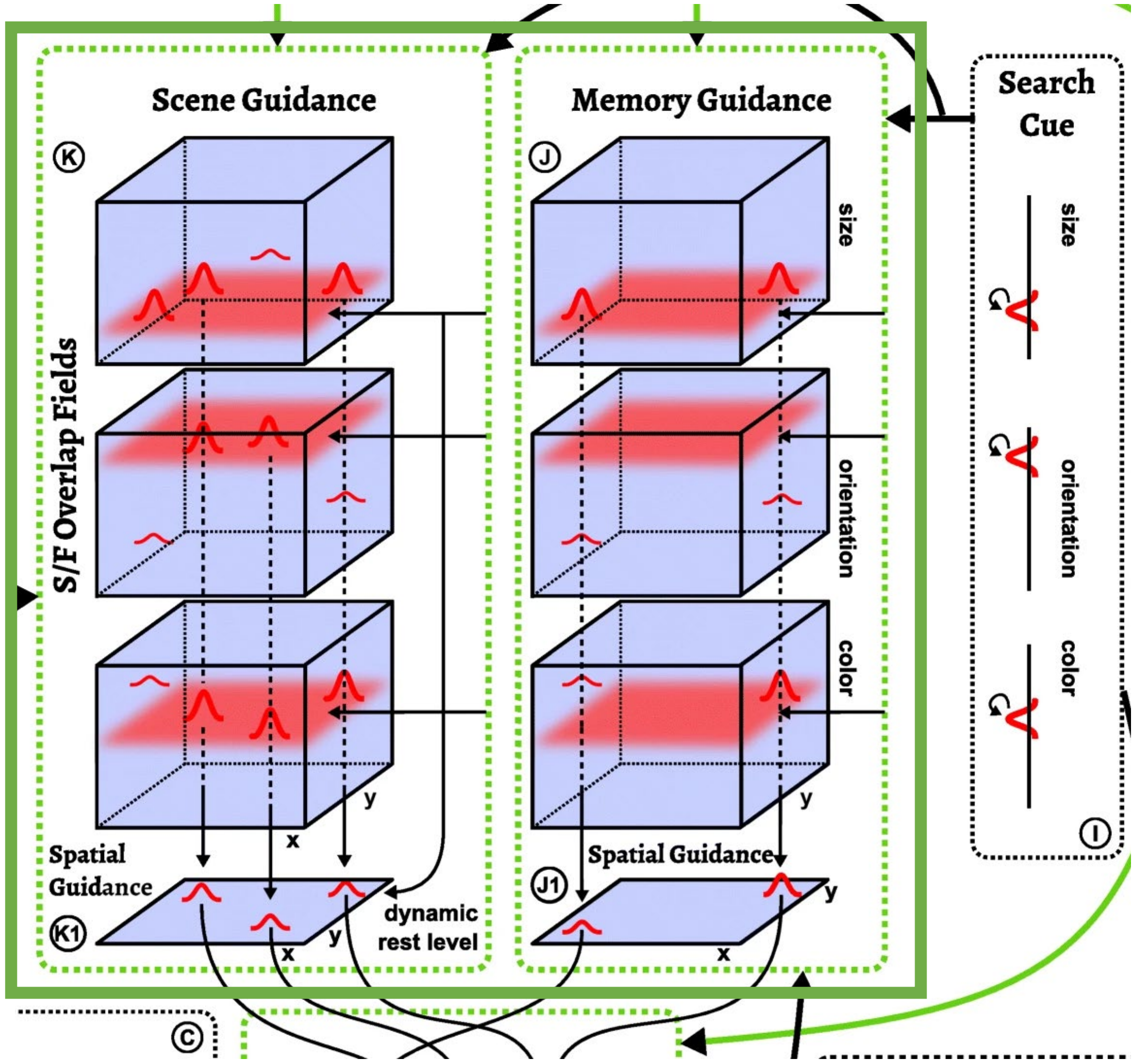
- The search task node drives a sub-network which increases the likelihood that attention will be focused on a location where all features of the search cue are present.
- This is **primarily achieved** through **top-down guidance** from **two sources**, the **visual scene** itself and **scene memory**.

Task 3: Visual search for cued feature conjunctions



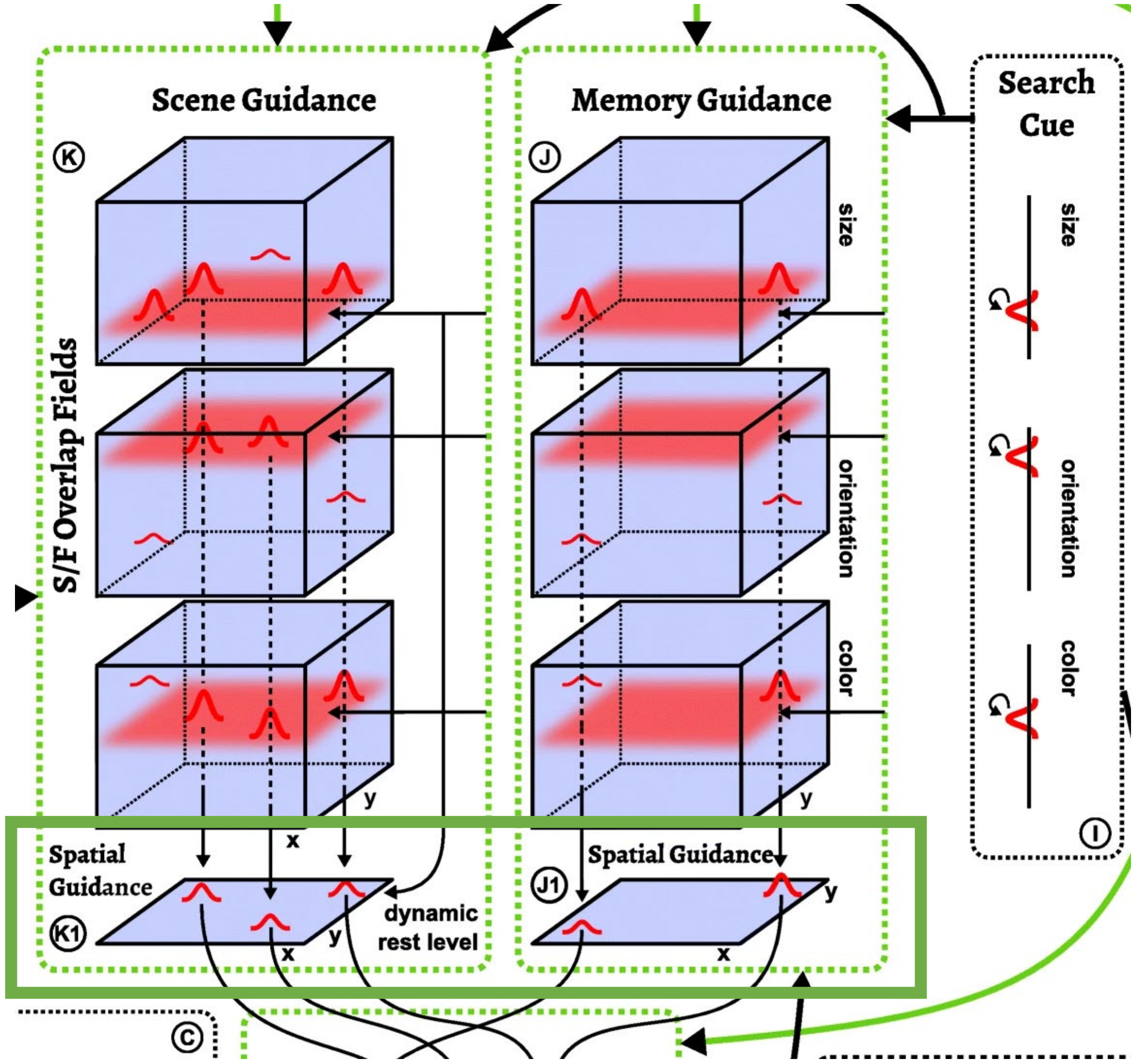
- **Each** of these components **includes** three three-dimensional **space/feature overlap fields** which **combine** sub-threshold **input** from the **scene maps or** the **memory maps** with feature **input** from the **search cue**.

Task 3: Visual search for cued feature conjunctions



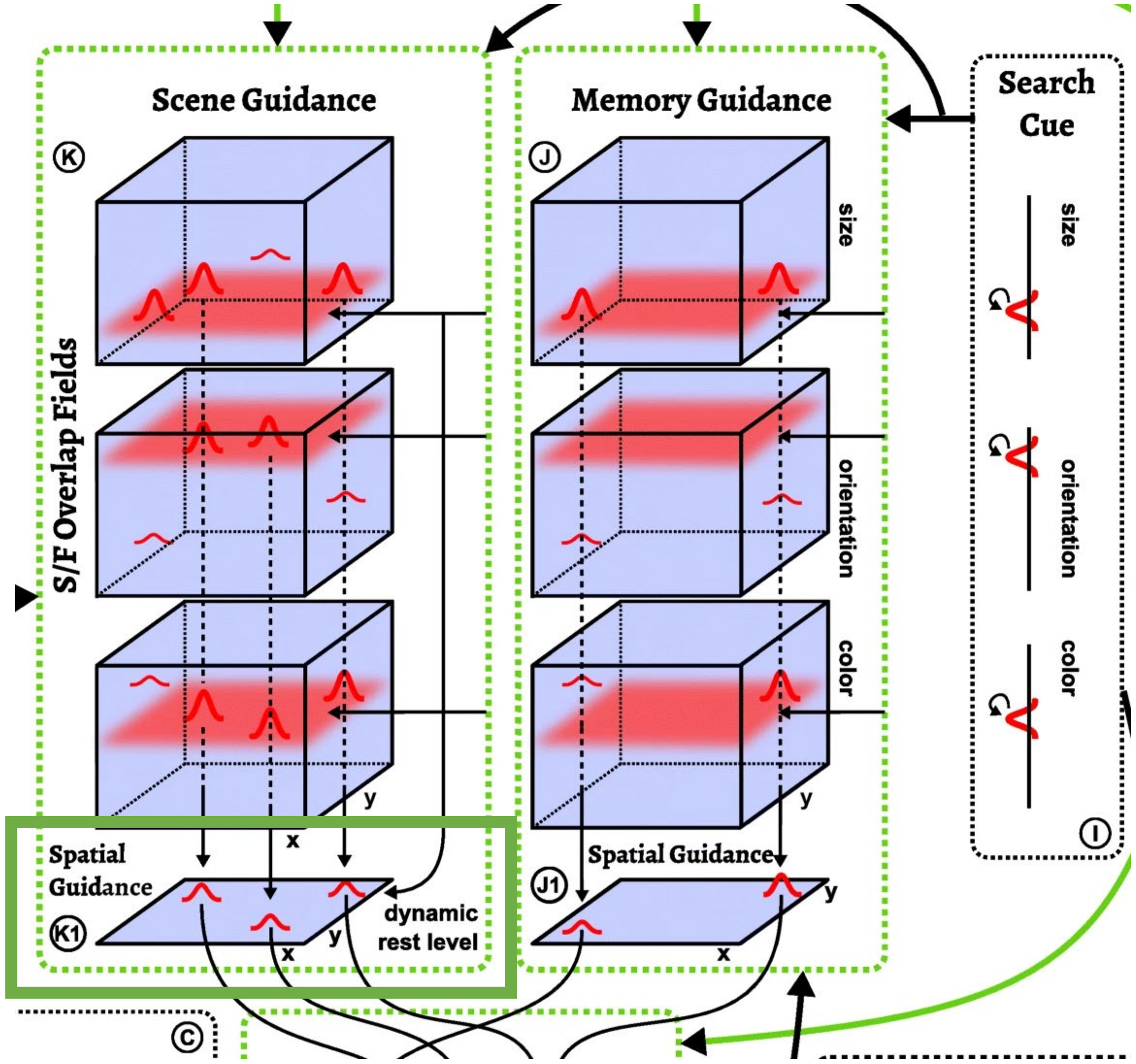
- Each of these components includes three three-dimensional space/feature overlap fields which combine sub-threshold input from the scene maps or the memory maps with feature input from the search cue.
- Supra-threshold **peaks emerge** at **locations** where there is **overlap** between the **cued feature** values and the **scene or memory**.

Task 3: Visual search for cued feature conjunctions



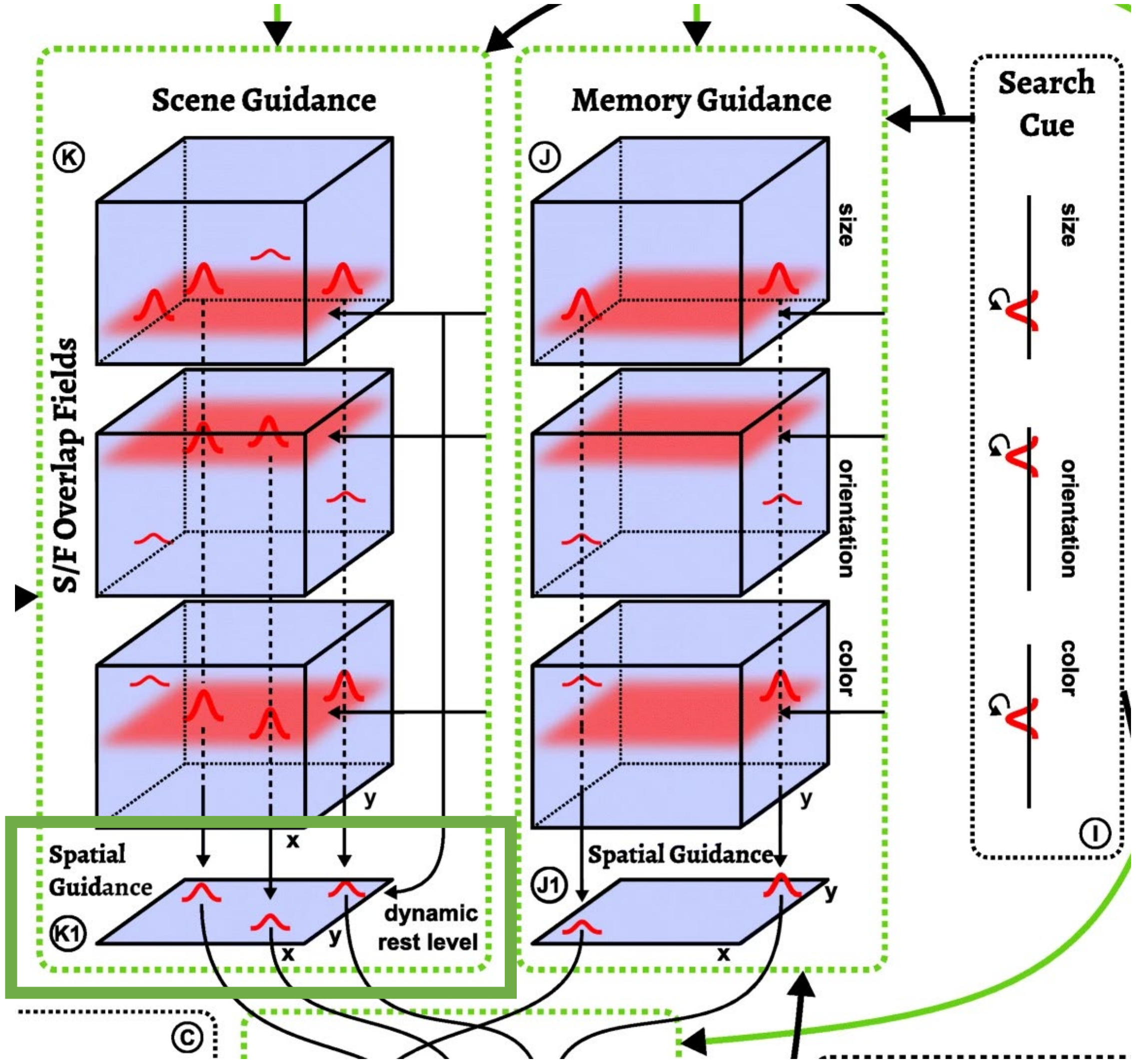
- These **peaks** are **projected** into two-dimensional **spatial guidance fields** which **bias attentional selection** in the **scene spatial selection field**.

Task 3: Visual search for cued feature conjunctions



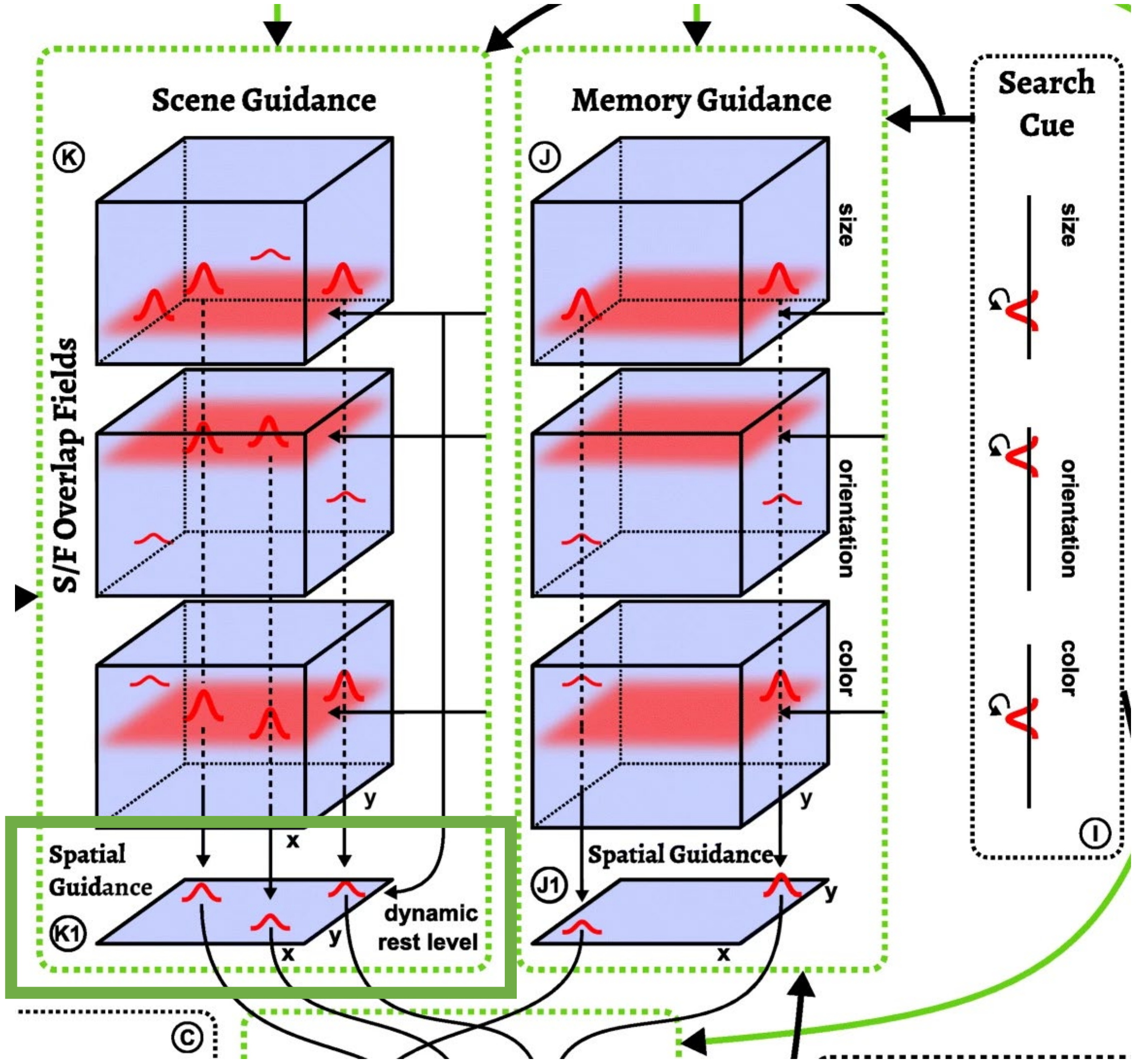
- These peaks are projected into two-dimensional spatial guidance fields which bias attentional selection in the scene spatial selection field.
- **Importantly**, the **resting level** of the **scene spatial guidance field** is **down-regulated dynamically** via **inhibitory connectivity** from each **search cue** field.

Task 3: Visual search for cued feature conjunctions



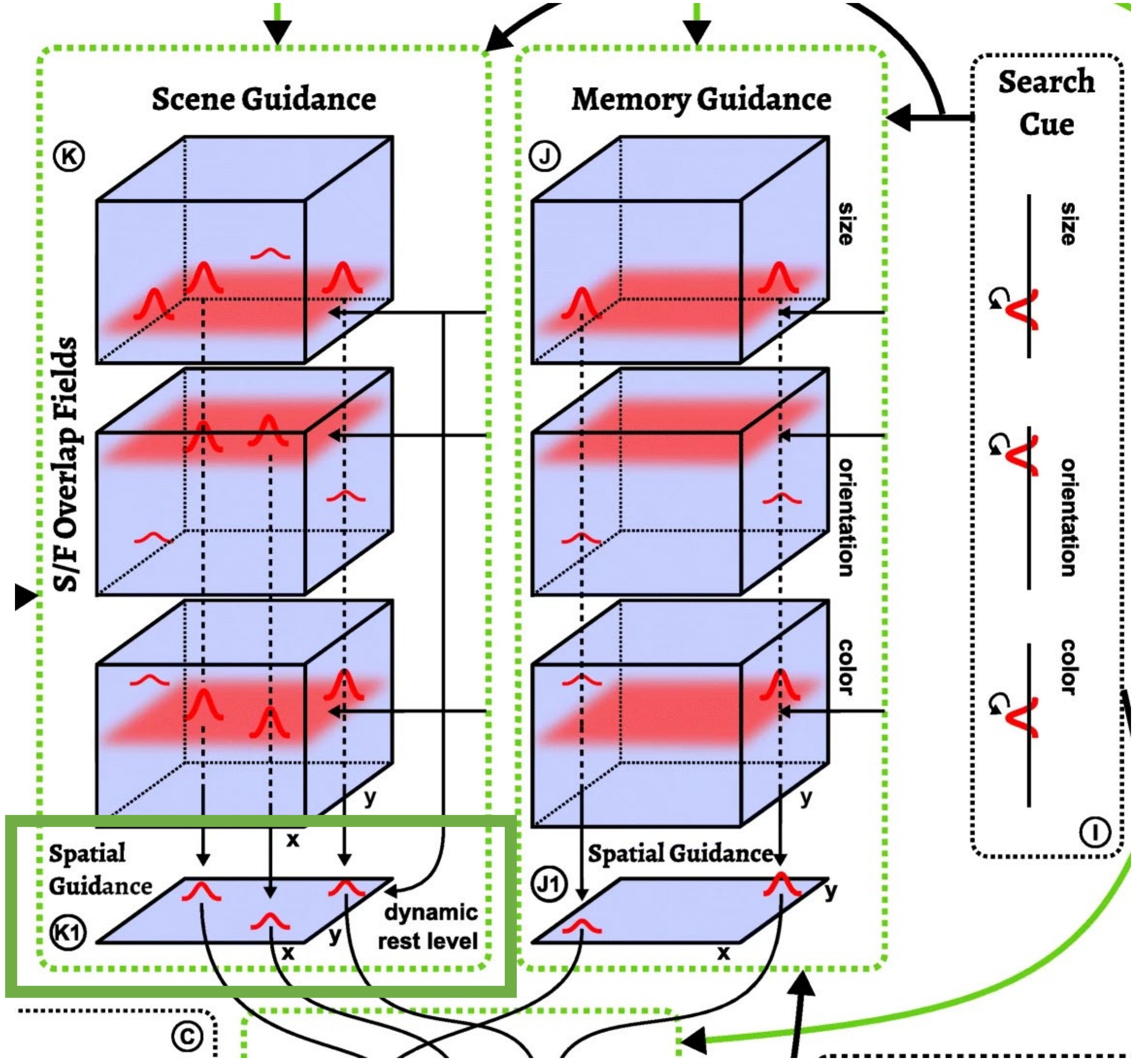
- The **resting level** thus **depends** on the **number** of **cued features**, decreasing as more search cue fields contain peaks.

Task 3: Visual search for cued feature conjunctions



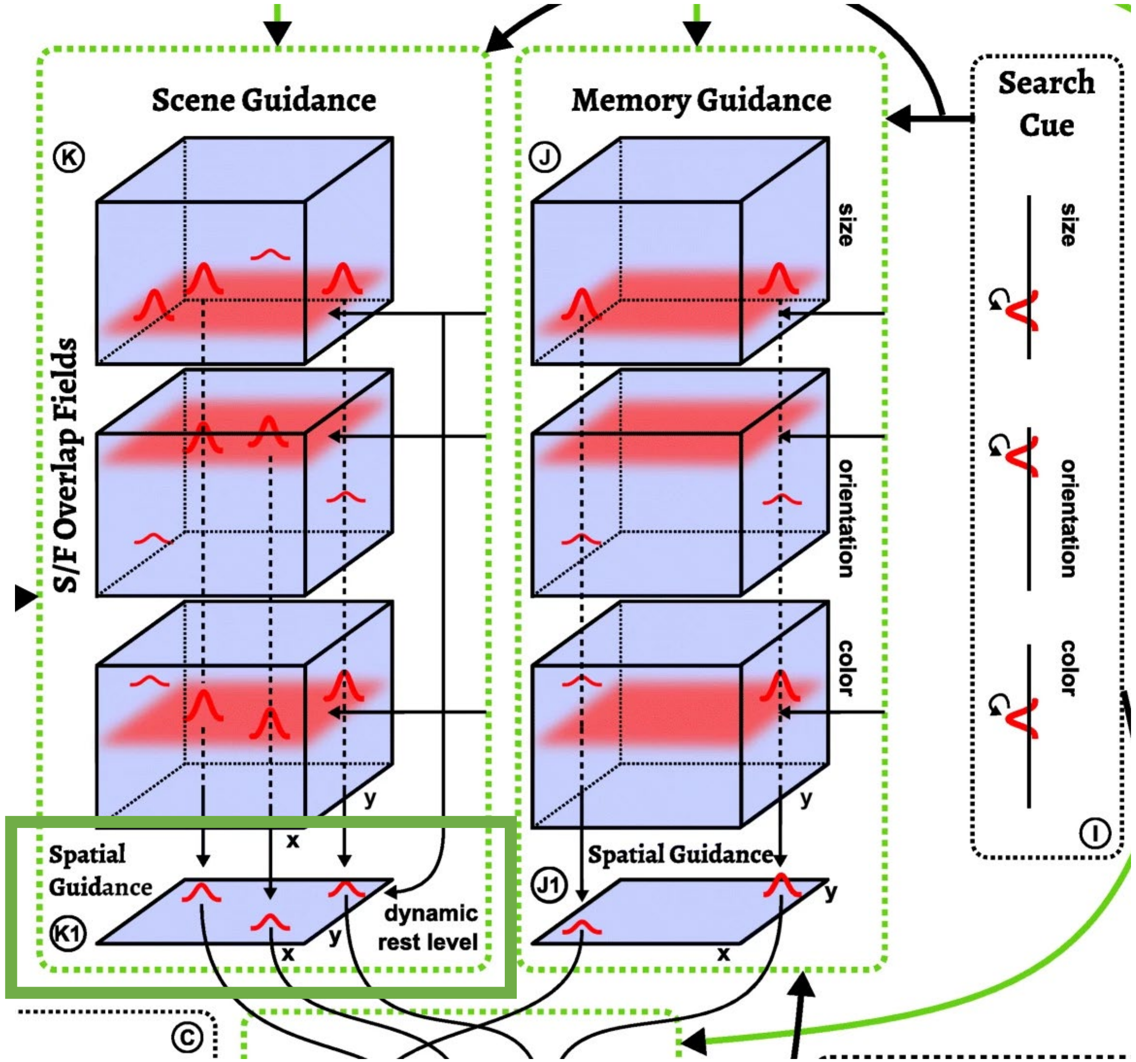
- The **strength** of the **inhibitory connections** is such that when only **one feature is cued** it **suffices** for **items to share only that cue feature in order to create peaks** in the scene spatial guidance field; when **more than one feature are cued**, **peaks emerge** for **all items** that **differ at most** in **one** of the **cued feature dimensions**.

Task 3: Visual search for cued feature conjunctions



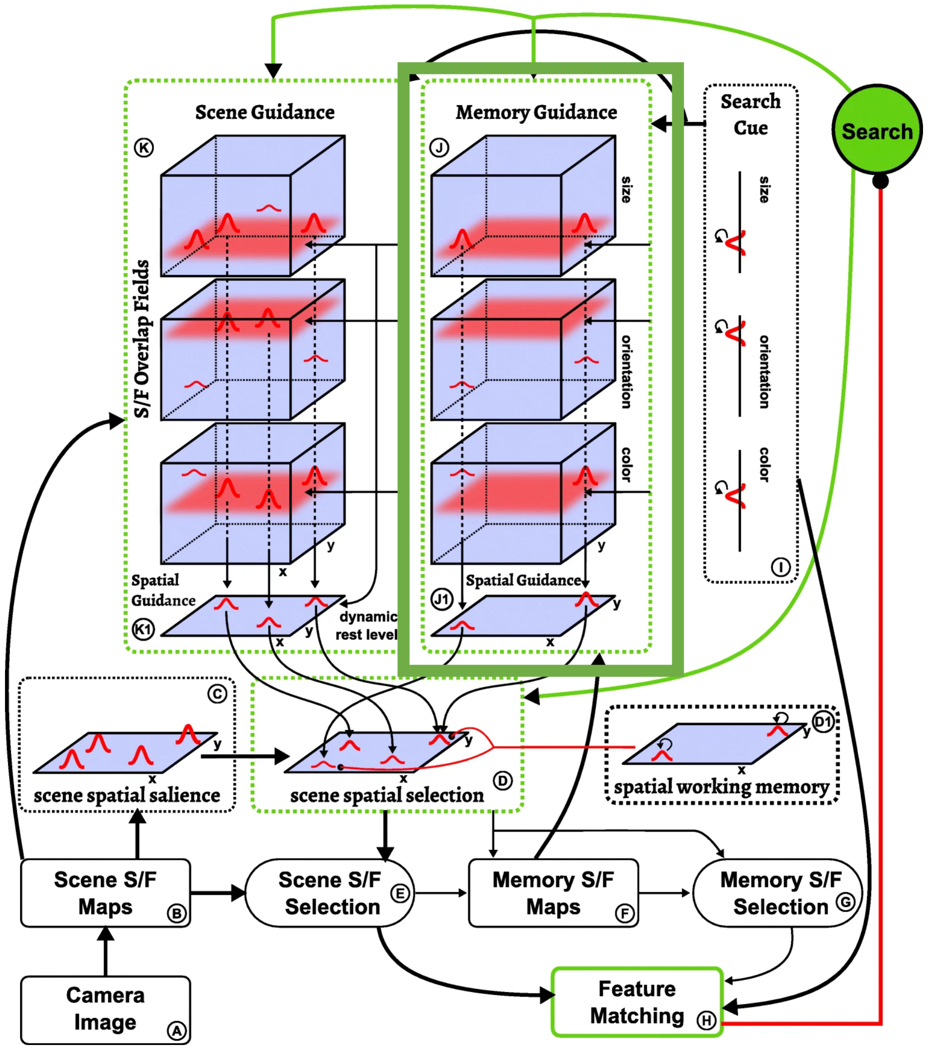
- Therefore, **attentional guidance** is **most effective** in **single feature search**, in which **peaks arise only** for **items** that **completely match** the **cue**.

Task 3: Visual search for cued feature conjunctions



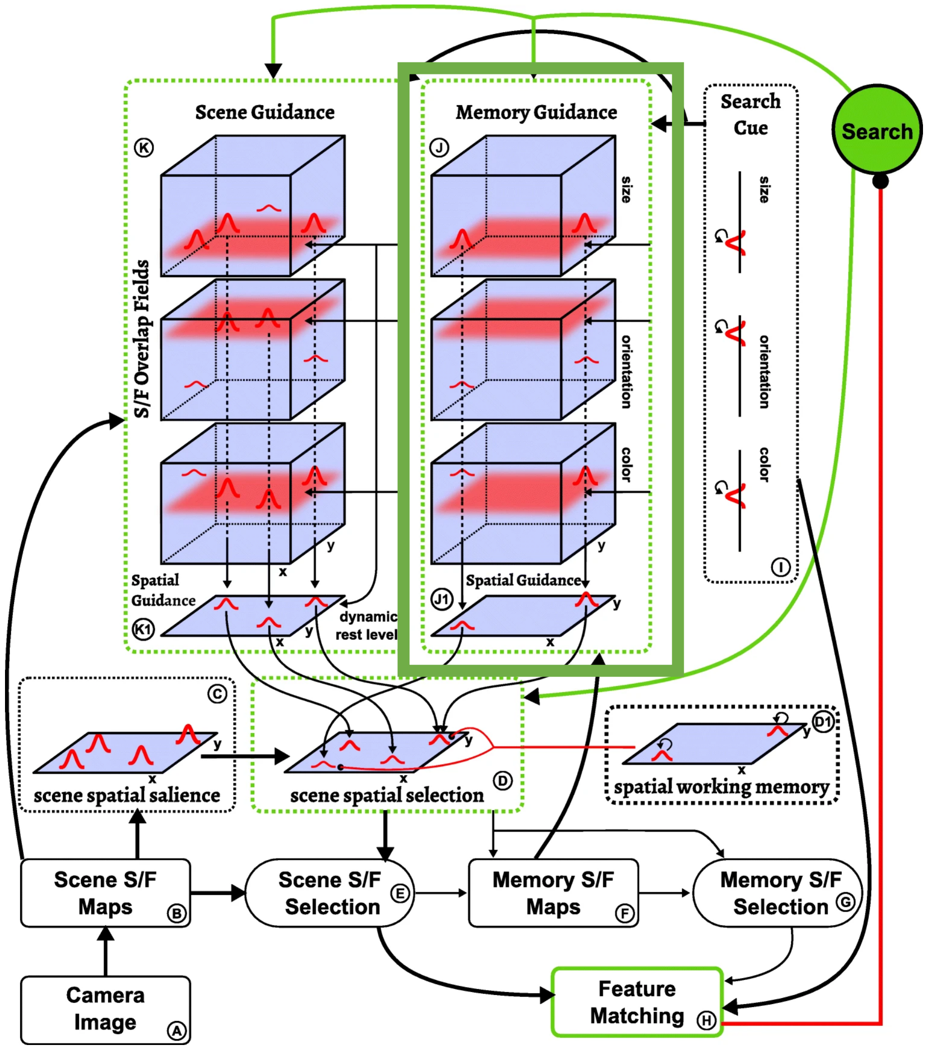
- Therefore, attentional guidance is most effective in single feature search, in which peaks arise only for items that completely match the cue.
- In **conjunctive search**, **non-target items may become active** as well, making conjunctive search **less effective** in this account.

Task 3: Visual search for cued feature conjunctions



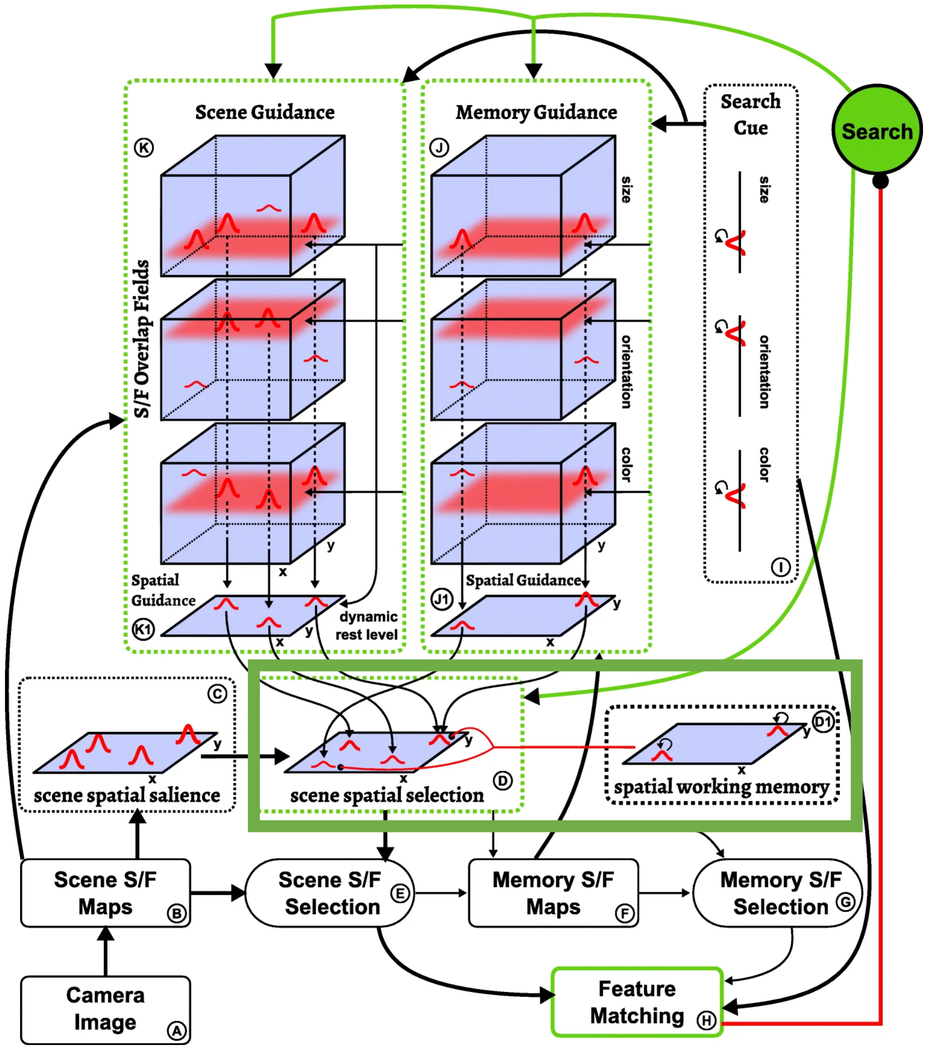
- The **influence** of **memory** on **attentional selection** described **thus far** is **purely excitatory** and based on the overlap of memory items with cue features.

Task 3: Visual search for cued feature conjunctions



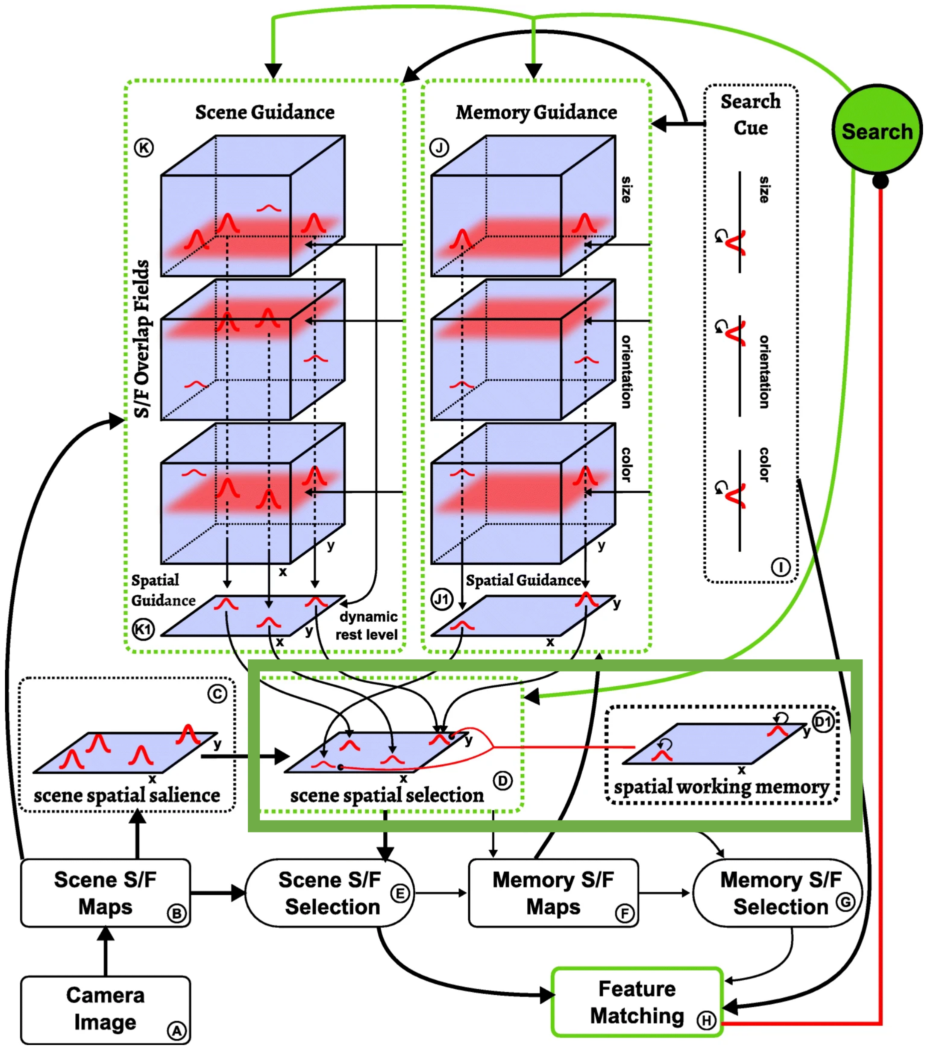
- The influence of memory on attentional selection described thus far is purely excitatory and based on the overlap of memory items with cue features.
- **This excitatory bias from memory explains** the overall **faster reaction times** in the **preview condition** of the **experiment.**

Task 3: Visual search for cued feature conjunctions



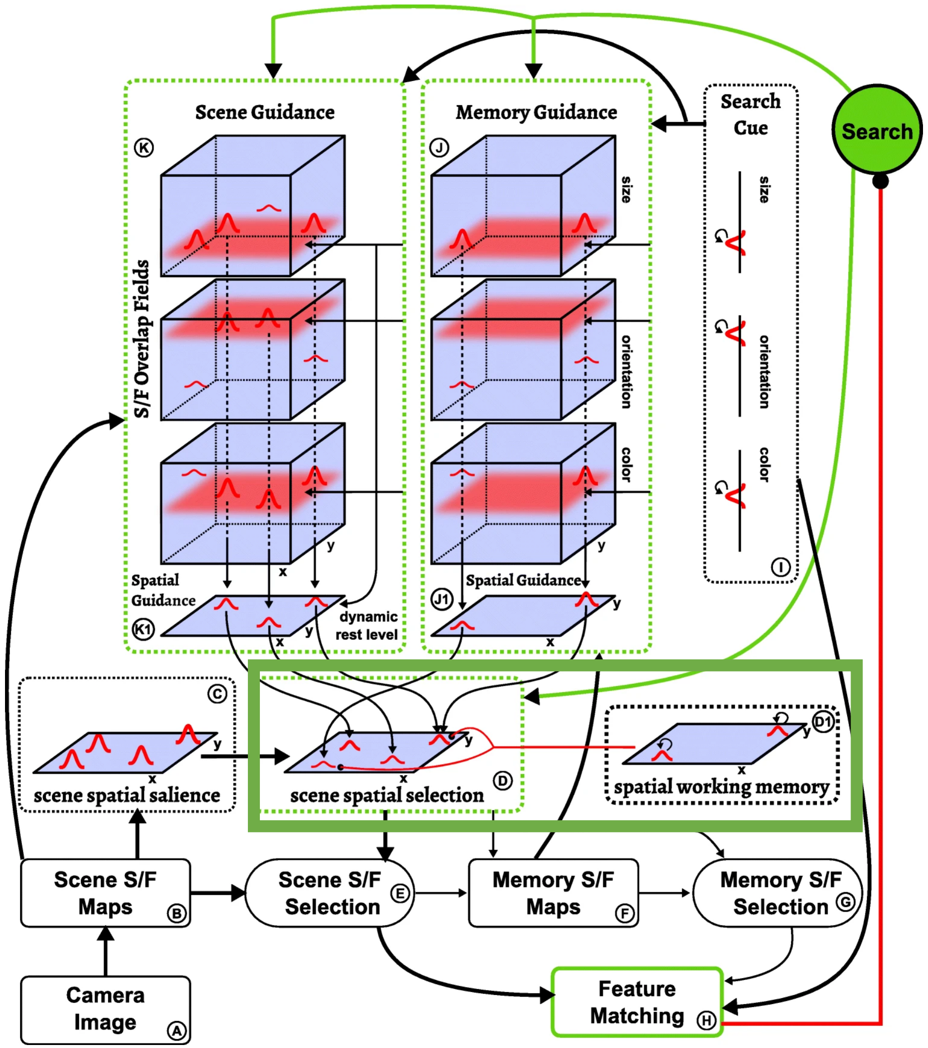
- An **additional**, **inhibitory influence** on attentional **selection** comes from the **spatial working memory** field, that represents **locations** that have been **committed** to **memory** during the **exploration** phase.

Task 3: Visual search for cued feature conjunctions



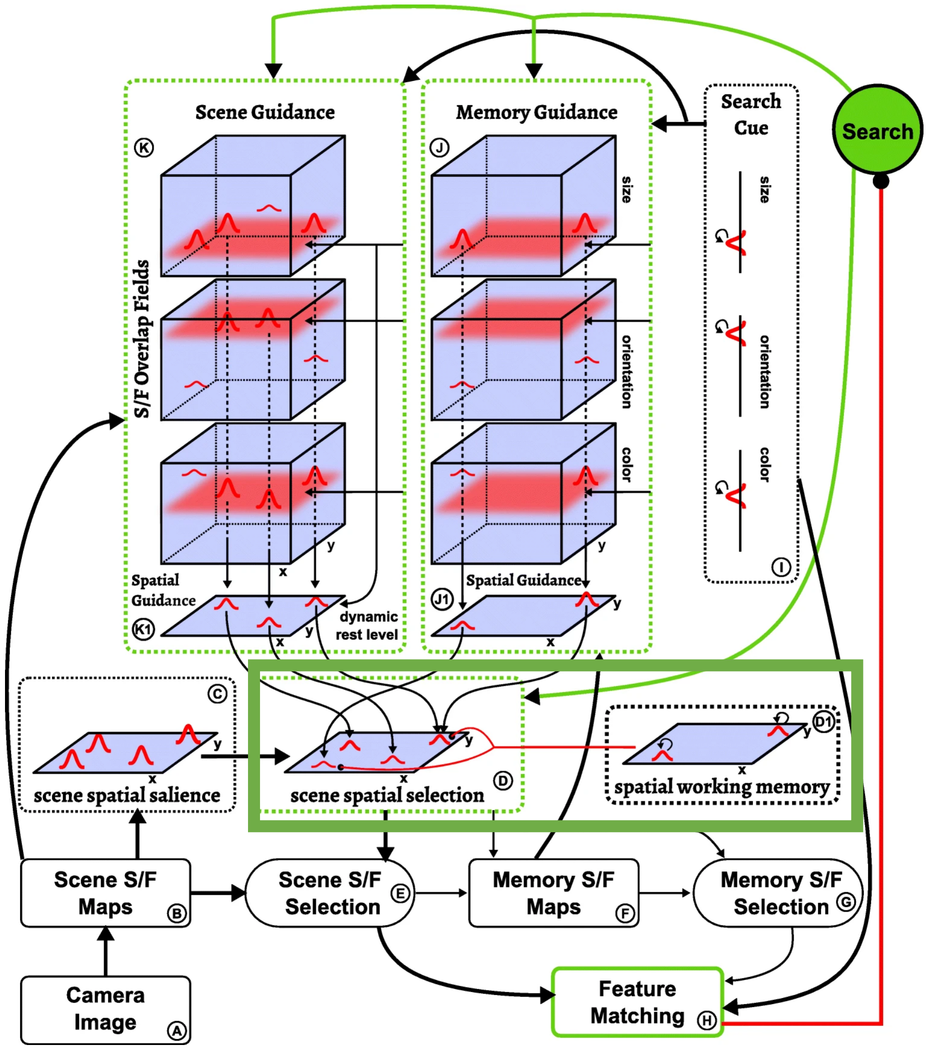
- An additional, inhibitory influence on attentional selection comes from the spatial working memory field, that represents locations that have been committed to memory during the exploration phase.
- **Their influence decreases** the **likelihood** that **attention revisits** such **locations**.

Task 3: Visual search for cued feature conjunctions



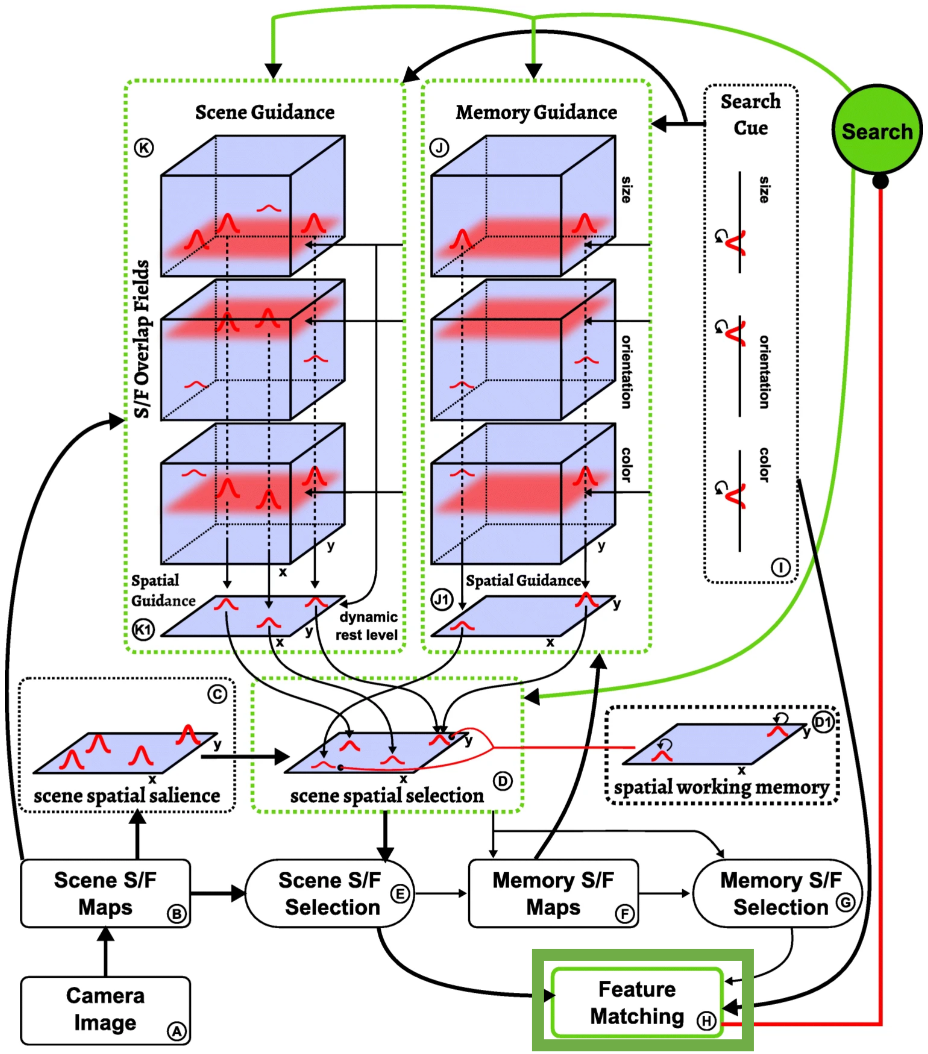
- The **inhibited locations** may **include items** that **match** the visual search **cue**. The **strength** of **inhibition** is **low** enough, however, to be **overruled** by **excitatory biases** from the other sources.

Task 3: Visual search for cued feature conjunctions



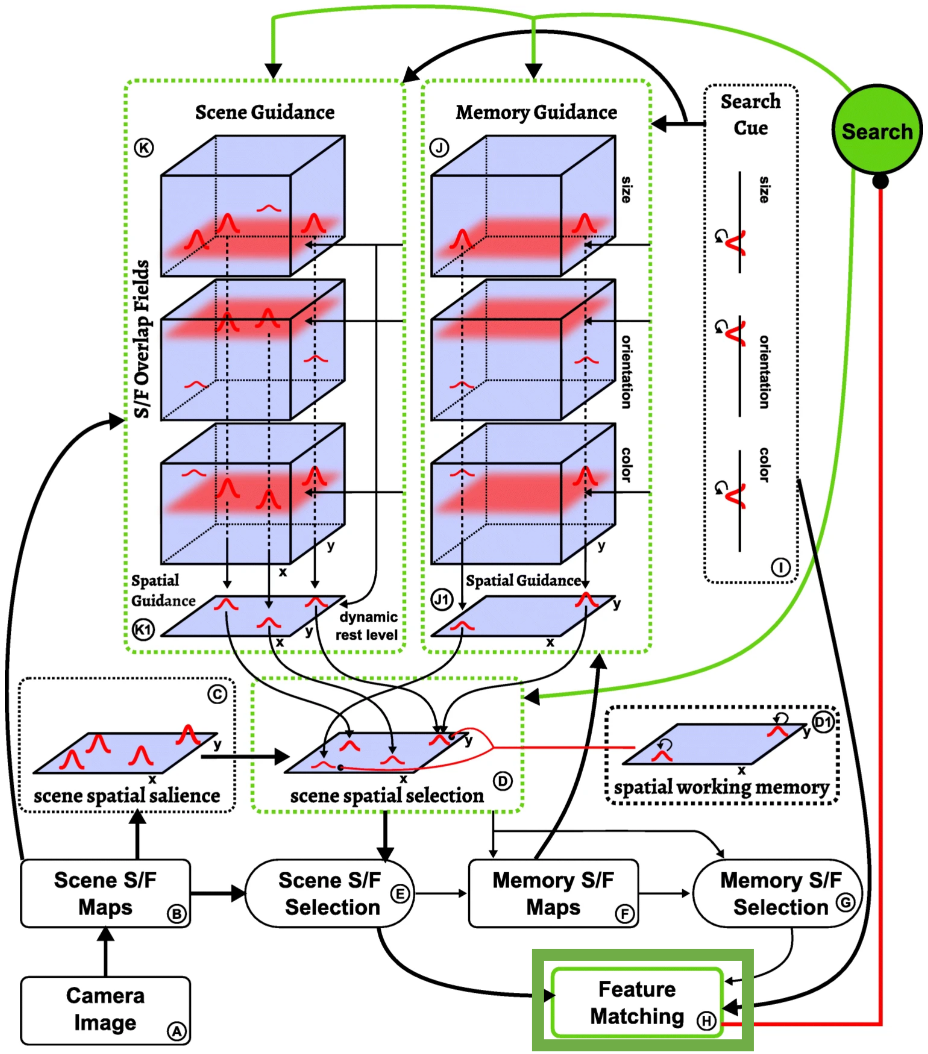
- The inhibited locations may include items that match the visual search cue. The strength of inhibition is low enough, however, to be overruled by excitatory biases from the other sources.
- This **inhibitory bias** from **spatial memory explains** the **increased efficiency** in the **preview condition** of the **experiment**.

Task 3: Visual search for cued feature conjunctions



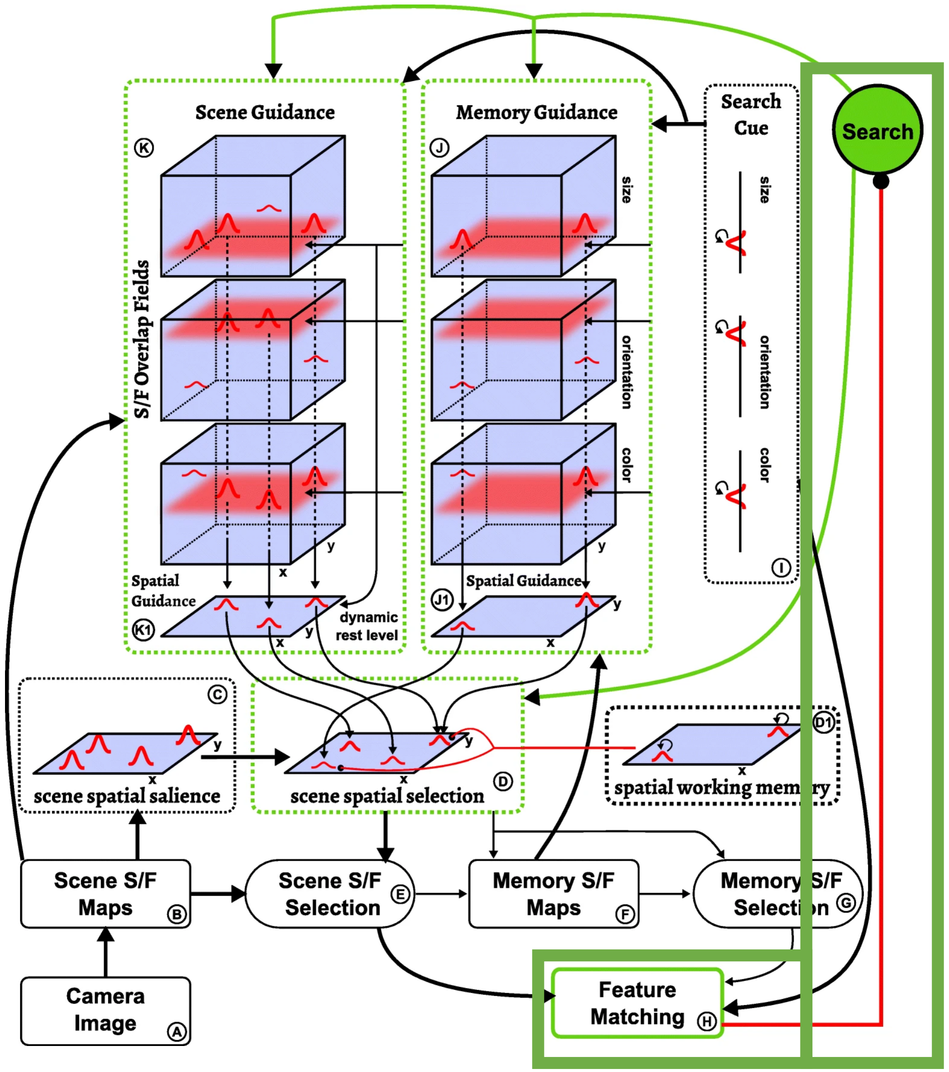
- The **visual search** process is **terminated** when the **features** at an **attended location match all** specified **cue features**.

Task 3: Visual search for cued feature conjunctions



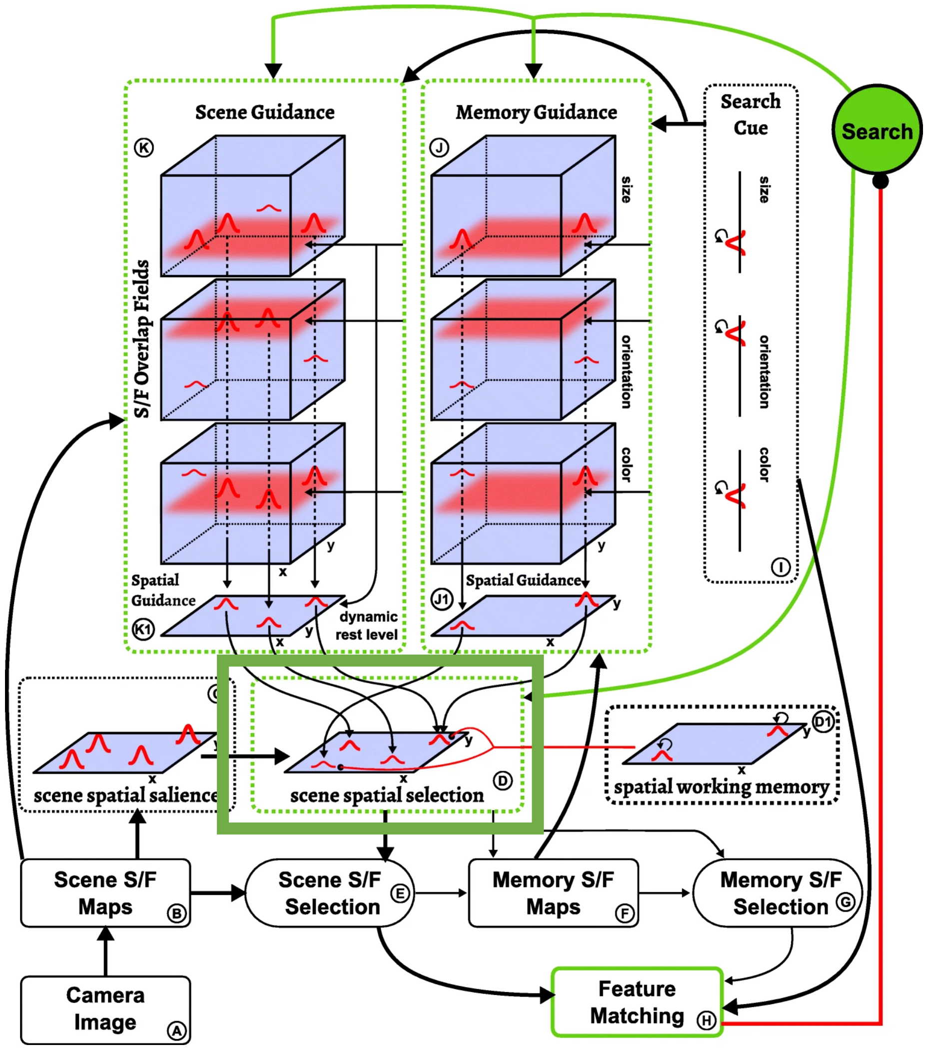
- The visual search process is terminated when the features at an attended location match all specified cue features.
- This is **detected** by **the feature matching component**, whose **CoS node activates** when such a **match occurs**, which **signals task completion**.

Task 3: Visual search for cued feature conjunctions



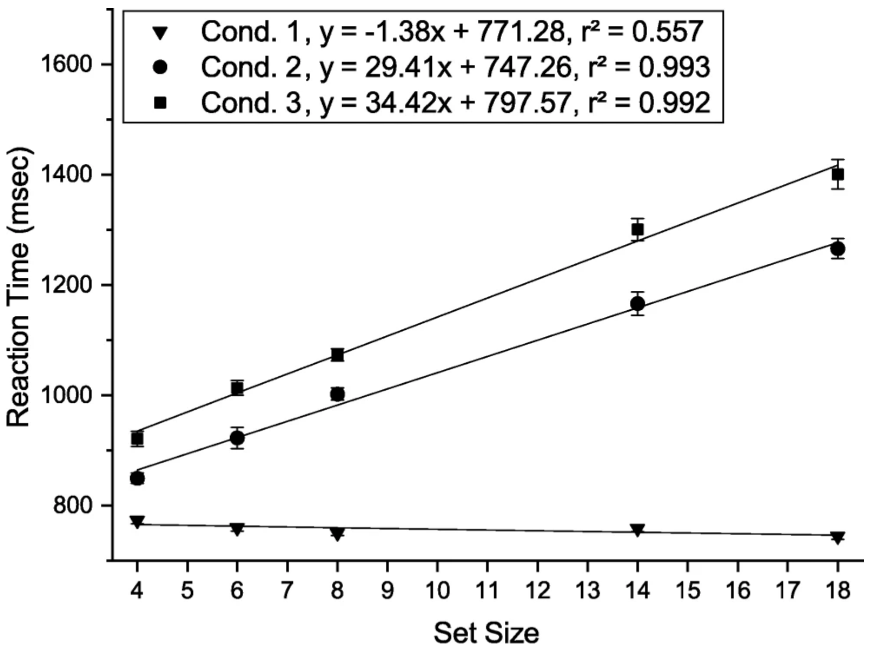
- If **instead one or more cued feature values** are **not present** in the **attended location**, the **CoD node** of the **feature matching component** becomes **active** and **inhibits** the **search task node**.

Task 3: Visual search for cued feature conjunctions

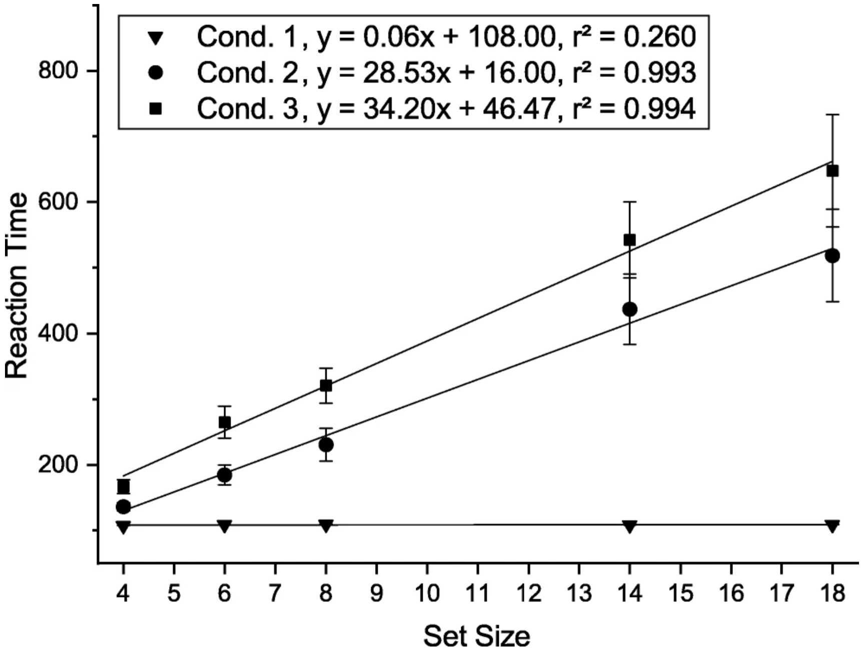


- This **destabilizes** the **scene spatial selection field**, which in turn **leads** to the **CoD itself** being **deactivated**, so that the **search task node** can **reactivate** and **drive** the attentional **selection** of a **new location**.

Model - Results



Experiment



Model

Grieben et al. Scene memory and spatial inhibition in visual search. Atten Percept Psychophys (2020)

Extension: Understanding the interplay between bottom-up processing and top-down guidance in visual search

Bottom-Up and Top-Down Attention

- The **capacity** of the brain to process sensory stimuli is **limited**

Bottom-Up and Top-Down Attention

- The capacity of the brain to process sensory stimuli is limited
- Neural **resources** are **focused** according to the current **contingencies**

Bottom-Up and Top-Down Attention

- The capacity of the brain to process sensory stimuli is limited
- Neural resources are focused according to the current contingencies
- This **cognitive process** is called **attention**

Bottom-Up and Top-Down Attention

- **Attention** can be **categorized** into two distinct functions

Bottom-Up and Top-Down Attention

- Attention can be categorized into two distinct functions

Bottom-up attention

Top-down attention

Bottom-Up and Top-Down Attention

- Attention can be categorized into two distinct functions

Bottom-up attention

- Attentional **guidance** driven purely by **external** factors

Bottom-Up and Top-Down Attention

- Attention can be categorized into two distinct functions

Bottom-up attention

- Attentional guidance driven purely by external factors
- **Saliency** of stimuli **depend** on their **inherent properties** relative to the background

Bottom-Up and Top-Down Attention

- Attention can be categorized into two distinct functions

Bottom-up attention

- Attentional guidance driven purely by external factors
- Saliency of stimuli depend on their inherent properties relative to the background
- E.g., **local** feature **contrasts** like red/green or sudden movement

Bottom-Up and Top-Down Attention

- Attention can be categorized into two distinct functions

Bottom-up attention

- Attentional guidance driven purely by external factors
- Saliency of stimuli depend on their inherent properties relative to the background
- E.g., local feature contrasts like red/green or sudden movement
- Is the phylogenetically **older system**

Bottom-Up and Top-Down Attention

- Attention can be categorized into two distinct functions

Top-down attention

Bottom-Up and Top-Down Attention

- Attention can be categorized into two distinct functions

Top-down attention

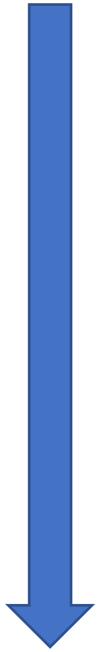
- Attentional **guidance** driven by **internal** factors

Bottom-Up and Top-Down Attention

- Attention can be categorized into two distinct functions

Top-down attention

- Attentional guidance driven by internal factors
- Like prior **knowledge**, current **task** or goal, etc…

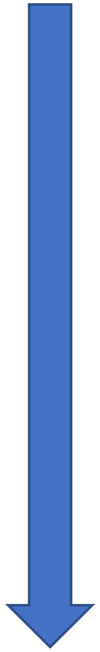


Bottom-Up and Top-Down Attention

- Attention can be categorized into two distinct functions

Top-down attention

- Attentional guidance driven by internal factors
- Like prior knowledge, current task or goal, etc…
- **Guidance** of **visual search**: e.g. the location of a known object is unknown in the current scene



The Binding Problem

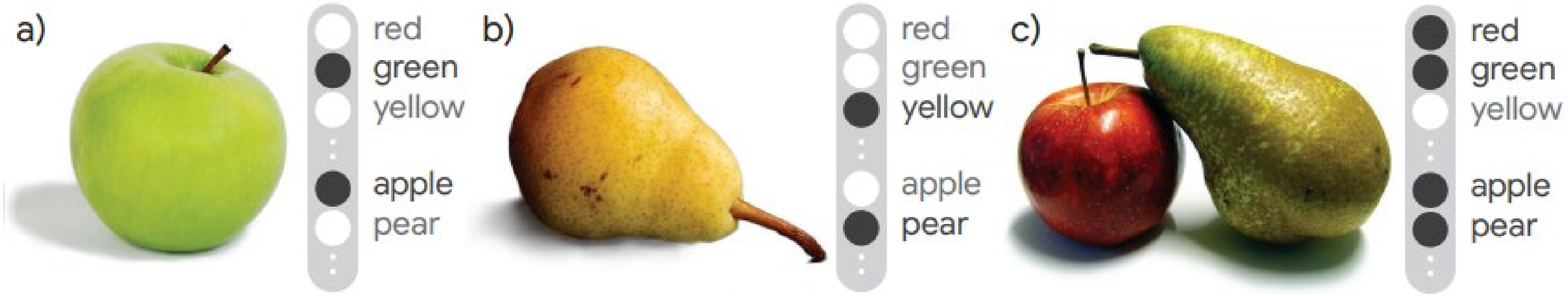
- Different attributes (**features**) of a stimulus (e.g., color, size, orientation) are **processed** by **different** areas of the **cortex**

The Binding Problem

- Different attributes (features) of a stimulus (e.g., color, size, orientation) are processed by different areas of the cortex
- Yet, they are **experienced** (in consciousness) as a **unity** (object)

The Binding Problem

- Different attributes (features) of a stimulus (e.g., color, size, orientation) are processed by different areas of the cortex
- Yet, they are experienced (in consciousness) as a unity (object)
- Artificial **neural networks** currently **ignore** this problem
  - *=> superposition catastrophe* (von der Malsburg, 1999)



The Binding Problem

- Different attributes (features) of a stimulus (e.g., color, size, orientation) are processed by different areas of the cortex
- Yet, they are experienced (in consciousness) as a unity (object)
- Artificial neural networks currently ignore this problem
  - *=> superposition catastrophe* (von der Malsburg, 1999)
- Yet, **binding** is highly **relevant** for correct **knowledge** representation

The Binding Problem

- Different attributes (features) of a stimulus (e.g., color, size, orientation) are processed by different areas of the cortex
- Yet, they are experienced (in consciousness) as a unity (object)
- Artificial neural networks ignore this problem
  - *=> superposition catastrophe* (von der Malsburg, 1999)
- Yet, binding is highly relevant for correct knowledge representation
- It is **unknown** how the **brain** correctly links up all the different features of complex objects

Does visual attention select objects or locations?

- The effects associated with **location**-based **attention** tend to be **large** and are found **consistently** across experiments

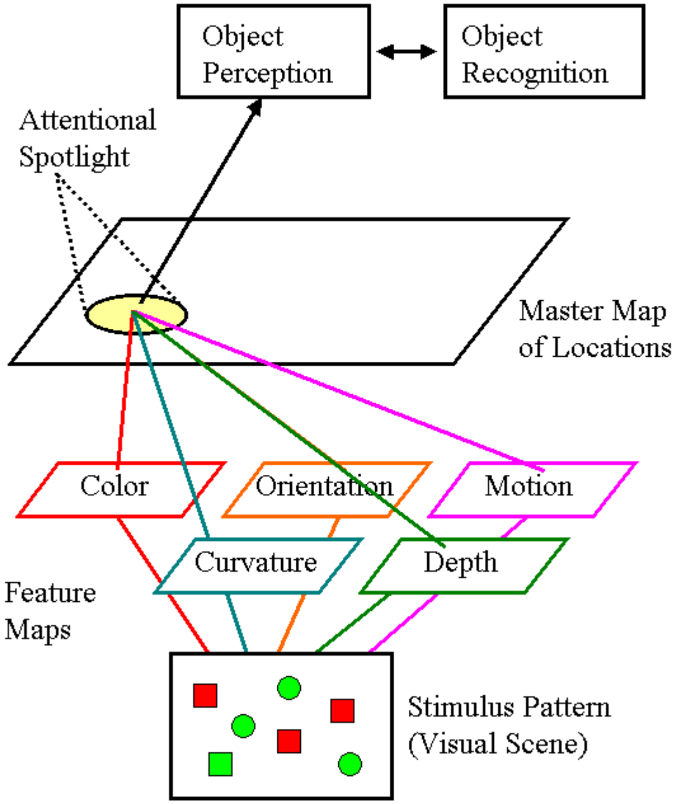
Does visual attention select objects or locations?

- The effects associated with location-based attention tend to be large and are found consistently across experiments
  - This **favors binding** through **attentional selection** of a location

Does visual attention select objects or locations?

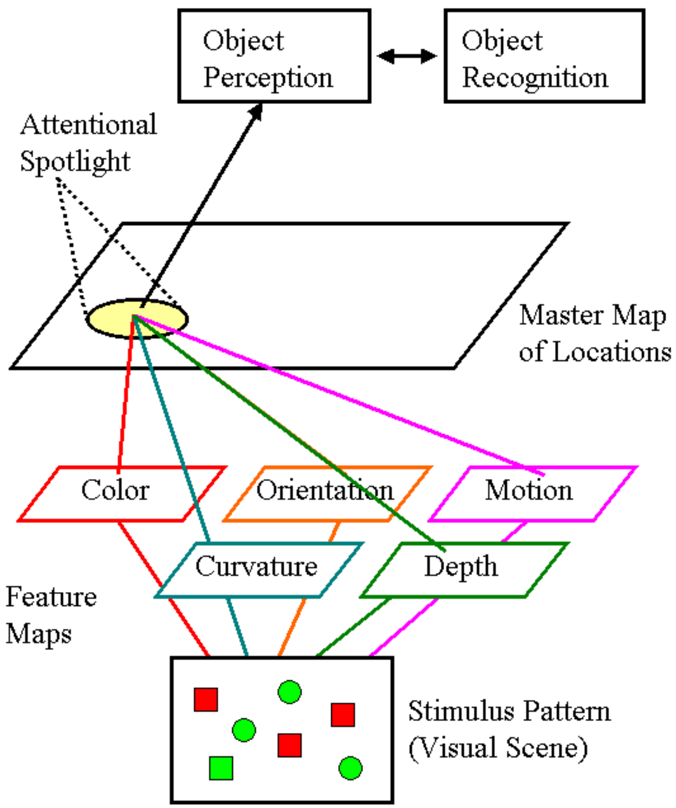
- The effects associated with location-based attention tend to be large and are found consistently across experiments
  - This favors binding through attentional selection of a location
  - **Feature integration theory** (Treisman & Gelande, 1980) is the **prevalent** theory

Feature Integration Theory (FIT)



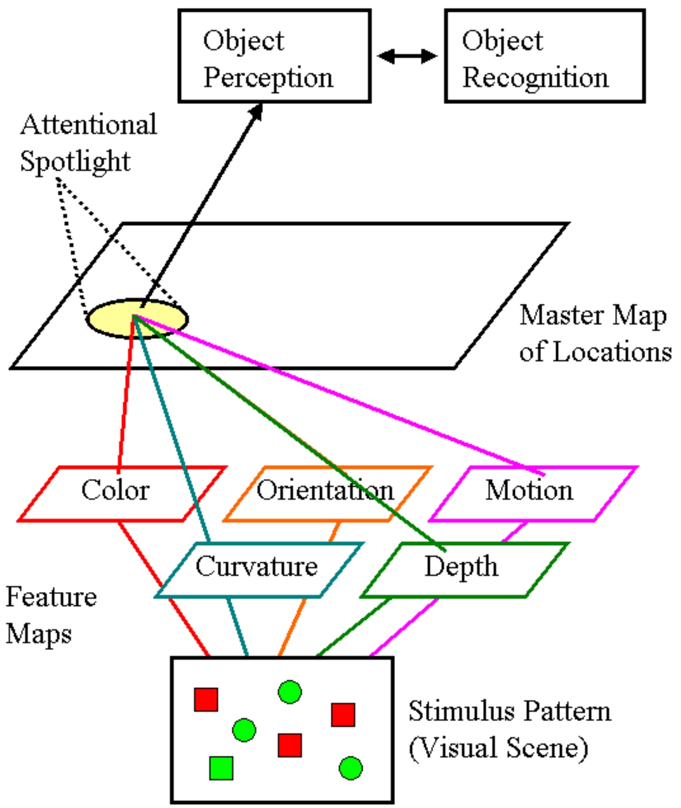
- The most **influential** psychological **model** of human visual **attention**

Feature Integration Theory (FIT)



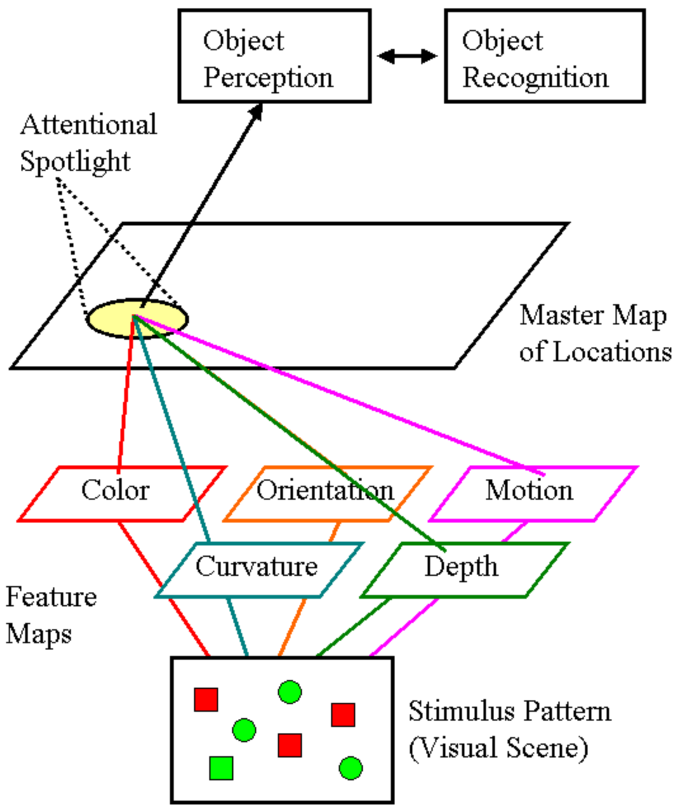
- The most influential psychological model of human visual attention
- Developed in **1980** by Anne **Treisman** and Garry **Gelade**

Feature Integration Theory (FIT)



- The most influential psychological model of human visual attention
- Developed in 1980 by Anne Treisman and Garry Gelade
- **Features** are extracted in **parallel** in a **preattentive** stage

Feature Integration Theory (FIT)



- The most influential psychological model of human visual attention
- Developed in 1980 by Anne Treisman and Garry Gelade
- Features are extracted in parallel in a preattentive stage
- **Objects** and their features are **bound** by **sequentially** attentional selection (attentional bottleneck)

Does visual attention select objects or locations?

- The effects associated with location-based attention tend to be large and are found consistently across experiments
  - This favors binding through attentional selection of a location
  - Feature integration theory (Treisman & Gelande, 1980) is the prevalent theory
- **Object**-based **attention** effects, however, are **small** and found **less consistently** across experiments

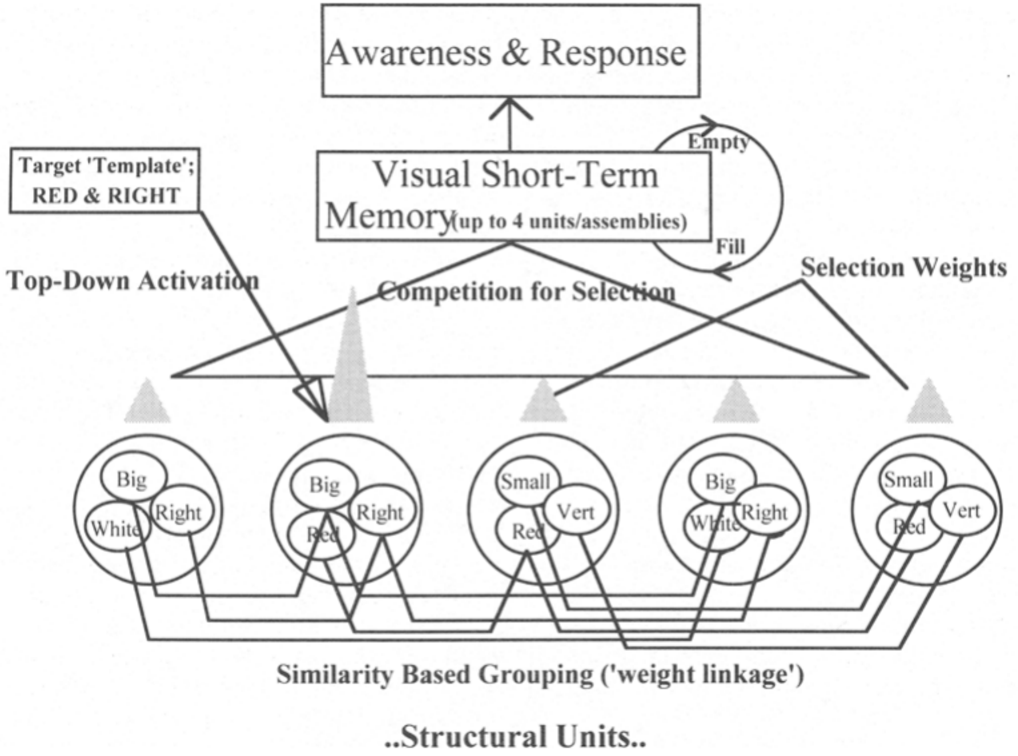
Does visual attention select objects or locations?

- The effects associated with location-based attention tend to be large and are found consistently across experiments
  - This favors binding through attentional selection of a location
  - Feature integration theory (Treisman & Gelande, 1980) is the prevalent theory
- Object-based attention effects, however, are small and found less consistently across experiments
  - This is seen as **evidence** for **binding without attention**

Does visual attention select objects or locations?

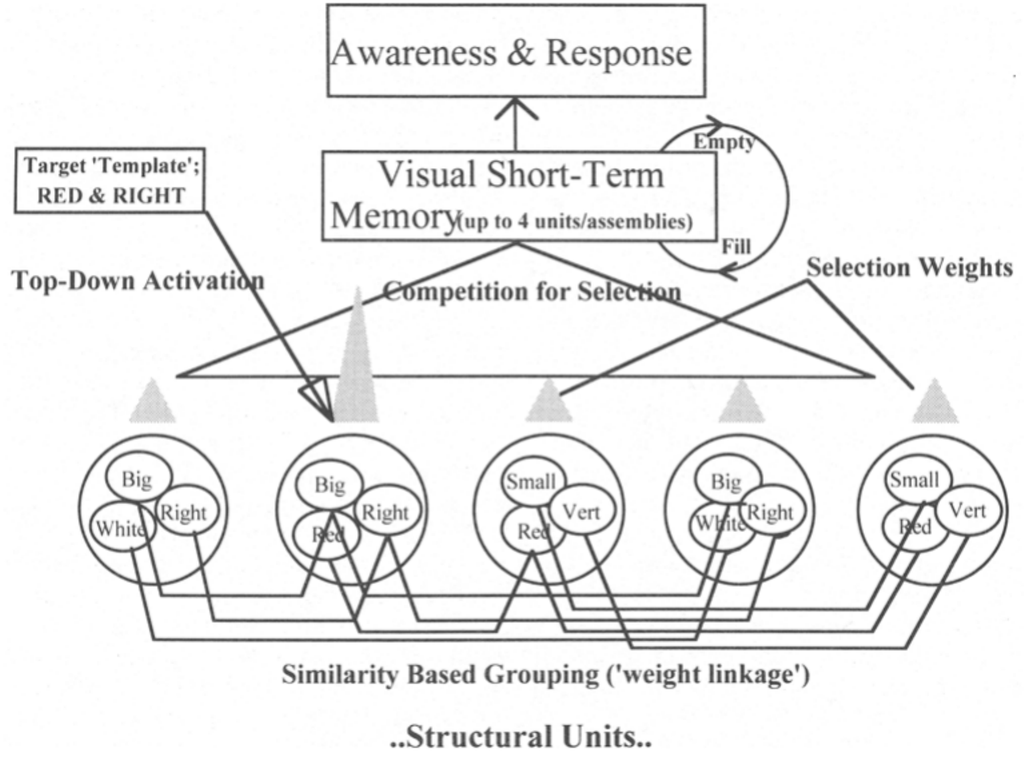
- The effects associated with location-based attention tend to be large and are found consistently across experiments
  - This favors binding through attentional selection of a location
  - Feature integration theory (Treisman & Gelande, 1980) is the prevalent theory
- Object-based attention effects, however, are small and found less consistently across experiments
  - This is seen as evidence for binding without attention
  - As **postulated** by **similarity theory** (Duncan & Humphreys, 1989)

Similarity Theory of Attention



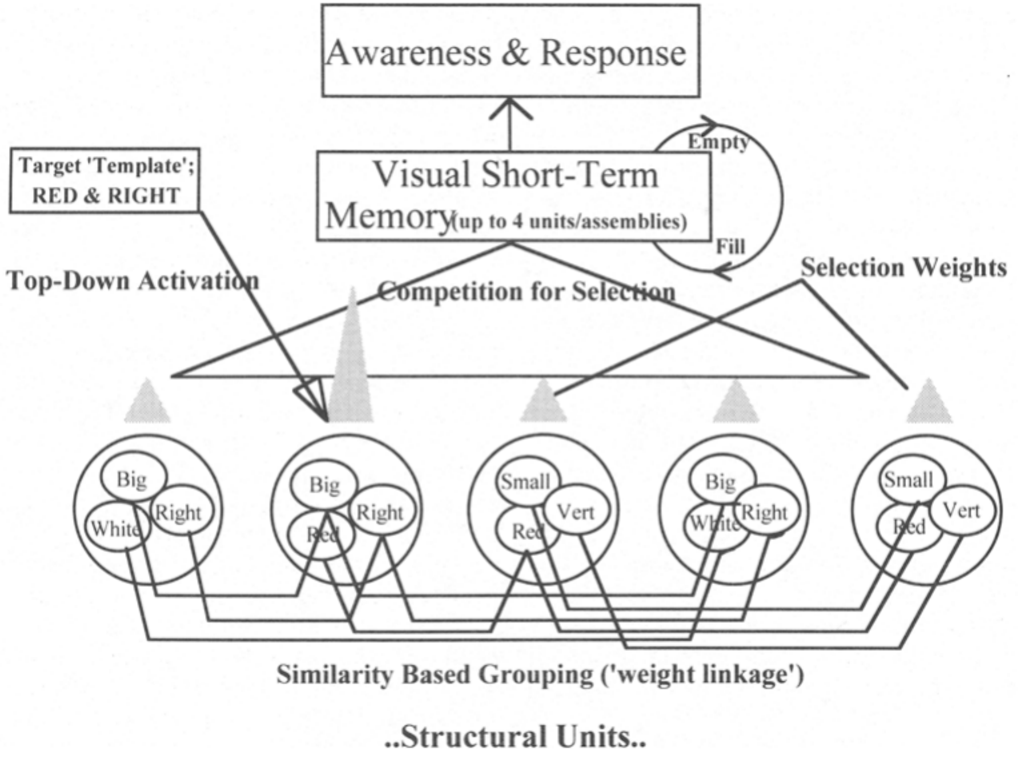
- **Duncan** and **Humphreys** (**1989**)

Similarity Theory of Attention



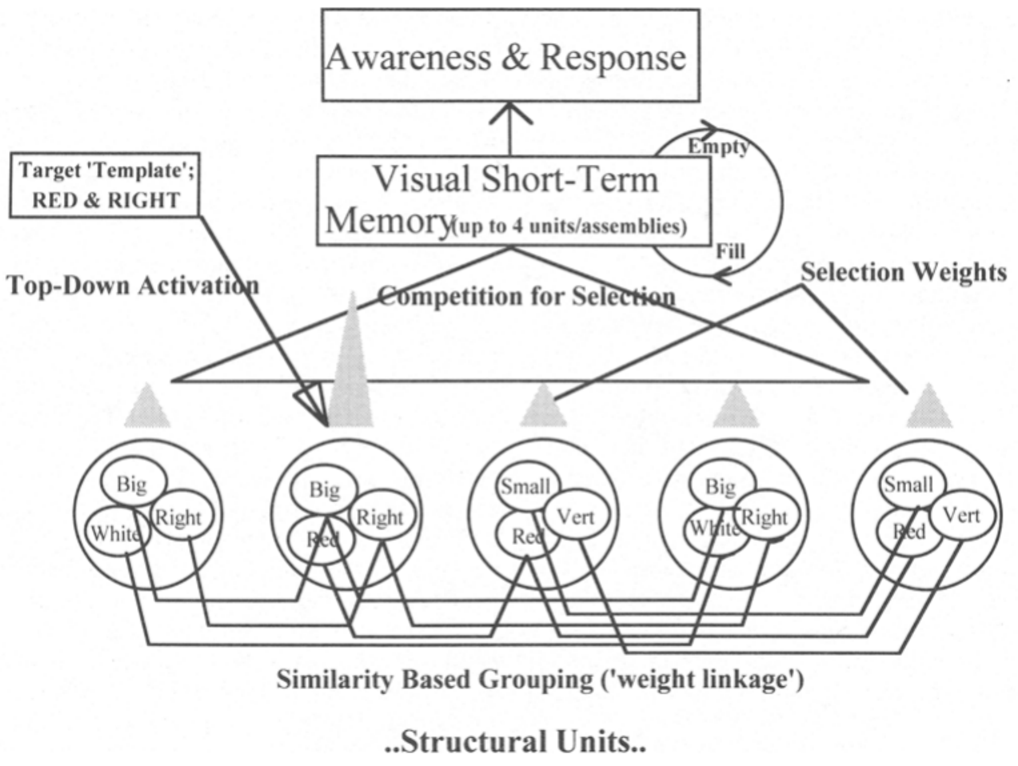
- Duncan and Humphreys (1989)
- Is an **alternative** theory to **FIT**

Similarity Theory of Attention



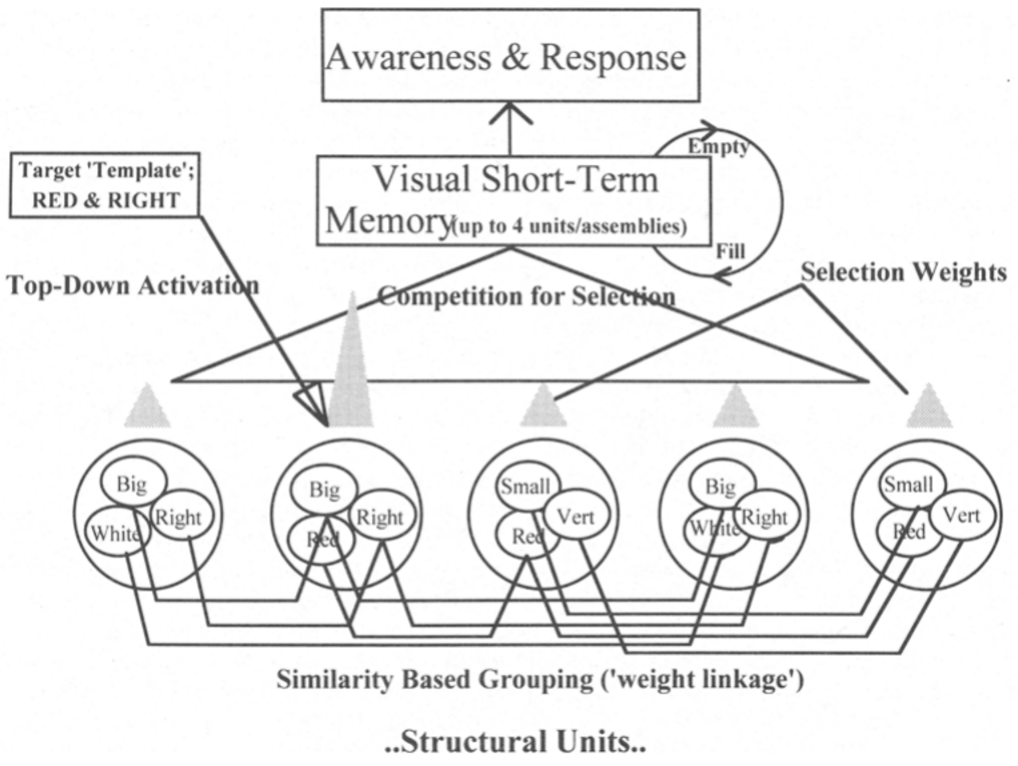
- Duncan and Humphreys (1989)
- Is an alternative theory to FIT
- **Objects** are **grouped** by **similarity**

Similarity Theory of Attention



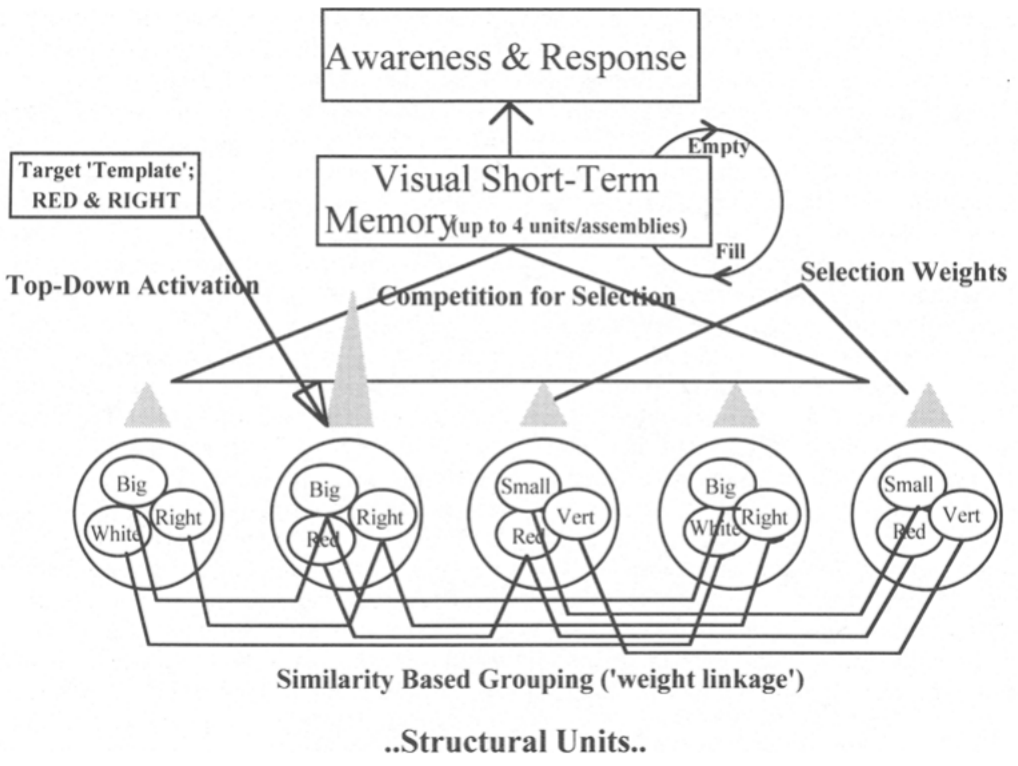
- Duncan and Humphreys (1989)
- Is an alternative theory to FIT
- Objects are grouped by similarity
- **Binding** of features **without attention**

Similarity Theory of Attention



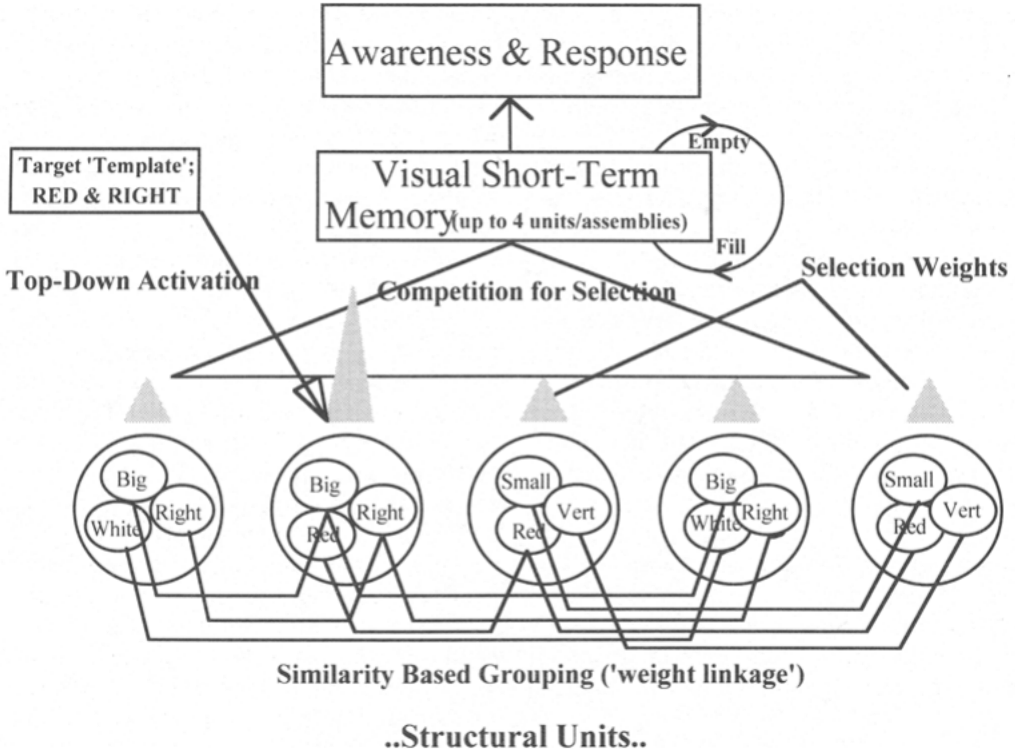
- **Similarity** between **targets** and **distractors** is the important **factor** for RTs

Similarity Theory of Attention



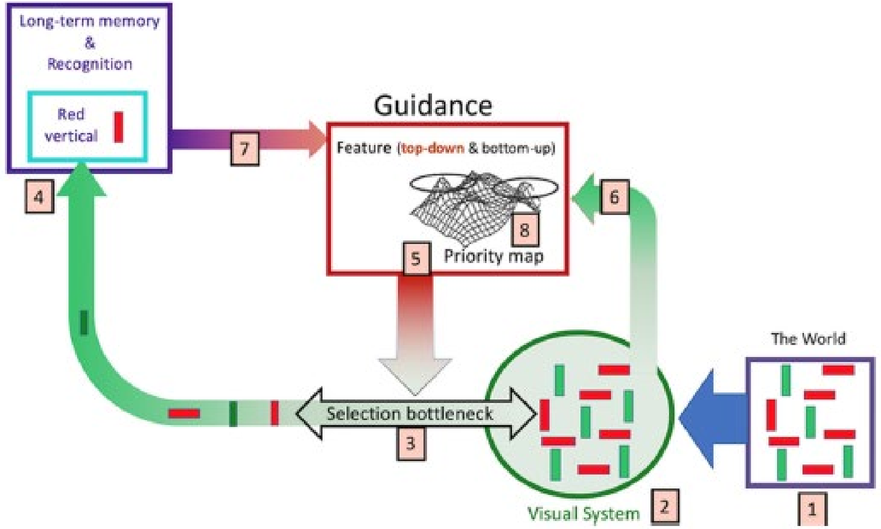
- Similarity between targets and distractors is the important factor for RTs
- The **capacity limit** of **VSTM** is the **origin** of the attentional **bottleneck**

Similarity Theory of Attention



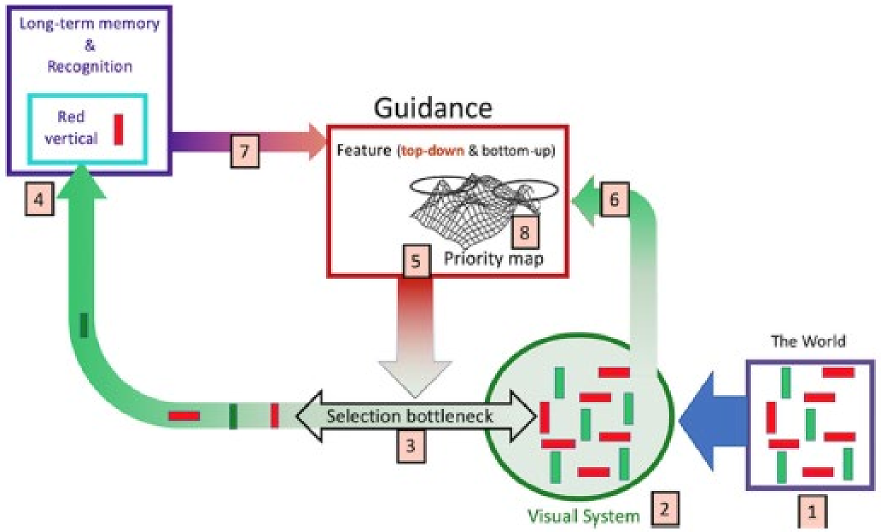
- Similarity between targets and distractors is the important factor for RTs
- The capacity limit of VSTM is the origin of the attentional bottleneck
- Some **findings cannot** be **explained** by **FIT**

Guided search (GS)



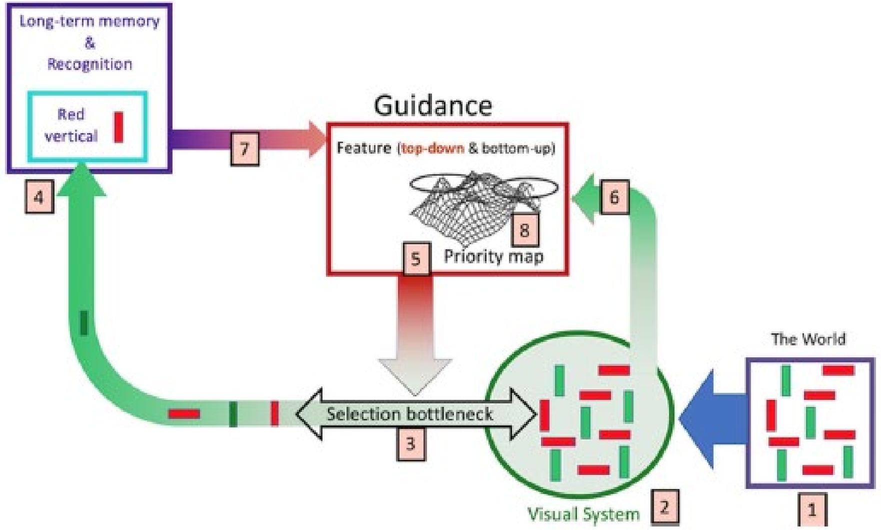
- By Jeremy **Wolfe (1994)**

Guided search (GS)



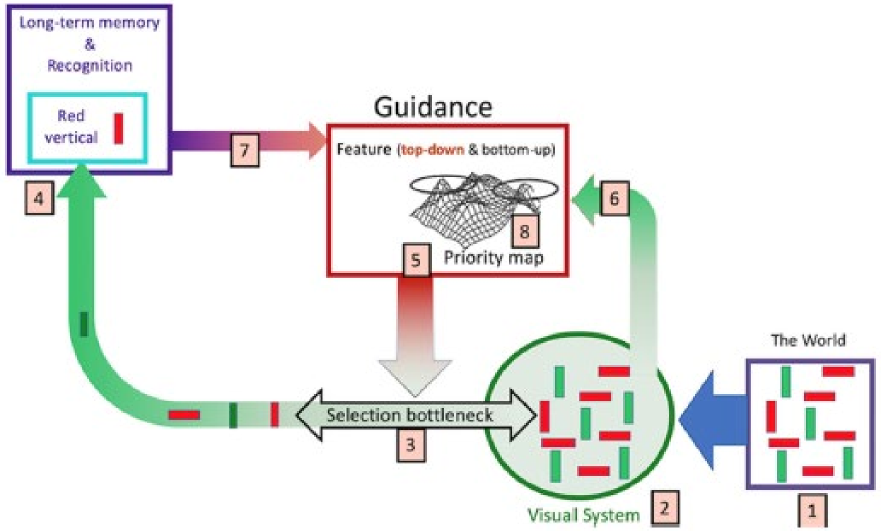
- By Jeremy Wolfe (1994)
- **Prevalent model** of visual search

Guided search (GS)



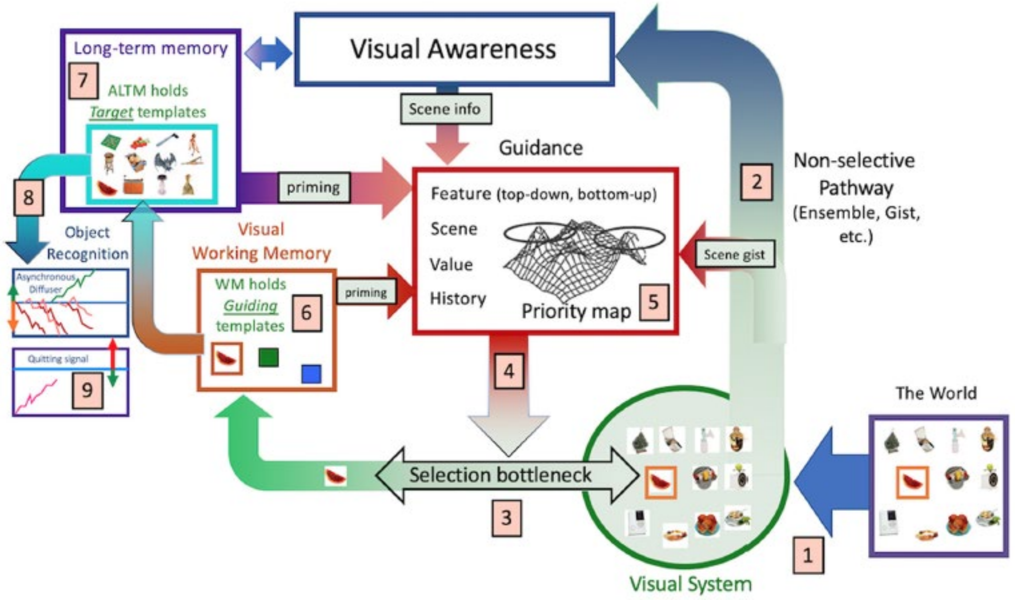
- By Jeremy Wolfe (1994)
- Prevalent model of visual search
- In the **spirit of FIT**, postulates **binding through attention**

Guided search (GS)



- By Jeremy Wolfe (1994)
- Prevalent model of visual search
- In the spirit of FIT, postulates binding through attention
- Was able to **explain** the **findings** that FIT failed to explain

Guided search (GS)



- By Jeremy Wolfe (1994)
- Prevalent model of visual search
- In the spirit of FIT, postulates binding through attention
- Was able to explain the findings that FIT failed to explain
- Still in **active development** (Wolfe, 2021)

Are Features bound with or without attention?

- Since **both** similarity theory and guided search delivered a **plausible** theory, the **question** remained **open**

Are Features bound with or without attention?

- Since both similarity theory and guided search delivered a plausible theory, the question remained open
- In **1998 Found** provided **evidence**, that a **third** feature that was correlated but irrelevant, could **improve the efficiency** of **conjunctive** visual **search**

Are Features bound with or without attention?

- Since both similarity theory and guided search delivered a plausible theory, the question remained open
- In 1998 Found provided evidence, that a third feature that was correlated but irrelevant, could improve the efficiency of conjunctive visual search
- **Found** considered its findings to be **consistent** with “preattentive binding” as proposed by the **similarity theory** and **not** with **guided search**

Are Features bound with or without attention?

- **Proulx** (**2007**) expanded on these considerations and **found** that **salient**, **task-irrelevant** singleton **features influence** search **efficiency**

Are Features bound with or without attention?

- Proulx (2007) expanded on these considerations and found that salient, task-irrelevant singleton features influenced search efficiency
- This led **Proulx** to **propose** that **both** GS and similarity theory **understate** the role of **bottom-up saliency** in **conjunction** searches

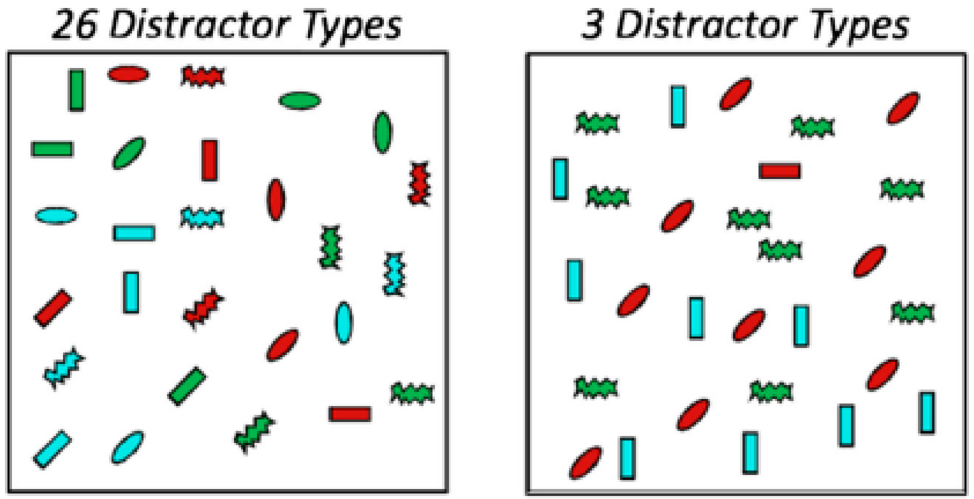
Are Features bound with or without attention?

- Proulx (2007) expanded on these considerations and found that salient, task-irrelevant singleton features influenced search efficiency
- This led Proulx to propose that both GS and similarity theory understate the role of bottom-up saliency in conjunction searches
- He **concluded** that **understanding** the **role** of **top-down** and **bottom-up** guidance is **crucial** for **models** of visual search

Are Features bound with or without attention?

- Proulx (2007) expanded on these considerations and found that salient, task-irrelevant singleton features influenced search efficiency
- This led Proulx to propose that both GS and similarity theory understate the role of bottom-up saliency in conjunction searches
- He concluded that understanding the role of top-down and bottom-up guidance is crucial for models of visual search
- And that on a **theoretical** level, the **surprising evidence** that bottom-up processing guides attention in conjunction search will **need to be addressed by models of visual search**

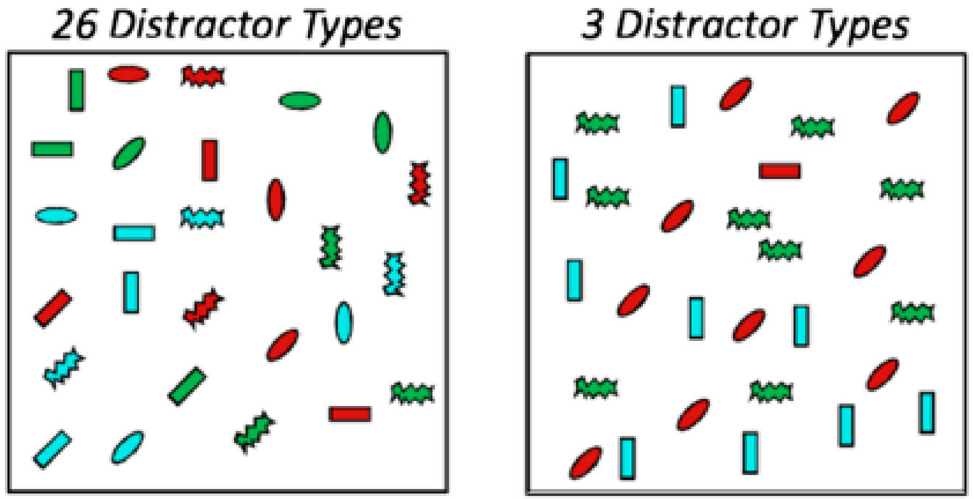
Triple Conjunction Visual Search



- **Nordfang** and **Wolfe** (**2014**) revisited **triple conjunction** searches

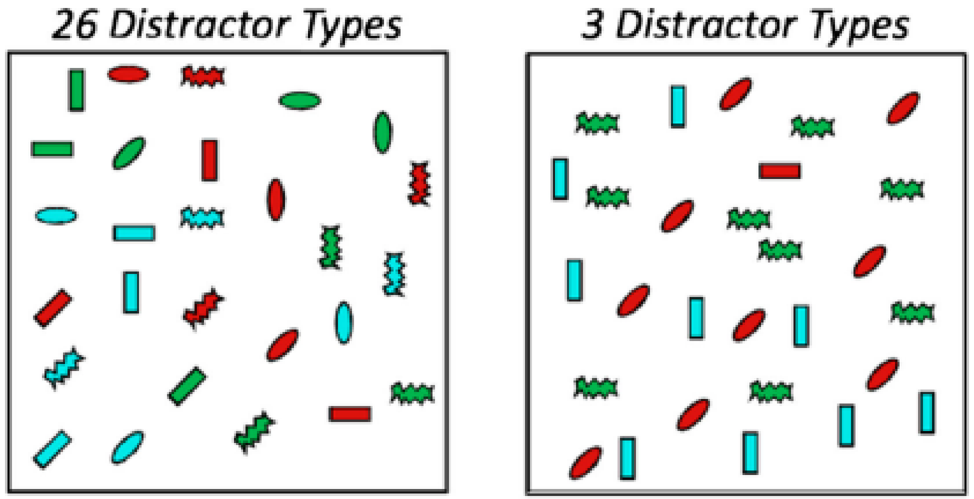
Nordfang and Wolfe. Guided search for triple conjunctions. Atten Percept Psychophys (2014)

Triple Conjunction Visual Search



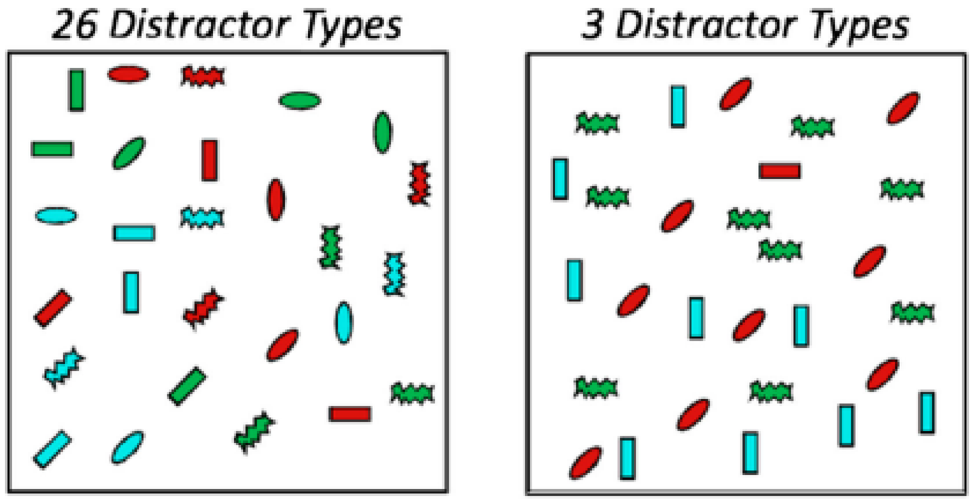
- Nordfang and Wolfe (2014) revisited triple conjunction searches
- They **found evidence** that both:

Triple Conjunction Visual Search



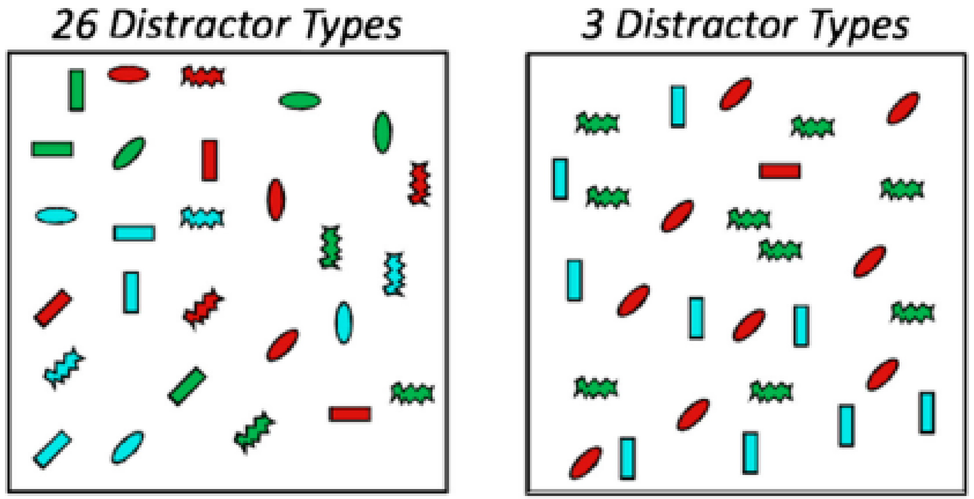
- Nordfang and Wolfe (2014) revisited triple conjunction searches
- They found evidence that both:
  - ***grouping***, the **number** of **different distractor groups** in a search display,

Triple Conjunction Visual Search



- Nordfang and Wolfe (2014) revisited triple conjunction searches
- They found evidence that both:
  - *grouping*, the number of different distractor groups in a search display,
  - and ***feature sharing***, the **number** of **features shared** between a **distractor** and the **target**,

Triple Conjunction Visual Search



- Nordfang and Wolfe (2014) revisited triple conjunction searches
- They found evidence that both:
  - *grouping*, the number of different distractor groups in a search display,
  - and *feature sharing*, the number of features shared between a distractor and the target,
- had a **substantial effect** on search **efficiency**

Triple Conjunction Visual Search

- They **concluded** that their **findings could** be **explained** by **preattentive binding**

Triple Conjunction Visual Search

- They concluded that their findings could be explained by preattentive binding
- **But** that **very efficient top-down** guidance based on a **nonlinear** ***sharing effect*** and/or **nonlinear** ***grouping effects*** in **bottom-up** salience **may also account** for the **observations** without resorting to preattentive binding

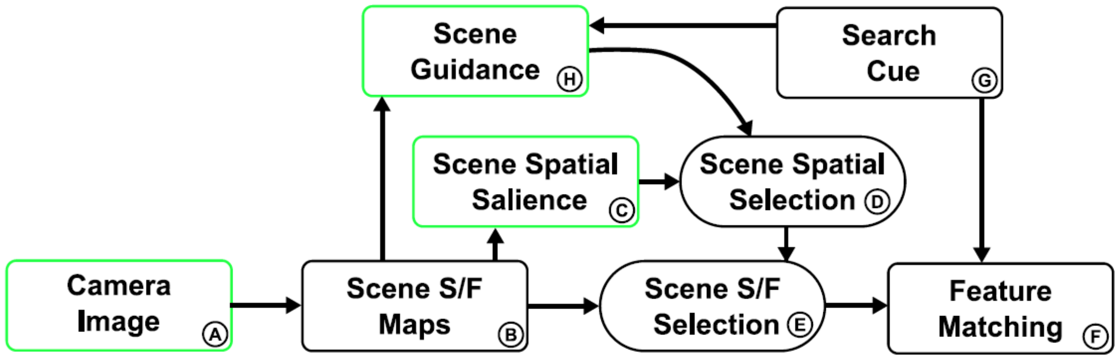
Triple Conjunction Visual Search

- They concluded that their findings could be explained by preattentive binding
- But that very efficient top-down guidance based on a nonlinear *sharing effect* and/or nonlinear *grouping effects* in bottom-up salience may also account for the observations without resorting to preattentive binding
- As they **expected** these to be **not trivial to model**, the **verification** of their **proposal** remained **open**

Triple Conjunction Visual Search

- They concluded that their findings could be explained by preattentive binding
- But that very efficient top-down guidance based on a nonlinear *sharing effect* and/or nonlinear *grouping effects* in bottom-up salience may also account for the observations without resorting to preattentive binding
- As they expected these to be not trivial to model, the verification of their proposal remained open
- Until today there is **no model** of visual attention and/or search **able** to fit or **explain** these intriguing **findings**

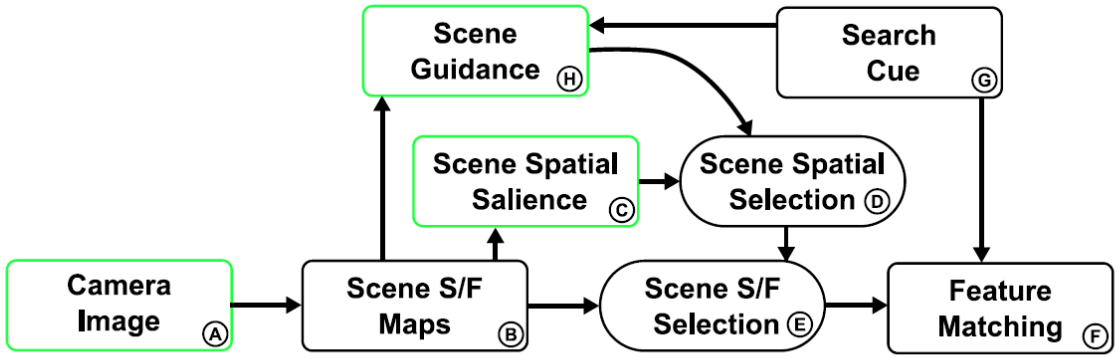
Model



- To **ease understanding**, we **reduced** our **previous** neural dynamic process **model** (Grieben et al., 2020) to its **visual search component only** (removing sub-networks related to scene memory and transient detection)

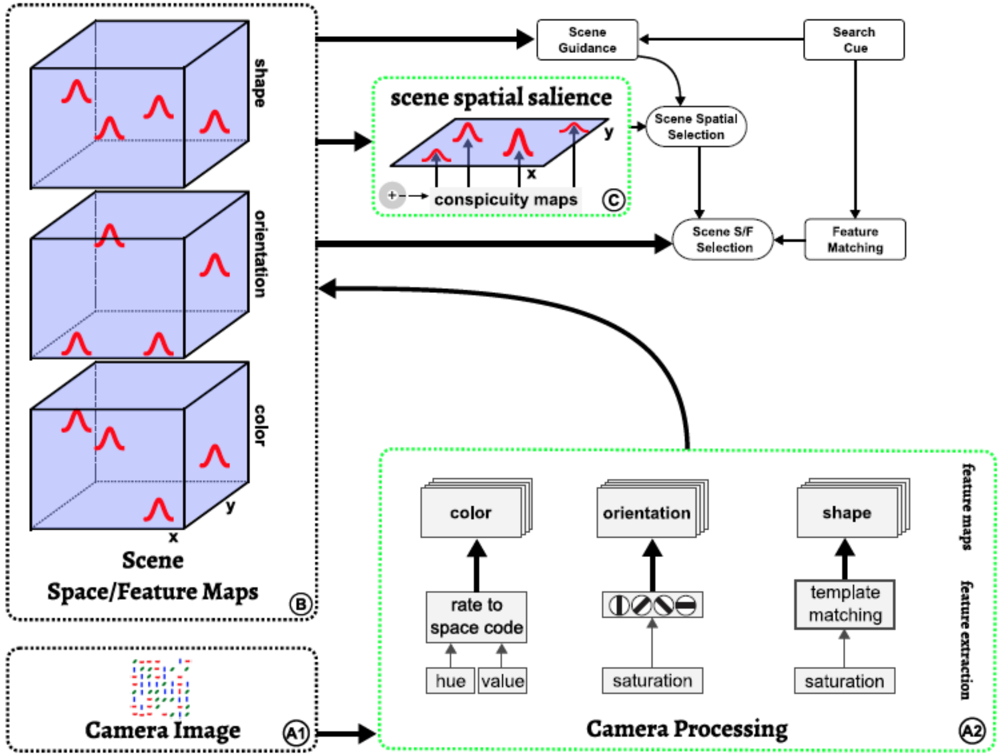
Grieben and Schöner. A neural dynamic process model of combined bottom-up and top-down guidance in triple conjunction visual search. CogSci (2021)

Model

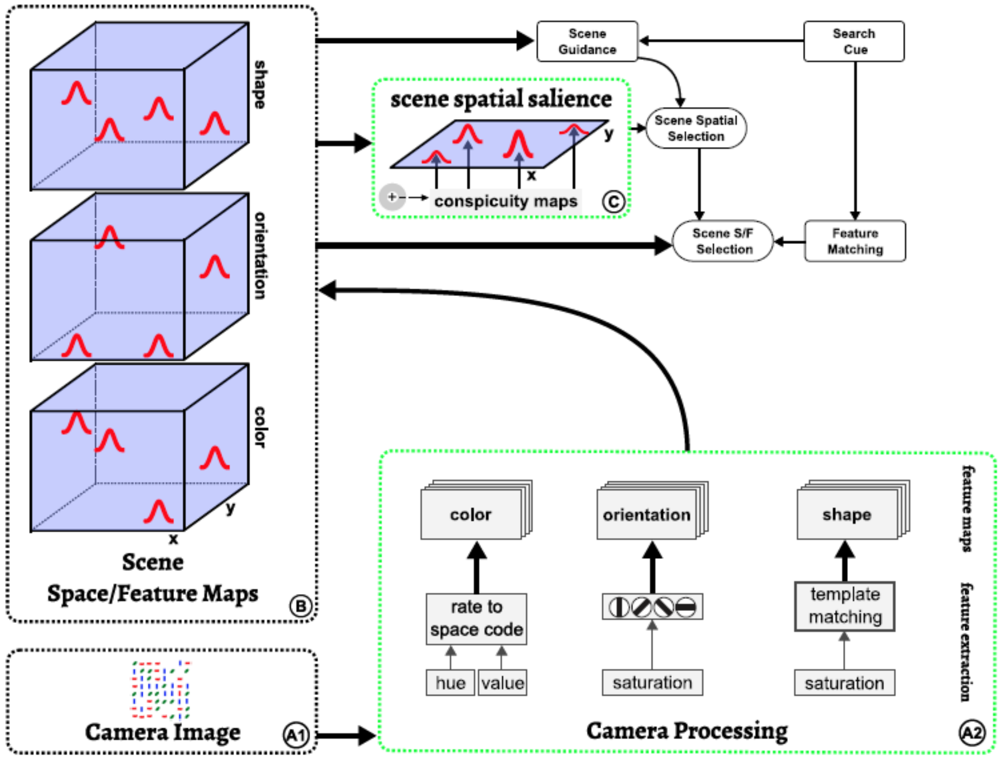


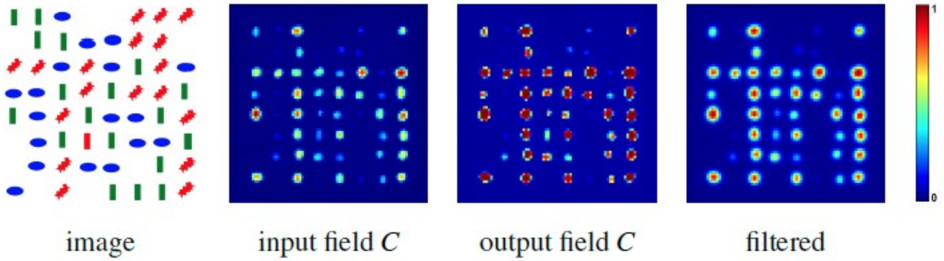
- To ease understanding, we reduced our previous neural dynamic process model (Grieben et al., 2020) to its visual search component only (removing sub-networks related to scene memory and transient detection)
- **Green** outlines **highlight** sub-networks **changed** with respect to the **previous model**

Feed-Forward Feature Maps and Salience Map

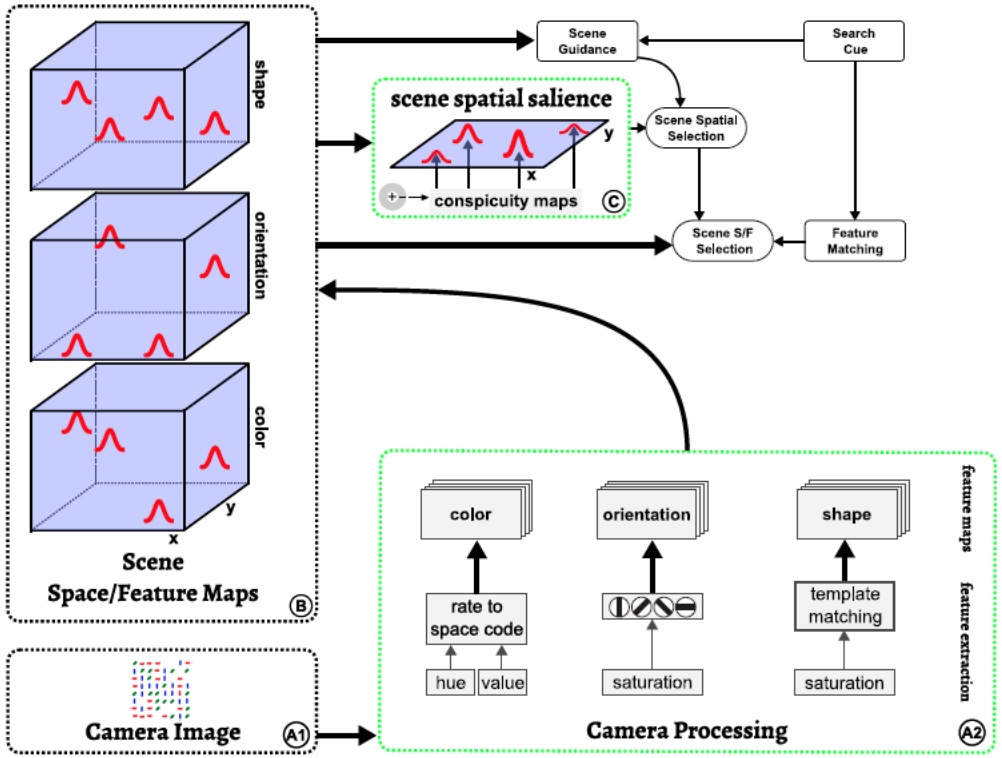


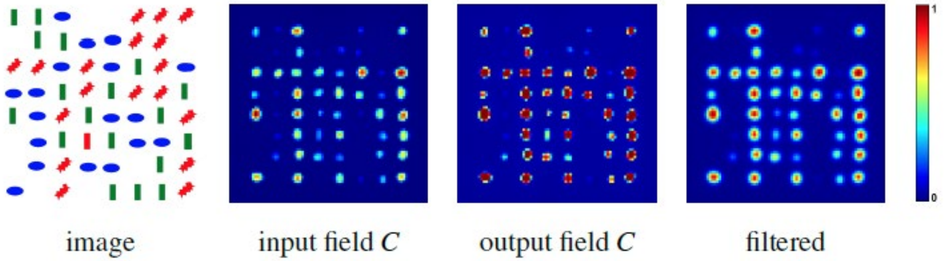
Feed-Forward Feature Maps and Salience Map





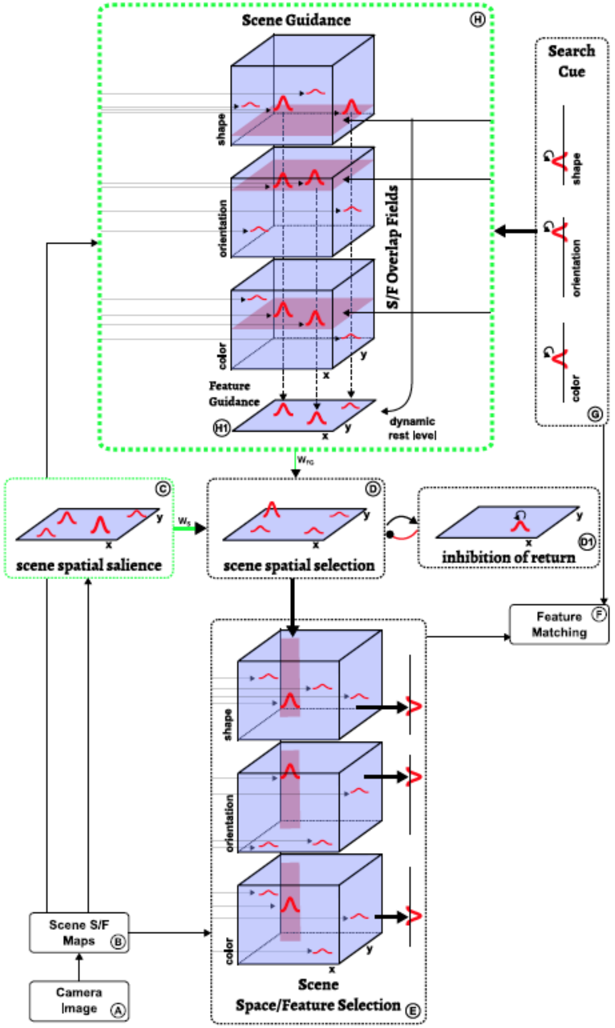
Feed-Forward Feature Maps and Salience Map



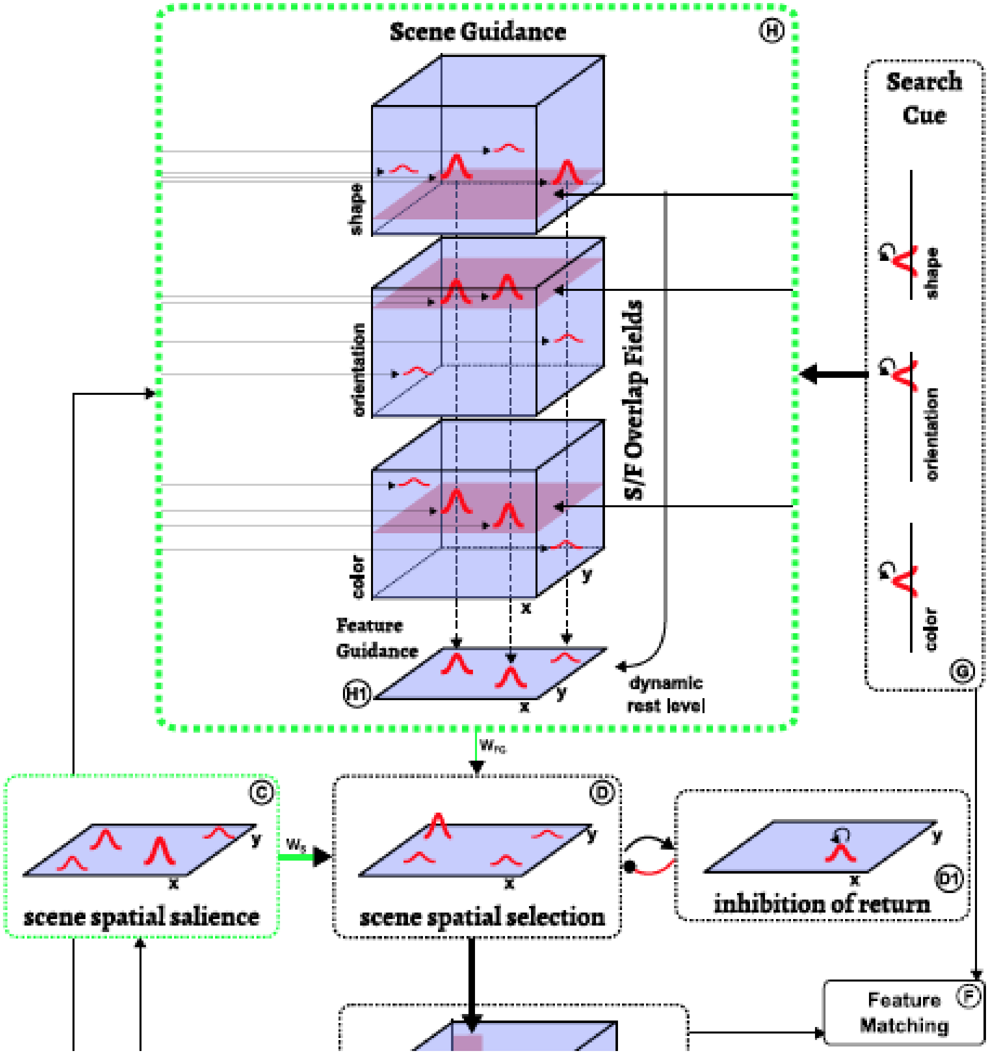


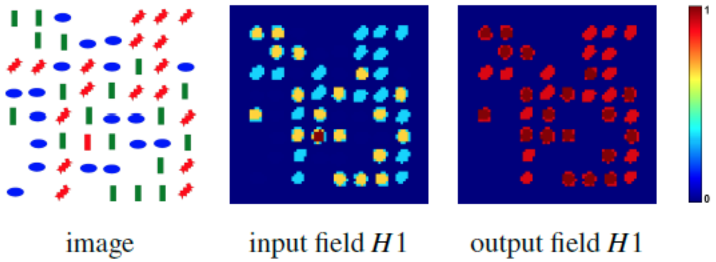
Responsible for the grouping effect

Attentional Selection and Visual Search

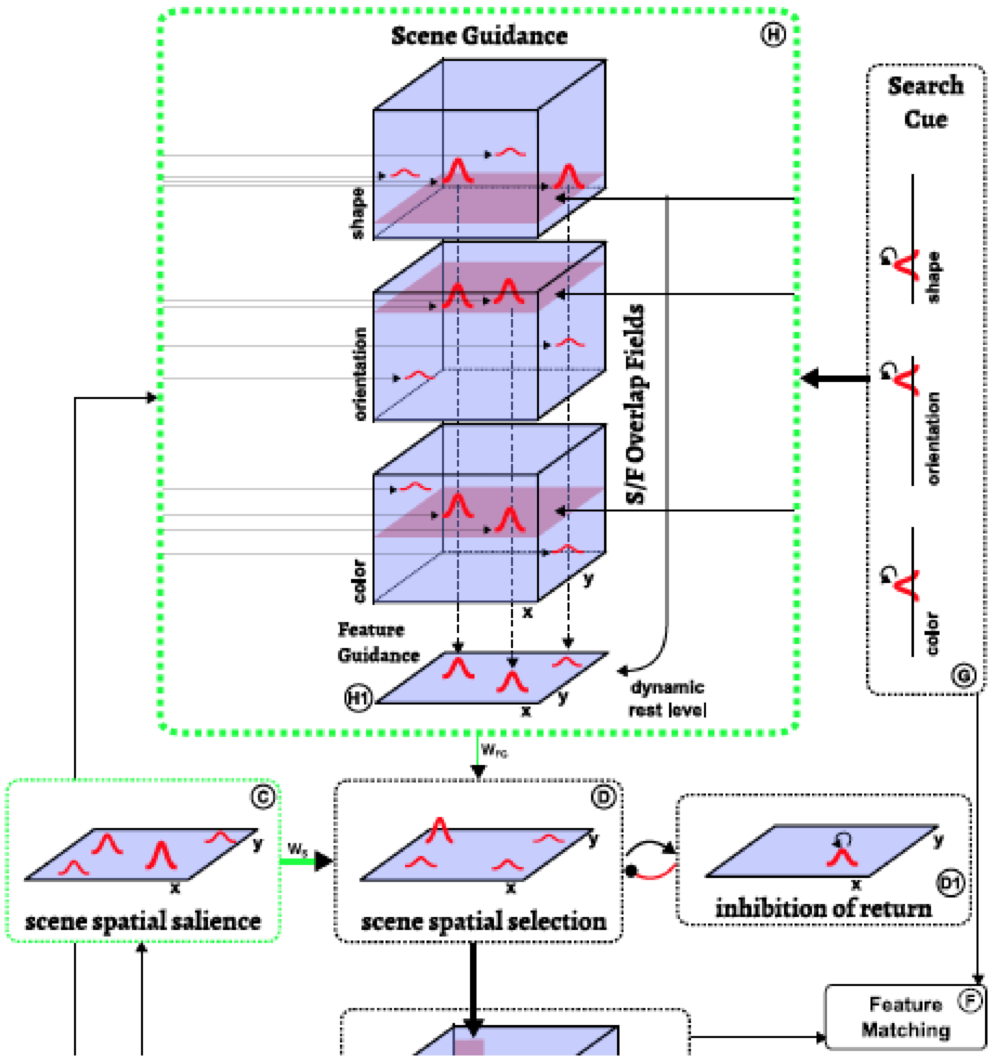


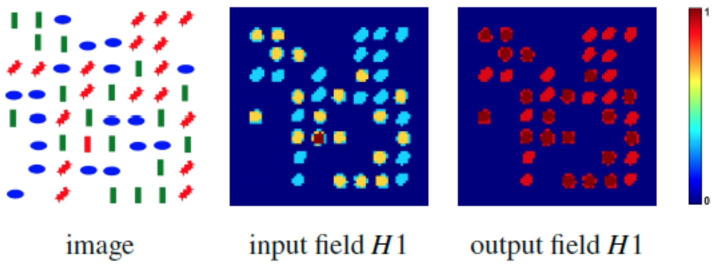
Attentional Selection and Visual Search





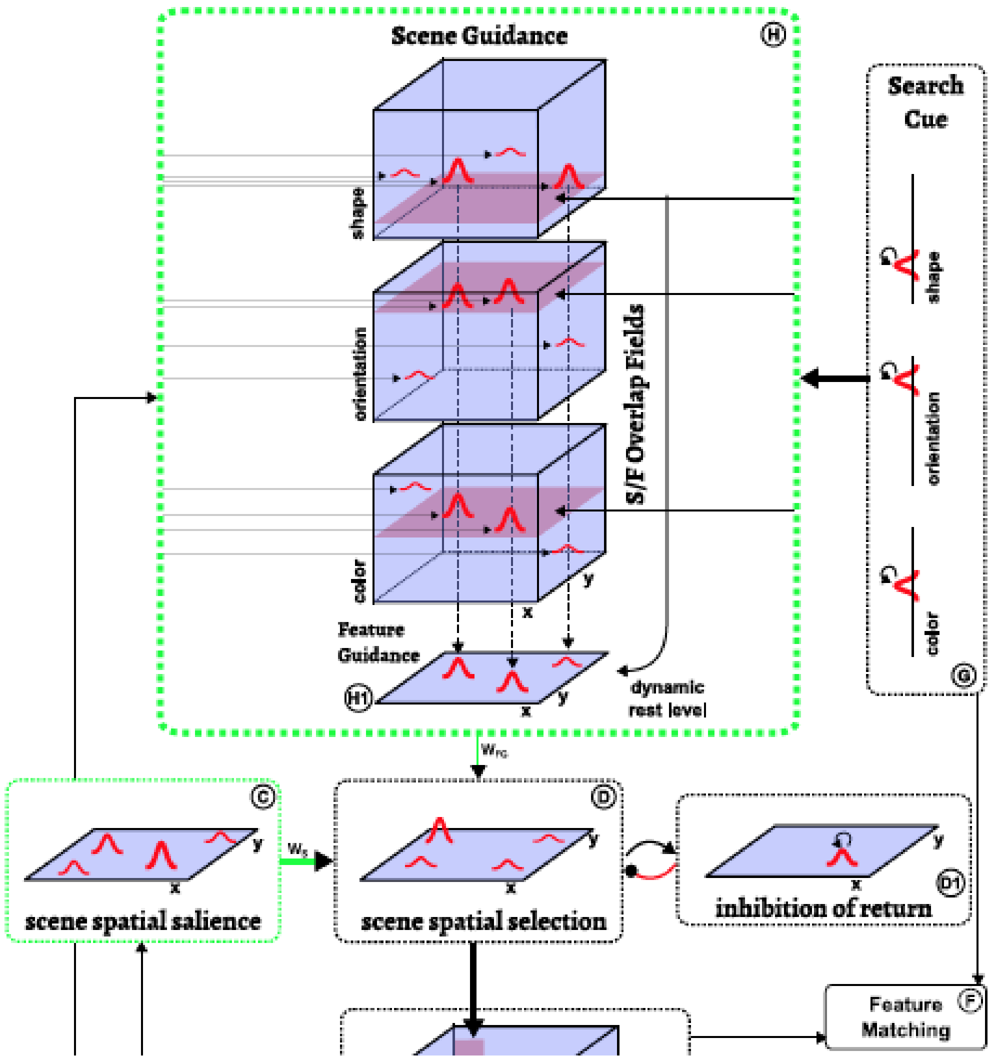
Attentional Selection and Visual Search

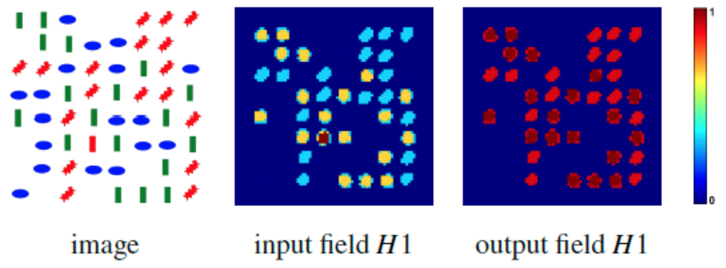


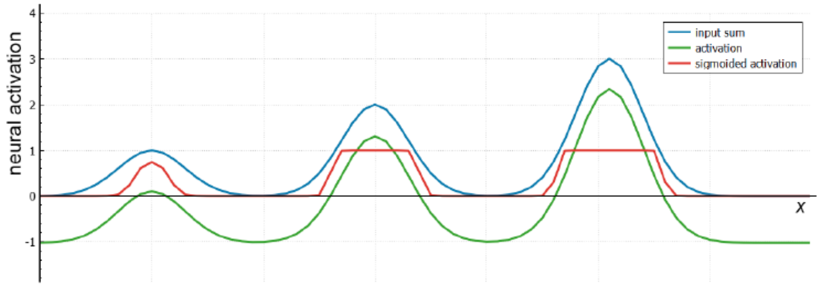


Responsible for the sharing effect

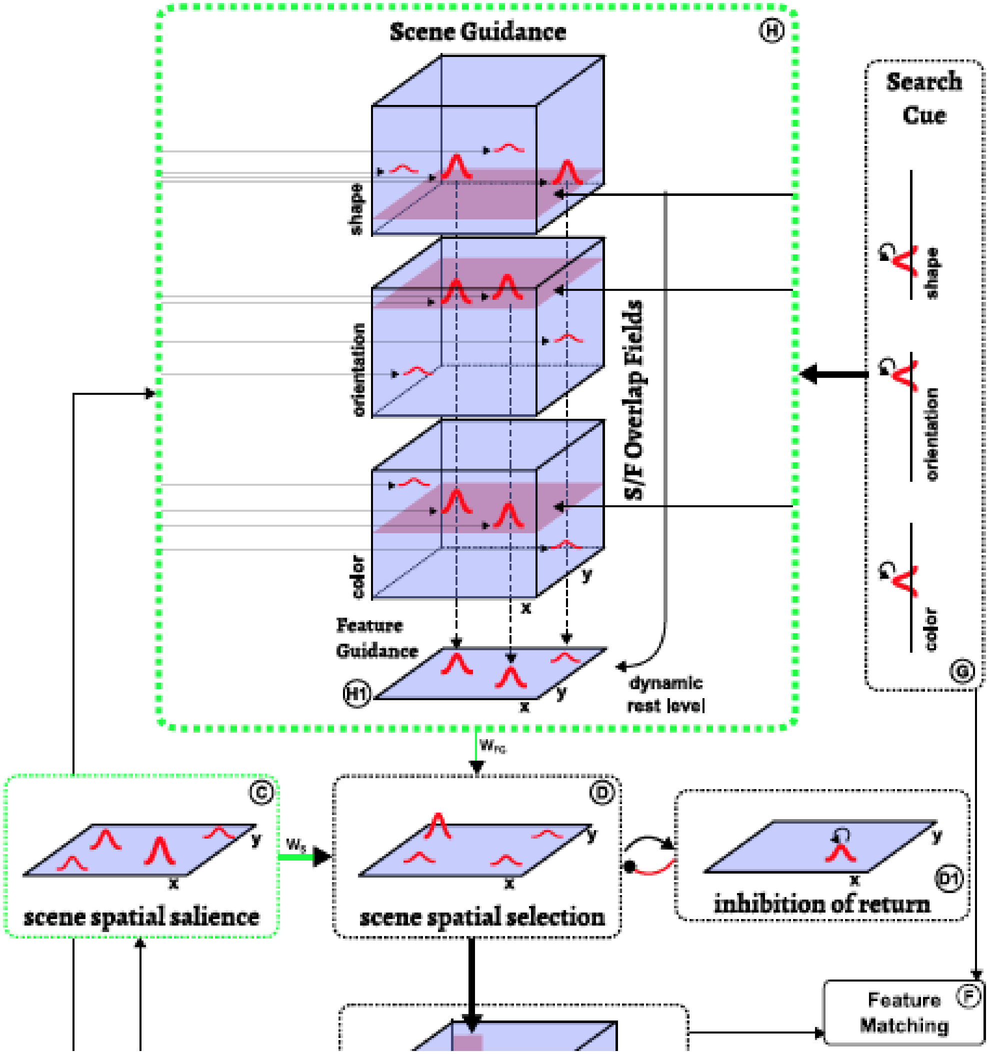
Attentional Selection and Visual Search

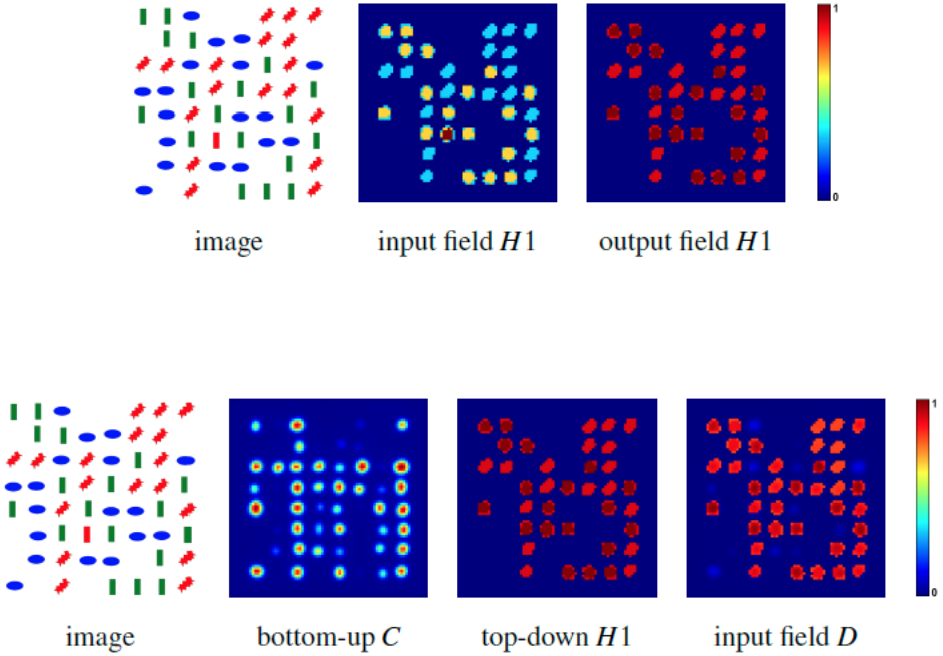




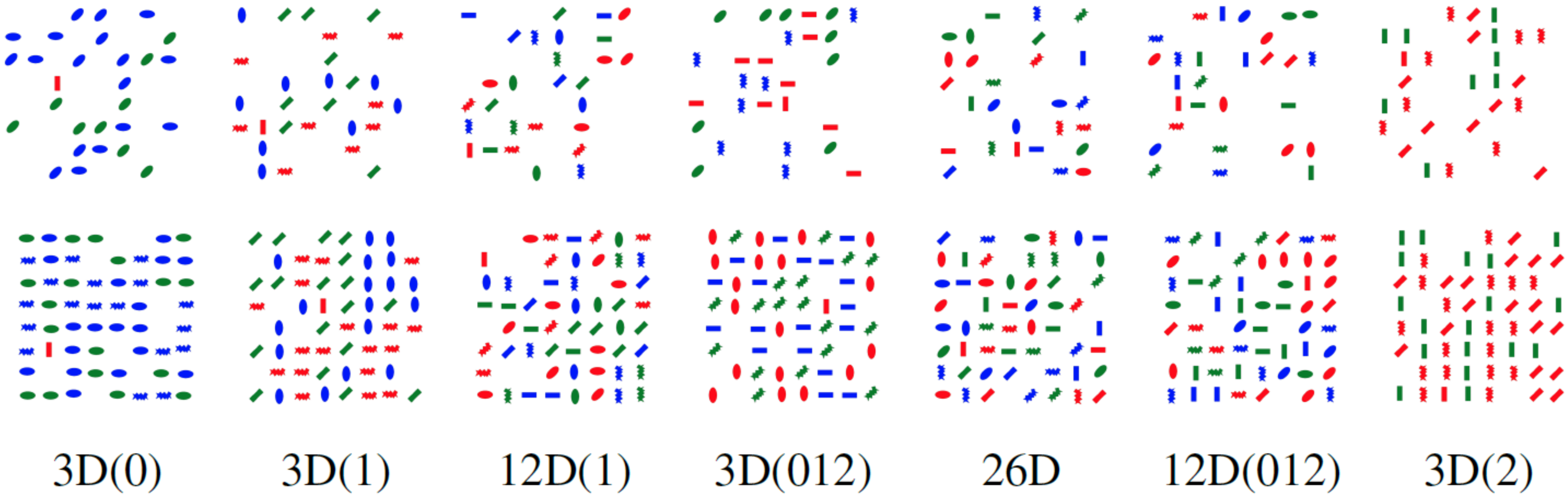


Attentional Selection and Visual Search





Results



Results

Table 1: The slopes of the RT × set size functions from the experiments, the previous model, and our model.

	Experiments (Nordfang & Wolfe, 2014)							Model (Grieben et al., 2020)		Model (this paper)	
	1a	1b	3	4	6	Slopes	$\bar{x}$	Slopes	$\bar{x}$	Slopes	$\bar{x}$
3D(0)					-1.2	-1.2	-1.2	0.0	0.0	0.0	0.0
3D(1)	2.0	4.0	2.4	3.0	2.4	2.0 - 4.0	2.8	0.0	0.0	1.1 - 2.8	1.9
12(1)			2.8	4.8		2.8 - 4.8	3.8	0.0	0.0	2.1 - 3.1	2.5
3D(012)	2.3	4.3		5.8	3.7	2.3 - 5.8	4.0	2.4 - 4.4	3.5	2.0 - 5.7	4.0
26D	4.9	6.5	3.4	6.2		3.4 - 6.5	5.3*	2.0 - 4.4	2.5	3.7 - 6.3	4.8
12D(012)			3.7	6.7		3.7 - 6.7	5.2*	2.2 - 4.4	3.5	3.9 - 6.7	5.3
3D(2)					19.8	19.8	19.8	8.2-15.1	11.2	19.8 - 22.3	21.2

* The mean for the 12D(012) condition is possibly misleading and the result of too few data points, since, from the direct comparison on a per experiment level it seems clear that this condition is presumably less efficient than condition 26D.

Conclusion

- In conclusion, the **model** provides a neural process account of the visual search paradigm that includes the **detection of the search cue from visual transients**, its **commitment to feature memory**, the **autonomous generation** of a **sequence** of attentional **selection decisions**, and the **matching** of the **cued feature** values to feature values extracted at each **attended location**.

Conclusion

- In conclusion, the model provides a neural process account of the visual search paradigm that includes the detection of the search cue from visual transients, its commitment to feature memory, the autonomous generation of a sequence of attentional selection decisions, and the matching of the cued feature values to feature values extracted at each attended location.
- The **model** accounts for conjunctive **searches** in a way that is **consistent** with the original notion of **binding through space**.

Conclusion

- I **showed experimentally** that allowing observers to first build a **scene** working **memory** before performing visual search not only **speeds** visual **search** as often reported, but also **increases** search **efficiency**, **an effect that has remained elusive for a long time**.

Conclusion

- I showed experimentally that allowing observers to first build a scene working memory before performing visual search not only speeds visual search as often reported, but also increases search efficiency, an effect that has remained elusive for a long time.
- I explained how this **effect emerges** from the time- and state-continuous **neural processes** in our **model**.

Conclusion

- We **extended our** neural dynamic process **model** for scene perception and top-down guided visual search (Grieben et al., 2020) to **account** for the feature **sharing** and **grouping effects** found by Nordfang and Wolfe (2014) for **triple conjunction** searches

Conclusion

- We extended our neural dynamic process model for scene perception and top-down guided visual search (Grieben et al., 2020) to qualitatively fit the feature sharing and grouping effects found by Nordfang and Wolfe (2014) for triple conjunction searches
- The **new version** of **our model accounts** for the **differences** between the conditions **observed** by Nordfang and Wolfe (2014) **without resorting to preattentive binding**

Conclusion

- We also **addressed** a **major theoretical weakness** of **models** of **conjunctive** visual **search** (Proulx, 2007)

Conclusion

- We also addressed a major theoretical weakness of models of conjunctive visual search (Proulx, 2007)
- Even though **bottom-up salience** may **disturb** the **efficiency** of top-down guided visual search, it is **crucial** for the visual **exploration** of a crowded **scene** in the **absence** of a **task**

Conclusion

- We also addressed a major theoretical weakness of models of conjunctive visual search (Proulx, 2007)
- Even though bottom-up salience may disturb the efficiency of top-down guided visual search, it is crucial for the visual exploration of a crowded scene in the absence of a task
- Through the **incorporation** of **bottom-up salience** our **model** is now **able** to **autonomously explore** the scene by bringing objects into the attentional foreground through **selective competition**, even in the absence of a task-induced top-down bias

Questions?

Thank you for your attention!