UNIVERSITÀ
DEGLI STUDI
FIRENZE

DOTTORATO DI RICERCA TOSCANO IN NEUROSCIENZE

CICLO XXXV

COORDINATORE Prof.ssa Maria Pia Amato

Assessing visual saliency of informative local features with psychophysics and eye movements

Settore Scientifico Disciplinare M-PSI/02

Dottoranda
Dott.ssa Castellotti Serena

Serena Castellotti

Tutore
Prof.ssa Maria Michela Del Viva

Coordinatore
Prof.ssa Maria Pia Amato

MMAmato

Anni 2019/2022

SUMMARY

Under fast viewing conditions, the visual system extracts salient and simplified representations of complex visual scenes. Eye movements optimize such visual analysis through the dynamic sampling of the most informative and salient regions in the scene. However, the definition of saliency, as well as its role in natural active vision, is still a matter of discussion. The present thesis is based on a recent *constrained maximum-entropy* model of early vision (Del Viva et al., 2013), that deals with the problem of the extraction of biologically relevant information from a large flux of input data in the shortest possible time for survival purposes. According to this model, the visual system produces an early saliency map of a visual scene selecting a limited number of local features, based on criteria of maximal entropy coupled with strict limitations on computational resources. The model, applied to natural images, extracts a set of optimal-information carrier features as candidate "salient" features.

The present thesis includes four different studies, which aim to assess the visual saliency of these optimally informative visual features, adding further evidence that confirms the predictions of the reference model. Particularly, we are interested in understanding the role of *optimal* visual features in the creation of a bottom-up *saliency map* that the oculomotor system could use to drive eye movements toward potentially relevant locations, therefore, ultimately, in their contribution to image reconstruction. In our experiments, the results obtained with *optimal* features are compared to those obtained with other features that do not meet the optimality criteria requested by the model, therefore discarded and considered *non-optimal*. Considering that luminance contrast has a central role in determining saliency in fast vision, we also compare the effects induced by *optimal* versus *non-optimal* features to those obtained with features of different luminance.

Before describing our experiments, in Chapter 1, the main properties of visual analysis with some notes about eye movements are described. Then the limitations of our visual system and the resulting need for data reduction in fast vision are discussed. Finally, Chapter 1 is mostly dedicated to the presentation of the *reference* model of early vision, with an extensive discussion of the main computational and behavioral results found in previous works.

After that, each chapter is dedicated to the presentation of a study. Although all the studies share the same final objective, each one has its own rationale and a specific research question to answer. Chapter 2 presents the literature about the *saliency map*, and then describes our Study 1 involving perceptual and eye movement tasks. In this study, *optimal* features were presented in isolation, to investigate whether they are considered visually more salient than other *non-optimal* features, even

in the absence of any meaningful global arrangement and semantic context. In Chapter 3, the topic of cover and overt attention has been summarized, and then Study 2 is presented, in which we implicitly tested the bottom-up saliency driven by *optimal* features by engaging participants in covert attentional and gaze-orienting cued tasks without explicitly requiring them to pay attention to stimulus saliency. Chapter 4 firstly discusses some saccades' properties and how visual distractors influence their trajectories. Then Study 3 is presented, in which we compared the effects on saccades trajectories produced by *optimal* vs. *non-optimal* features used as distractors in a saccadic task, considering the magnitude of curvature as a measure of feature saliency. Finally, Chapter 5 described the problem of occluded objects in real scenes and how our visual system can recognize the whole image based only on little fragmented information. Our Study 4 presented here, explore whether optimal features also play a significant role in more natural settings, investigating the contribution of optimal local information contained in a few visible fragments to image discrimination in fast vision. Chapter 6 is finally dedicated to the discussion of the results obtained in our studies.

Overall, the results show that *optimal* features are considered visually salient, they can automatically attract attention, they interfere with the path of saccades, and they partially contribute to image discrimination. On the other end, *non-optimal* features do not produce the same effects. These findings suggest that optimally informative local features get preferential treatment during fast image analysis and automatically guide attention and eye movements to create a bottom-up *saliency map*.

Note that, according to the *reference* model, *optimal* features represent a compromise between the amount of information they carry about the visual scene and the cost for the system to process them; whereas *non-optimal* features used in our experiments are individually the most informative, but their use also implies large computational costs. Our findings then suggest that not only the amount of information but also the need of saving computational resources takes a significant role in shaping what the visual system considers to be *salient.*

Very interestingly, all the effects found with *optimal* features are similar to those obtained with high-luminance features, suggesting that the saliency determined by information maximization criteria produces effects comparable to those due to luminance-based saliency.

Let me also mention that, in our studies, we employ some novel paradigms that may be useful tools to test the relative saliency of different stimuli in future research.

To conclude, the findings presented in this thesis suggest that visual saliency may be derived naturally in a system that, under the pressure of fast visual analysis, operates maximum information transmission under computational limitation constraints, as predicted by the reference model.

Table of contents

List of figures

List of supplementary figures

List of tables

Publications

Some of the findings reported in this thesis have been published in peer-reviewed journals.

The study presented in Chapter 2 is included in the following publication:

- Castellotti, S., Montagnini, A., & Del Viva, M.M. (2021). Early visual saliency based on isolated optimal features. Front. Neurosci. 15:645743. https://doi.org/10.3389/fnins.2021.645743

The study presented in Chapter 3 is included in the following publication:

- Castellotti, S., Montagnini, A., & Del Viva, M. M. (2022). Information-optimal local features automatically attract covert and overt attention. Scientific Reports, 12(1), 9994. https://doi.org/10.1038/s41598-022-14262-2

The study presented in Chapter 4 is included in the following publication:

- Castellotti S., Szinte, M., Del Viva, M. M., & Montagnini, A. (2023). Saccadic trajectories deviate toward or away from optimally informative visual features. IScience, 107282. https://doi.org/https://doi.org/10.1016/j.isci.2023.107282

The study presented in Chapter 5 is included in the following publication:

- Castellotti S., D'Agostino, O., & Del Viva, M.M. (2023). Fast discrimination of fragmentary images: the role of local optimal information. Frontiers in Human Neuroscience, 17. https://doi.org/10.3389/fnhum.2023.1049615

Chapter 1

Introduction

1. INTRODUCTION

1.1 Visual analysis

1.1.1 The visual system: from retinal processes to cortical structures

Human visual perception is an active process that can be defined as the ability to receive and interpret the information that our eyes capture. Visual processing begins in the retina, the innermost light-sensitive layer at the back of the eye, consisting of several layers of interconnected neurons (**Figure 1A**; Baylor, 1987; Mayer & Dowling, 1988). Photoreceptors, located in the outermost layer, absorb light and convert it into a neural signal, an essential process known as phototransduction (for a review, see Luo et al., 2008). There are two main types of photoreceptor cells, each differently contributing to the formation of the visual image. Rods are very sensitive to low-luminance levels as well as to luminance variations, allowing good vision in dim conditions (scotopic vision), but they are not sensitive to color and present a low spatial resolution. Cones are much less sensitive to light and only function in bright conditions (photopic vision), have a high spatial resolution, and are responsible for the perception of color. Primates have only one type of rod but three kinds of cone photoreceptors, distinguished by the range of wavelengths to which they respond: the L (long-wave), M (medium-wave), and S (short-wave) cones. The rods constitute the majority of the receptors in our retina, and they are mainly concentrated in the periphery, while the density of cones peaks in the center of the retina (fovea) and rapidly decreases away from it (**Figure 1B**; Rodieck, 1998).

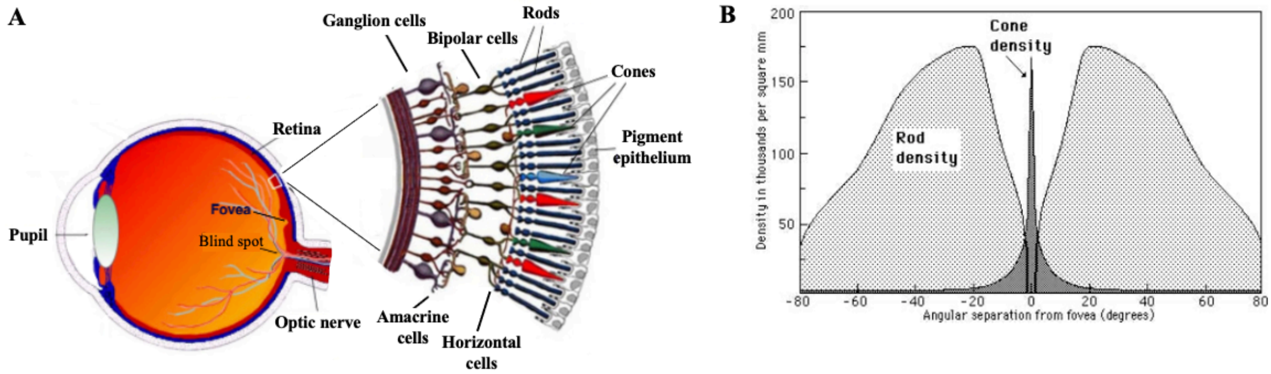

Figure 1. Neurons in the human retina. (A) Representation of retinal layers. Figure adapted from (Cavaletti & Marmiroli, 2009). **(B) Distribution of photoreceptors in the retina.** Distribution of rods and cones plotted as a function of the distance from the center of the fovea. Figure adapted from (Østerberg, 1937).

Photoreceptors signals are synaptically transmitted to bipolar cells, which in turn connect to ganglion cells in the innermost layer. The site where axons of ganglion cells converge is devoid of photoreceptors and thus corresponds to a blind spot in the visual field of each eye. In addition to this vertical pathway, the retinal circuit includes many lateral connections provided by horizontal cells and amacrine cells (Figure 1A; Cavaletti & Marmiroli, 2009).

Retinal ganglion cells are the output neurons of the retina, and their axons form the optic nerve from which the central geniculo-cortical pathway starts (**Figure 2**; Ferster & Lindström, 1983; Garey & Powell, 1971; Spatz, 1979). The optic nerve extends to a midline crossing point, the optic chiasm. Beyond the chiasm fibers from the temporal hemiretinas proceed to the ipsilateral hemisphere; fibers from the nasal hemiretinas cross to the contralateral hemisphere. This partial decussation of fibers ensures that all the information about each hemifield is processed in the visual cortex of the contralateral hemisphere.

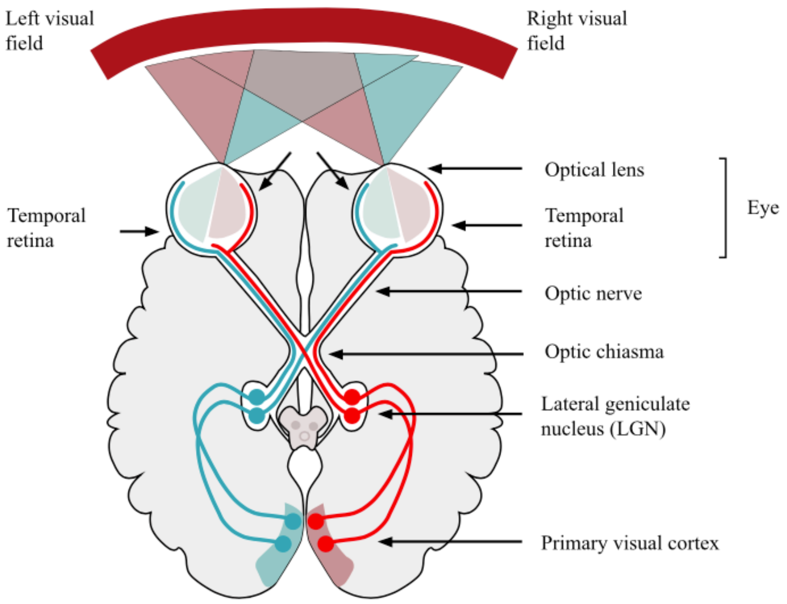

Figure 2. Geniculo-cortical pathway. Visual pathway from the retina to the primary visual cortex.

Axons then join in the optic tract that extends to the lateral geniculate nucleus (LGN) of the thalamus, consisting of six layers, each receiving input from either the ipsilateral or the contralateral eye. The inner two layers are magnocellular layers (M-cells), while the outer four layers are parvocellular layers (P-cells) (Kaplan et al., 1990; Valberg & Lee, 1992). M-cells are larger and have a faster conduction speed than P-cells (Maunsell et al., 1999), and they respond best to achromatic stimuli of low spatial and high temporal frequencies, whereas P-cells respond best to chromatic stimuli of high spatial and low temporal frequencies (Davis et al., 2006). The LGN neurons then project through the optic radiation to the primary visual cortex (V1), located in the occipital lobe.

1.1.1.1 Properties of the primary visual cortex

V1 has a very well-defined map of the spatial information in vision (*retinotopic map;* Holmes, 1918). The topography of visual input is largely conserved along the ascending projections, while it is not so for the metric relationships. Namely, two close regions in the visual scene are projected to close areas in the visual cortices, though their distance is transformed depending on their eccentricity with respect to the fovea: the projection of the foveal area is magnified with respect to the peripheral areas (*cortical magnification;* Cowey & Rolls, 1974; Daniel & Whitteridge, 1961; Slotnick et al., 2001).

The primary visual cortex is divided into six functionally distinct layers, of which the IV receives most of the visual input from the LGN (Douglas & Martin, 2007; Rockel et al., 1980; Shepherd, 2004). M- and P-cells of the LGN send information to different sublayers of layer IV. The parvocellular pathway allows the perception of fine details, colors, and large changes in brightness. The magnocellular pathway carries information about large, fast things (low spatial frequency, high temporal frequency), it is colorblind and seems to be crucial for movement analysis (Nassi & Callaway, 2009; Pokorny, 2011).

The signals from the two eyes are kept separate in layer 4, whereas below and above this layer, most cells receive information from both eyes (Hubel & Wiesel, 1962, 1968). These *binocular* neurons integrate the signals from both the right and left eyes and create the perception of depth, contributing to the creation of stereopsis from binocular disparity (Qian, 1997; Scholl et al., 2013).

A main characteristic of the visual cortex is its organization into columns of specialized neurons: cells with similar functional properties, such as orientation selectivity and ocular dominance, are located close together in columns (Hubel & Wiesel, 1959, 1962, 1963).

V1 sends its main output to a set of higher-order visual areas (V2, V3, V4, V5), also organized as neural maps of the visual field (Felleman & Van Essen, 1991; Hubel & Wiesel, 1965; Maunsell & Newsome, 1987). Then, cortical processing of visual information continues mainly through two major pathways (Goodale & Milner, 1992); a ventral pathway toward the temporal lobe carries information about what the stimulus is, a dorsal pathway into the parietal lobe (and then to the frontal lobes) carries information about where the stimulus is, information that is critical for guiding movement.

1.1.1.2 Receptive fields along the visual pathway

In the visual system, a neuron's receptive field represents a small window on visual space. Receptive field properties change from layer to layer along the visual pathway (**Figure 3**; Nassi & Callaway, 2009).

The receptive fields of retinal ganglion cells have a center-surround organization (on-center and off-center; **Figure 3A – left panel**; Kuffler, 1953), leading to high sensitivity to borders (for a review see, Kim et al., 2021). Neurons in the lateral geniculate nucleus have similar receptive fields. These cells respond optimally to an appropriately placed spot of light of just the right size (**Figura 3A– right panel**).

Receptive field changes in the primary visual cortex, having an elongated structure (Hubel & Wiesel, 1959, 1962, 1968; Hubel, 1982). This structure holds an important mechanism in the brain's analysis of visual form. Indeed, the key property of these neurons is the selectivity for the orientation of contours. Hubel and Wiesel classified V1 cells into *simple* and *complex* cells (Hubel & Wiesel, 1959, 1962). The receptive fields of simple cells, receiving afferents from various LGN cells (Movshon et al., 1978b), are subdivided into antagonistic regions separated by parallel straight lines (**Figure 3B – left panel)**. The optimal stimulus, either a bar or edge, was easily predictable from the geometry of the receptive field, so a stationary line stimulus worked optimally when its boundaries coincided with the boundaries of the subdivisions for this cell type (**Figure 3B – right panel**). Complex cells have linear receptive fields with specific orientation axes too, but their receptive fields are larger and do not have well-defined excitatory and inhibitory zones (**Figure 3C – left panel**), so the exact location of the stimulus does not appear essential. It has been suggested that complex cells are built up from many simple cells (Hubel & Wiesel, 1962; Movshon et al., 1978a). Some complex cells are also direction-selective, in the sense that they respond only when the stimulus moves in one specific direction (**Figure 3C – right panel**). In V1, the different cortical cells that receive their afferents from the same point of the retina have similar receptive fields but different orientation axes. In this way, for each point of the retina, all the orientation axes are presented in the cortex.

The size of a receptive field varies both according to its eccentricity (its position relative to the fovea) and the position of neurons along the visual pathway (Kandel et al., 2013; Nassi & Callaway, 2009). Receptive fields with the same eccentricity are relatively small at early levels in visual processing and become progressively larger at higher levels. The size of the receptive field is expressed in terms of degrees of visual angle (the entire visual field covers nearly 180°). In the early stages of visual processing, the receptive fields near the fovea are the smallest. The receptive fields for retinal ganglion cells that monitor portions of the fovea subtend approximately 0.1°, whereas those in the visual periphery reach up to 10°. V1 neurons are commonly <1° in the fovea, while neurons in the peripheral field representation of some extrastriate visual areas may have receptive fields >100° (Smith et al., 2001; Zeki, 1978).

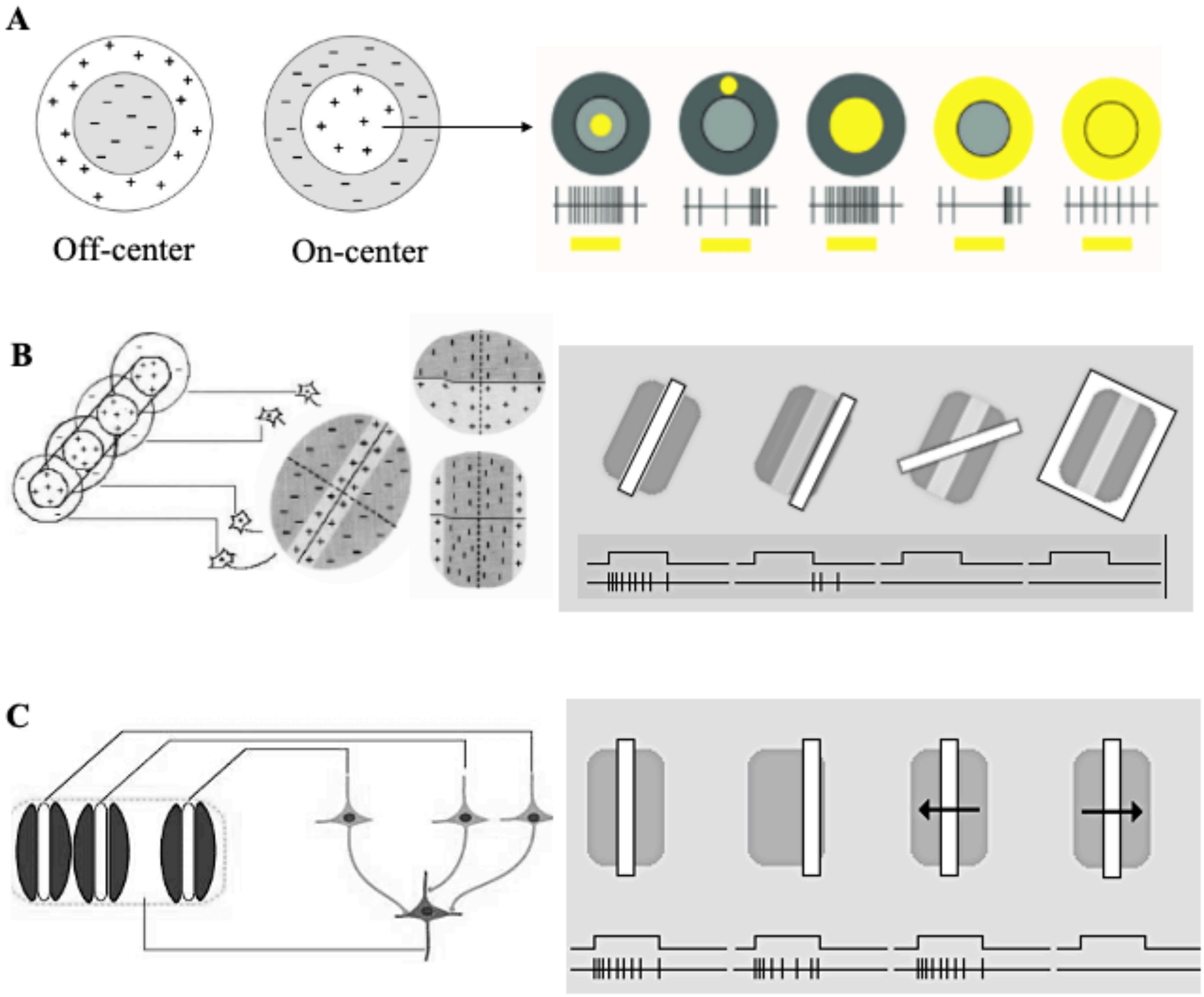

Figure 3. Receptive fields from the retina to V1. (A) Receptive fields of retinal ganglion cells and LGN cells. Left panel: Circular receptive fields characterized by a central inhibitory (center-off) or excitatory (center-on) zone. Right panel: activation of an on-center cell when a spot of light is turned on in its different region. Figure adapted from (Kim et al., 2021). **(B) Receptive fields of simple V1 cells.** Left panel: simple cell's receptive fields built in the cortex by collecting responses from LGN cells. Right panel: activation of a simple cell receptive with excitatory (light grey) and inhibitory sub-regions (dark grey). The horizontal lines indicate the onset and offset of stimulation; the vertical lines indicate nerve impulses. Figure adapted from (Hubel, 1982). **(C) Receptive fields of V1 complex cells.** Left panel: complex cell's receptive fields built from many simple cells. Right panel: complex cells are not fussy about the stimulus position, as long as it falls somewhere inside the receptive field, and some respond to specific motion direction. Figure adapted from (Hubel, 1982).

1.1.2 Visuo-oculomotor system and eye movements

For the purpose of this thesis, it is also important to mention that some subsidiary projections of visual information, separate from the main ascending thalamocortical pathway, also exist. Among these, the retinotectal pathway (**Figure 4**), passing through the Superior Colliculus (SC), in the midbrain, is particularly relevant for active vision (for a review, see Munoz & Everling, 2004). Indeed, the SC plays a major role in the initiation and control of eye movements, and it is traditionally associated with reflexive orienting behaviors. This view, which has been increasingly debated recently (Hall & Moschovakis, 2003), is supported by the observation that the SC in humans receives direct visual input from the retina and directly projects to the brainstem reticular formation that in turn controls the oculomotor motor-neurons, completing a subcortical independent loop (Sadun et al., 1986). This pathway continues to the pontine formation in the brain stem and then to the extraocular motor nuclei. SC also receives information from other cortical areas, particularly the primary visual cortex, the posterior parietal cortex and frontal eye fields (Cerkevich et al., 2014; Glimcher, 2001; Lock et al., 2003; Munoz & Everling, 2004; Schiller, 1984). The cortical input from V1 to SC could be functional to the creation of the visual saliency map. Indeed, as will be discussed in the following chapters, recent data support the hypothesis that, in primates, V1 creates a saliency map from visual input and then the exogenous guidance of attention is realized by the SC, which can select the most salient location as the target of a gaze shift (Zhaoping, 2002, 2016).

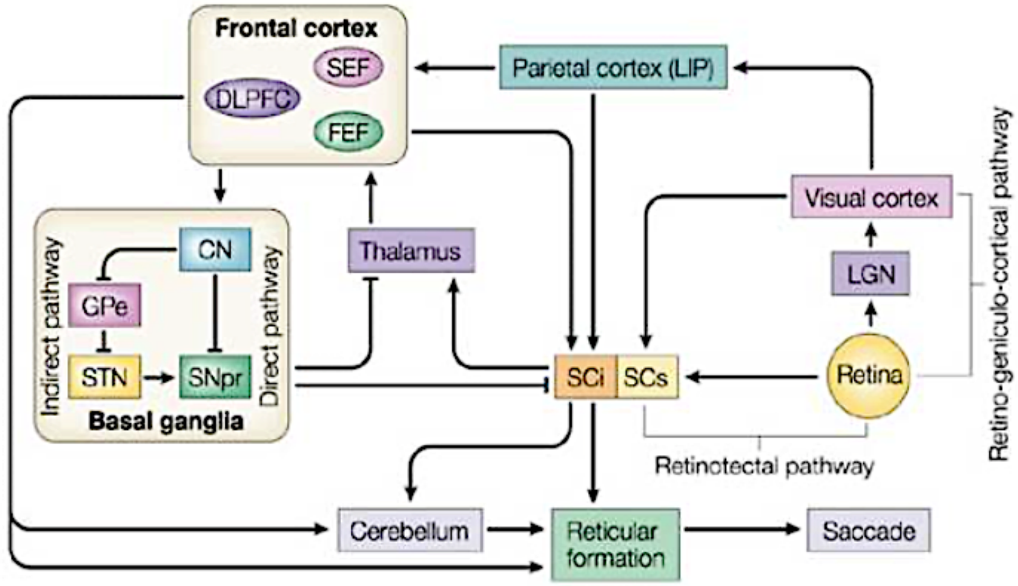

Figure 4. Brain saccadic network. Figure adapted from (Munoz & Everling, 2004).

Humans and non-human primates explore their visual surroundings by means of rapid shifts of eye-gaze, or saccades, in order to align the higher acuity region of the retina with regions of interest in the scene. Therefore, the nature and the quality of the visual information in input depend crucially on the direction of our gaze. In turn, the mechanisms that orient the eyes in space rely on visual input as well as on higher-level cognitive factors (e.g., the intention to look for something in particular). The decision of when and where to make a saccade is behaviorally important and is usually made in the cerebral cortex. Such intimate interaction between a highly sophisticated visual system and an efficient oculomotor system has evolved, for primates, into the unique skills of active vision.

When we explore a new visual scene, or when some changes occur at a given location of our visual field, it is of primary importance to select a maximally informative visual stimulus as the target for the next saccade. Maximization of the rate of incoming visual information is intuitively a main principle to drive efficient oculomotor control (Najemnik & Geisler, 2005, 2008). Trying to achieve this goal by simply increasing the rate of eye movements would not be an efficient strategy, though, since useless or wrong saccades have a cost. In fact, only poor visual information can be acquired during a saccade, due to the phenomenon of saccadic suppression. The accurate selection of the target for gaze re-direction is therefore crucial for vision. In other terms, in ecological situations, an important decision has to be taken regarding the relevance of a particular visual stimulus among many others in the scene.

In the next introductory paragraphs, we will discuss the problem of which parts of the visual scene have to be prioritized for fast and efficient image reconstruction in the brain, which is the starting point of this thesis.

1.2 Fast vision: visual system limitations and the need for data reduction

The visual system needs to analyze the visual scene efficiently in a short time – in the order of ten milliseconds – as fast image recognition is crucial for survival (Hare, 1973). A huge amount of information from the external world is potentially available, at any moment, to the visual system, thus the latter needs to quickly extract the most relevant elements for initiating adaptive behaviors. In fact, rapid and reliable detection of visual stimuli is essential for triggering autonomic responses to emotive stimuli and for orienting towards interesting or potentially dangerous stimuli (Hare, 1973). Previous studies demonstrated that the speed of visual processing is very high; about 100 ms for animals and face processing (Kirchner & Thorpe, 2006), and only 30 ms for images showing affective contents (Whalen et al., 1998).

The amount of information that needs to be processed in that limited amount of time is significant. It has been estimated that the capacity of transmission of photoreceptors in the retina is about 20 Gb/s for each eye, and it drastically decreases at the level of optic nerve fibers (about 4 Gb/s) with a final neural ratio of nearly 124:1 (Echeverri, 2006).

These data must be joined with the limits on the brain's capacity to process visual information. A considerable amount of energy is indeed required to create an accurate representation of the visual scene in the shortest possible time (Lennie, 2003; Levy & Baxter, 1996) largely due to the rate at which neurons produce spikes (Attwell & Laughlin, 2001). All this evidence highlights the existence of an early information bottleneck (Atick, 1992).

Therefore, it is widely believed that the visual system operates a strong data reduction at an early stage of processing (Attneave, 1954; Barlow, 1961), by creating a compact summary of the most relevant information in the input. These relevant features can be then handled by further levels of processing.

1.3 Some models of early visual processing

Early models of fast vision describe the initial stages of visual information processing as the extraction of a "sketch" based on a limited number of ''salient" features' (Marr, 1982; Morgan, 2011). Sketches, containing a drastically reduced amount of information, are then simplified but informative representations of visual scenes.

Past models of efficient coding of information in early vision were based on reducing redundancy (Atick, 1992; Barlow, 1961; Olshausen & Field, 1996, 2004). In the visual system, the images that fall upon the retina when viewing the natural world have a relatively regular statistical structure, which arises from the contiguous structure of objects and surfaces in the environment (Simoncelli & Olshausen, 2001; Simoncelli, 2003). Field (1987) has shown that the receptive field properties of simple-cells in primary visual cortex (V1) are well suited to this structure, in that they produce sparse representations (Field, 1987). Based on these findings, Olshausen and Field proposed that a coding strategy that maximizes sparseness is sufficient to account for all properties of receptive fields of simple cells in mammalian primary visual cortex (Olshausen & Field, 1996, 2004). They showed that a learning algorithm that attempts to find sparse linear codes for natural scenes will develop a complete family of localized, oriented, bandpass receptive fields (**Figure 5**). The resulting sparse image code provides an efficient representation for later stages of processing because it possesses a higher degree of statistical independence among its outputs.

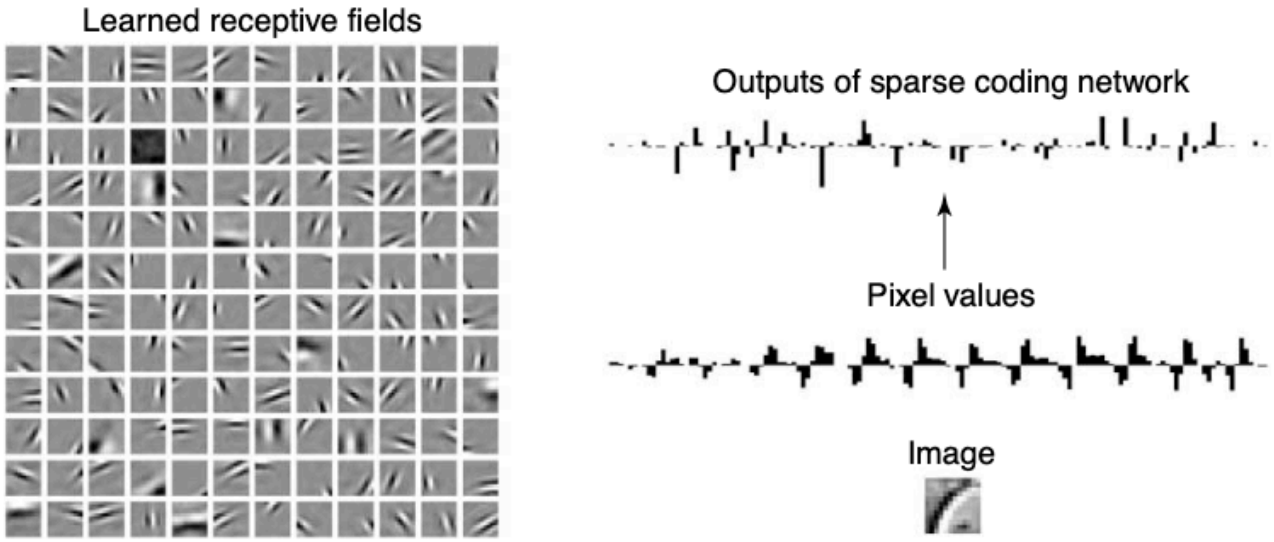

Figure 5. Sparse coding of natural images (Olshausen & Field, 1996, 2004)**.** On the left is represented a set of receptive fields that are learned by maximizing sparseness in the output of a neural network. Each patch shows the receptive field of a model neuron within a 12x12 pixel image patch. The network was trained on approximately half a million image patches extracted from whole images of natural scenes. The receptive fields that emerge from training are spatially localized, oriented, and bandpass (i.e., selective to a spatial structure at a particular scale), similar to cortical simple cells. On the right is represented an example image patch and its encoding by the sparse coding network. The bar chart directly above the image patch shows the pixel values contained in the patch. These input activities are transformed into a much sparser representation in the output of the network, shown in the bar chart at the top. The value of an output unit corresponds (roughly) to the degree of similarity between its receptive field and the input image. As the receptive fields are matched to the structures that typically occur in natural scenes, an image can usually be fully represented using a small number of active units. Figure retrieved from (Olshausen & Field, 2004).

This model, as well as other model based on efficient coding of information by reducing redundancy (Atick, 1992; Barlow, 1961; Olshausen & Field, 1996), take an approach based on preserving the majority of the available information and do not lead to large reduction factors, with the extraction of few salient features.

On the other end, other models of visual feature extraction are not focused on data reduction, whereas they are generally based on the a-posteriori knowledge of specific physiological details (Marr & Hildreth, 1980; Morrone & Burr, 1988; Watt & Morgan, 1983).

For example, Marr and Hildreth (1980) proposed a theory of edge detection starting from the evidence that changes in natural images occur over a wide range of scales, thus any single filter can be simultaneously optimal at all scales. A way of dealing separately with the changes occurring at different scales is taking local averages of the image at various resolutions and then detecting the changes in intensity that occur at each one. Therefore, to realize this idea, they first choose as the optimal filter a Gaussian filter, localized in the spatial domain and with limited bandwidth in the frequency domain. The first step in this model is to filter the image with an appropriate Gaussian filter (**Figure 6A** and **6B**). Secondly, to detect the intensity changes, authors considered that, wherever an intensity change occurs, there will be a corresponding peak in the first directional derivative, or equivalently, a zero-crossing in the second directional derivative of intensity. Therefore, the task of detecting these changes can be reduced to that of finding the zero-crossings of the second derivative D^2 of intensity, in the appropriate direction by means of the Laplacian ∇^2 operator, which allows for finding the edges of the image (**Figure 6C** and **6D)**. To detect changes at all scales, it is necessary only to add other channels, and to carry out the same computation in each. Intensity changes in images arise from surface discontinuities, reflections, or illumination boundaries, and all these objects have the property of being spatially localized. For this reason, the zero-crossing segments of the different channels do not appear to be independent, and rules can be deduced to combine them in an image description. In particular, to combine information from different channels, it is essential to ensure that the zero-crossings from independent channels of similar size coincide (spatial coincidence assumption). If this were not the case, they would probably be due to distinct physical surfaces or phenomena. Therefore, coincidence of zero-crossings across scales provides the basis for a schematic description of the image (the "primal sketch"). It follows that the minimum number of channels required is two, and that, assuming that the two channels are reasonably separated in the frequency domain, and their zero-crossings agree, they can be taken to indicate the presence of a border in the image.

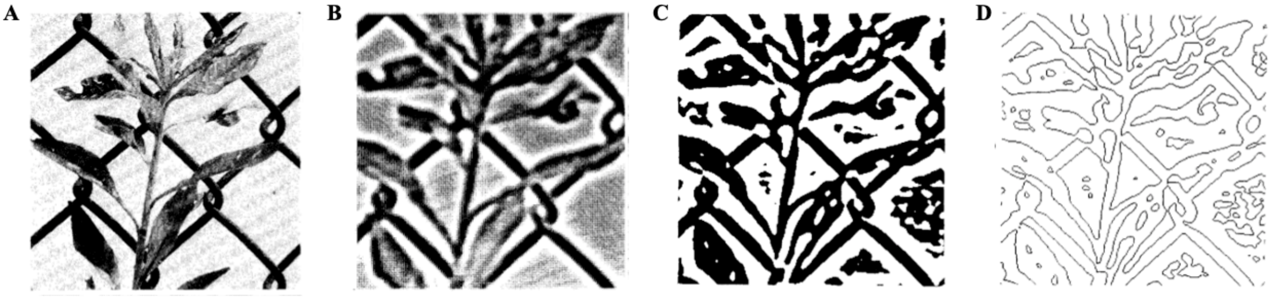

Figure 6. Example of edge detection from Marr and Hildreth model (1980). The image in **(A)** has been filtered in **(B)** with the ∇^2G filter (zero is intermediate grey), then in **(C)** positive values are shown with white and negative with black, and finally in **(D)** only zero-crossing segments appear. Figure adapted from (Marr & Hildreth, 1980)**.**

A critique of this model has been raised by Morrone and Burr (1988), arguing that it is designed to detect edges, not lines, even though these are salient features too. The authors claimed that the requirement for the coincidence of zero-crossings across scales eliminates the inappropriate marking of lines, which occur at different positions at different scales, but can also eliminate real edges under certain conditions. One example is when two edges occur nearby: zero-crossings at larger scales will occur midway between the edges. Their model of feature detection is defined "phase-dependent energy model" (Morrone & Burr, 1988) and it is based on the evidence showing that visual detectors in V1 have even- and odd-symmetric receptive fields (Hubel, 1982; see Figure 3B). The authors started from a new definition of lines and edges which consider their local Fourier representation (Morrone & Burr, 1988). They suggested that the points in a waveform that have unique perceptual significance as "lines" and "edges" are the points where the Fourier components of the waveform come into phase with each other. The basic operators of the model are pairs of filters of equal amplitude spectra but orthogonal in phase: one filter type has an even-symmetric line-spread function (i.e., a Fourier phase spectrum of 0 – cosine phase), the other an odd-symmetric line-spread function. (i.e., a Fourier phase spectrum of π /2 – sine phase). A useful property of lines and edges is that they occur at points of the waveform where the arrival phases of the Fourier components are maximally similar. The authors then suggested that the visual system could locate features of interest by searching for maxima of local energy. The local energy function locates the position of image features, both edges and lines, but gives no information about the type of feature. To identify the feature type, the system has to evaluate the value of the average arrival phase at that point, which determines the nature of the feature: values near zero correspond to a line, and values near π /2 correspond to an edge (i.e., a response from filters with even-symmetric fields will signal a line; a response from filters with odd-symmetric fields will signal an edge). If both filter types respond at the peak of local energy, both edges and lines are seen, either simultaneously or alternating in time. The model was tested with a series of images and shown to predict well the position of perceived features and the organization of the images.

Overall, these type of models successfully describes how the visual system extracts salient features but they are based on the a-posteriori knowledge of specific physiological details, rather than on considerations of information compression efficiency (Marr & Hildreth, 1980; Morrone & Burr, 1988; Watt & Morgan, 1983), therefore they have a reduced predictive power.

In the next paragraph, I'll present a recent model of early visual feature extraction, aimed at significantly reducing information through the selection of salient features (Del Viva et al., 2013) without assuming the known properties of underlying physiological mechanisms.

1.4 Constrained-maximum entropy model of early visual features extraction

1.4.1 The assumptions of the reference model

The reference model of this thesis has been formulated by Del Viva, Punzi, and Benedetti in 2013 (Del Viva et al., 2013). It stems from the problem discussed in the previous paragraphs that can be summarized as follow: the extraction of biologically-relevant information from a large flux of input data in the shortest possible time for survival purposes.

The need for extracting a small amount of ‘‘relevant’’ information from a large input flux of data is not unique to vision (Ristori & Punzi, 2010; Smith & Lewicki, 2006), although vision may be one of the fields where the requirements are particularly severe. The model can be indeed applied not only to vision but to any information processing system (natural or artificial) that has to reduce input information within precise computational constraints while transmitting into the output as much information as possible. Only the application of the model to vision will be discussed in this thesis.

The model is based on some very general assumptions and aimed explicitly at reducing information through the selection of salient features (**Figure 7A**). First, the model assumes that, at an early stage, the reduction of the huge input data flow is achieved by filtering only those pieces of input data matching a reference set of features, disregarding any other information (*pattern matching*). Second, there is only a fixed number of visual features that the system can recognize in input (*limited capacity*). Third, the model imposes a tight upper bound on the total amount of data that can be produced as output to be transmitted to the next stages of processing (*fixed output bandwidth*). Finally, the system is optimized to preserve the maximum amount of information (*maximum entropy output*).

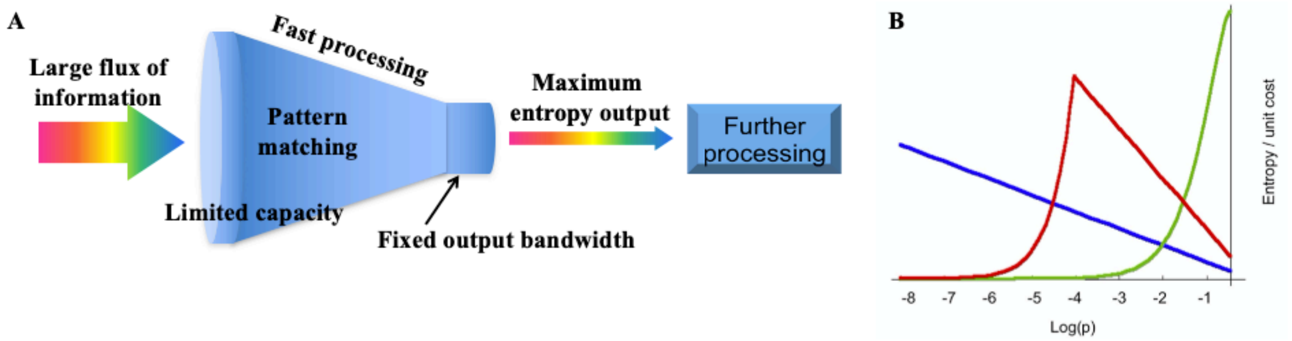

Figure 7. The reference model of data reduction by Del Viva et al. (2013). (A) Schematic representation of the information filter proposed by the reference model. The visual system acts as a filter that recognizes and selects the meaningful features of the input, dropping all the remaining information. The number of recognizable features in the input is limited and the information in the input has to be transmitted to the next processing stages with minimal energy consumption. The system must be optimized to produce the maximum entropy output. **(B) Entropy yield per unit cost plotted as a function of the probability of occurrence in the input of each feature.** Green curve: limited storage and unlimited bandwidth ($N=100$, $W=\infty$); Blue curve: limited bandwidth and unlimited pattern storage capacity ($W=0.001$, $N=\infty$); Red curve: limited bandwidth and storage ($N = 100$, $W = 0.001$). Parameter values and the vertical scale are arbitrarily chosen for illustration. Figure retrieved from (Del Viva et al., 2013).

The authors specified that the model is purely functional, and do not concern with the details of how this computation is implemented. The functionality of this abstract pattern-filtering model is completely defined by its reference set of features. This also means that the model discussion does not need to be concerned with the specific computation used in their recognition, nor with its localization within any specific anatomical structure.

In order to test the prediction of the model, the reference set of visual features must be determined precisely. One problem is that the number of possible a-priori choices for the reference set of visual features is very large. As already discussed, possible approaches, often used in developing similar models in the literature, are based on known properties of neuron receptive fields, or on considerations of performance in the reconstruction of visual scenes (Marr & Hildreth, 1980; Morrone & Burr, 1988). The authors choose instead a different approach, focusing only on the requirement that the set of features must be information-efficient. That is, they assumed that the system is optimal from the point of view of delivering the maximum amount of information to the following processing stages. Therefore, they choose the feature set producing the largest amount of entropy allowed by the given limitations of the system.

Adopting the principle of maximum entropy as a measure of optimization together with the imposed strict limitations to the computing resources of the system, allowed them to completely determine the choice of the feature set from the knowledge of the statistical properties of the input data.

In the next paragraphs, the derivation of the general model function will be described in detail.

1.4.2 The selection function

Let p_i be the probability that a given portion of the input data matches a specific pattern *i*, out of a set *Q* of mutually exclusive patterns, such that $\sum p_{\mathrm{i}} = 1$, when *i* runs over all *Q*. The pattern-recognition system can be thought of as an array of *N* pattern-matching elements, each of them capable of recognizing the occurrence of the single pattern *i*, providing a single output bit, that signals the presence of the pattern in the input. This system would produce, on average, an information output equal to $-p_i log(p_i)$ – that is, it is a source of entropy $-p_i log(p_i)$. Neglecting correlations, the total entropy of the system is simply $\sum^{N}_{i} -p_i log(p_i)$. In absence of other constraints, maximization of the total entropy would be attained by simply including all possible patterns. This solution would imply transferring to the output the whole information in the original input, with just a change of format.

The key to a meaningful answer is the explicit inclusion of the limitations of the system. Let's assume that the system can recognize up to a maximum number *N* of distinct patterns; to obtain the maximum entropy output under this constraint, patterns should be chosen to maximize the function $-p_i log(p_i)$. This is a large probability, and in practice, it is likely to lead to selecting the patterns with the highest probability of occurrence in the input (see green line in **Figure 7B**).

However, the output flux of the system is also bounded, due to bandwidth limitations, and the choice of the most probable patterns could quickly exceed this limit. On the other end, if the system would have limited bandwidth but unlimited pattern storage capacity, the pattern selected would be the rarest (see blue line in **Figure 7B**).

In order to account for both constraints, a "worst-case" cost has been associated to each pattern, defined as the larger of the "storage cost" 1/N and the "bandwidth cost" $\mathrm{p_i}$/W, where W is the maximum allowed total rate of pattern acceptance, $\Sigma\, p_i < W$. Therefore, an entropy yield per unit cost is given for each pattern by:

$$f(p) = \frac{-p \log(p)}{\max(1/N,\ p/W)}$$

The optimal performance of the filtering system is then attained by choosing the set of patterns such that $f(p_i)>c$, where c is determined by the computational limitations: $\int_{f(p)>c} \delta(p)\mathrm{d}p < N$ and $\frac{1}{N_{tot}} \int_{f(p)>c} p\delta(p)\mathrm{d}p < W$, where $\delta(p)$ is the density of patterns having the probability of occurrence p, normalized to the total number N_{tot} of patterns in Q. The quantity $\frac{1}{N_{tot}} \int_{f(p)>c} p\delta(p)\mathrm{d}p$ is the average fraction of image elements that match successfully and get preserved in the output - its inverse is the *compression factor* achieved by the filtering algorithm.

It is then obtained an unambiguous and general heuristic recipe to determine the set of *optimal* patterns that a generic pattern-filtering system should use in order to achieve maximum information preservation under the given constraints.

As can be seen in **Figure 7B** (red curve), the function *f(p)* presents a rather sharp maximum; the set of *optimal* features will be thus concentrated in a limited range of values of *p* around the maximum of *f(p)*, which occurs at $p = W/N$; it will therefore depend on both available storage size and bandwidth.

1.4.3 Extraction of *optimal* visual features from natural image statistics

To apply the model to vision, authors considered the simplest possible set Q of base features, defined as all possible configurations of 3*3 square pixel matrices in black-and-white images (1-bit depth). The reduction of input images to only two levels is a corollary of the central idea of compression by pattern filtering proposed by the model (Del Viva et al., 2013): the number of possible patterns, assumed to be a limited resource, increases exponentially with the number of allowed levels (that is 2^{n*N} where n is the number of bits and N the number of pixels) – and so does the amount of computing needed to calculate them. Therefore, using a large number of grey levels in the model would be not only unpractical but also would defeat its very purpose of saving computational resources. For the same reason, the authors chose to implement the model by defining as a feature a 3x3-pixel image partition.

A public database of 560 calibrated natural pictures has been used (Olmos & Kingdom, 2004), and each image (768x576 pixel) was digitized to 1-bit luminance (black/white), by setting the threshold at its median luminance value (see **Figure 8A**; for more examples see Figure 9A). The probability distribution of all the 3x3 pixel patterns has been calculated and then the set of *optimal* features has been identified following the model function (**Figure 8B**). For the extraction of a set of *optimal* features, algorithm parameters were N=50 and W=0.05 (**Figure 8C)**. These parameters were chosen based on the consideration that the algorithm relies on the idea of a strong compression at the minimum possible computational price. Thus, they considered as a reasonable upper bound to N a value of 10% of all possible distinct patterns and picked 50 features over the total 512 distinct features possible in the basic 3*3 model; and they considered a compression factor of at least 20 setting W=0.05 as a constraint. This set of features (**Figure 8C)** has been used as the *optimal* features set in the experiments described in the following chapters.

In the figure below, other sets of features used in the original work are also shown (**Figure 8D, Figure 8E,** and **Figure 8F**). In the experiments carried out in our studies and described in the following chapters, the set of features shown in Figure 8E was also used. Indeed, these 50 specific features have been used as *non-optimal* features and compared to *optimal* features with many different paradigms. This set included the features with the lowest probability in the statistical distribution of all possible features, those that are discarded by the constrained-maximum entropy model due to large storage occupation.

By looking at Figure 8, it turns out that about 70% of the *optimal* features (Figure 8C) selected by the algorithm can be classified as edges, bars, or end-stops, of various orientations as found in primary visual areas (Hubel, 1982; Hubel & Wiesel, 1968, 1962, 1959; see Figure 3B). Others are interpretable as corner detectors. Conversely, most of the *non-optimal* features discarded by the model selection have either an irregular structure resembling visual noise (Figure 8E), or uniform luminance (Figure 8F), with lower resemblance to known visual features. Thus, the biologically plausible structure of the features seems to derive from very general principles of information maximization and computational limitations.

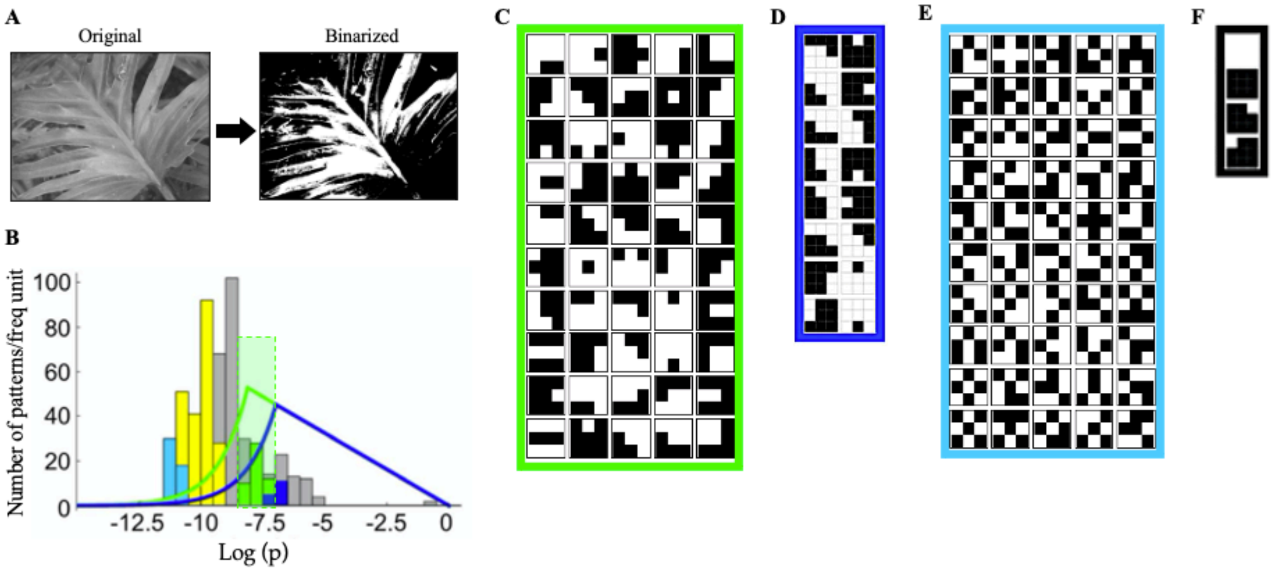

Figure 8. Visual features extraction from natural images. (A) Example of images. One example of an original 256 grey-levels image (Olmos & Kingdom, 2004) digitized into black and white. **(B) Selection of *optimal* features.** The histogram shows the probability distribution of the 512 possible 3x3 1-bit pixel patterns. The curves are the model selection functions for W=0.05 and two different values of N - green: N= 50 (*optimal* features in (C)), blue: N= 15 (*optimal* features in (D)). Green and blue histograms are the probability distributions of corresponding selected patterns. Cyan and yellow histograms are the distributions of low-probability patterns. **(C-D) *Optimal* features.** Visualization of 50 and 15 *optimal* features (3X3-pixels) selected by the model function (green and blue curves in (B)). **(E) *Non-optimal* features.** Visualization of 50 *non-optimal* features (corresponding to cyan histogram in (B)). These features are those with the lowest probability of occurrence. **(F) Highest-probability features.** Visualization of the four features with the highest probability of occurrence. Figure adapted from (Del Viva et al., 2013).

1.4.4 Human contrast sensitivity to *optimal* features

To obtain direct evidence that the human visual system assigns a privileged role to model-predicted *optimal* features in image-reconstruction processing, the authors first performed psychophysical measurements of contrast sensitivities for the detection of single isolated 3X3 pixel features.

Human contrast sensitivity was measured for all possible 1-bit 3X3 pixel features, scanning the entire range of probabilities found in natural images, with a 2IFC procedure. In each trial, participants were required to indicate the interval containing the feature (n = 3, trials = 300). The contrast of the features was randomly chosen from trial to trial within a set of predetermined values in the range 0.01 to 0.22. In this experiment such small features subtended about 6X6 min of arc, allowing to target early visual processing stages. These are very likely the anatomical substrate of the hypothesized filter because data compression must be done very early in the visual stream to be effective. Although early visual structures comprise multiple cell types, with different receptive field sizes (Nassi & Callaway, 2009), here, for simplicity, a single small scale is considered. However, this small scale is consistent with receptive field sizes found in human V1, which are about 15’ in the fovea (Smith et al., 2001) and become progressively larger with eccentricity and through the hierarchy of visual areas (Zeki, 1978).

The results showed that observers’ contrast sensitivity peaks within a limited probability range, corresponding to the probability of occurrence of *optimal* features, in agreement with the predictions of the model.

The authors also specified that these results are not simply a consequence of the band-pass behavior of the human contrast sensitivity as a function of spatial frequency. They claimed that, although there is a mild correlation between the probability of features occurrence and their spatial frequency content, for which the rarer patterns contain on average more of the higher spatial frequencies, there is a significant overlap of spatial frequency content between the patterns over the whole range of probability. Also, at any rate, the spatial frequency of all features is comprised between 9 cycles/deg and 27 cycles/deg, while the maximum sensitivity lies at about 7 cycles/deg in their illumination conditions. Therefore, the spatial frequency spectrum lies entirely above the frequency of maximum human sensitivity. Spatial frequency sensitivity considerations would instead predict a low-pass behavior, resulting in an increasing function with probability, which is very different from what has been observed.

For the purposes of the studies presented in the next chapters, in which *optimal* (Figure 8C) and *non-optimal* (Figure 8E) features are used individually as stimuli, we performed a quantitative analysis of the differences in spectral properties between the two sets (see **Supplementary Material**).

1.4.5 Discrimination of images based on *sketches*

The model not only predicts enhanced sensitivity to *optimal* features, that was verified, but also predicts that early visual processing utilizes only the parts of the images that match the reference set of *optimal* features to create a compressed but informative internal representation of the scene. To test this prediction, the authors created *sketches* from the images, by keeping only those features of the binarized image matching one of the features of the reference set (Figure 8C), blanking all other parts. Sketches were prepared from the same images database used in determining the reference features set (**Figure 9A, 9B**). These sketches represent the prediction of the output of the early visual processing stage that the model trying to simulate (**Figure 9C**).

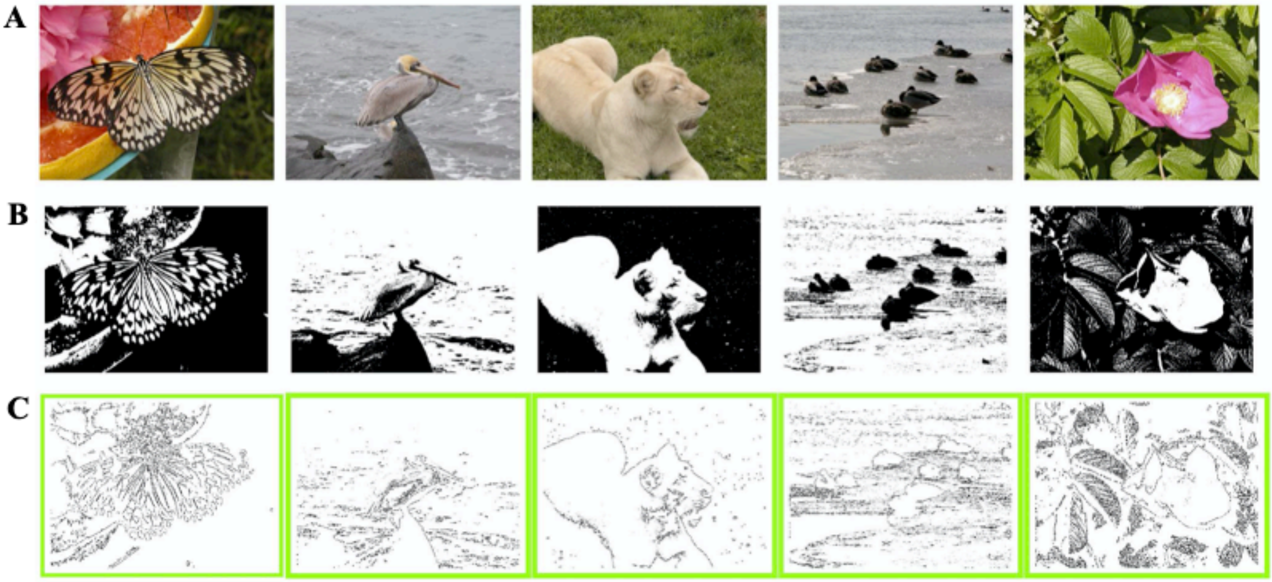

Figure 9. Examples of images and their *sketches* obtained with *optimal* features. (A) Full-color images. Natural images from the reference database (Olmos & Kingdom, 2004). **(B) Digitized images.** Digitized versions of images in (A). **(C) *Sketches.*** Sketches obtained from the images in (B), using the *optimal* features set of Figure 8C. Figure adapted from (Del Viva et al., 2013).

By a close inspection of the *sketches* obtained with the *optimal* features sets clearly emerged that they retain most of the salient features of originals, in spite of a substantial reduction of information (Figure 9C).

To quantify this qualitative observation, the authors measured the effectiveness of these sketches in allowing human observers to identify natural images under fast viewing conditions. If the early visual system really selects only those specific *optimal* features for its processing, then the *sketches* should elicit nearly the same response as complete images. Specifically, these "salient" sketches were used as stimuli in a discrimination experiment (2AFC paradigm), where they were shown for only 20 ms, in order to probe the early stages of visual analysis (Thorpe et al., 1996). The stimulus was followed by a random noise mask (500 ms); the backward masking paradigm is used to prevent iconic memory and interrupt further analysis of briefly presented stimuli by neural structures (Enns & Lollo, 2000; Herzog, 2016; Macknik & Martinez-Conde, 2004). Subsequently, two digitized images were presented side-by-side for 700 ms, one of them being the unfiltered image corresponding to the sketch, and the other a distractor, randomly selected from the dataset (**Figure 10A**). The observers were asked to identify the correct match between the sketch and the original image. As a control, the full image with 256 grey levels instead of the sketch was shown for 20 ms (**Figure 10B**), and the same image and a distractor, also with 256 grey levels, were shown in the task. Other sketches obtained varying the model function parameters were also used in this task: sketches built based on alternative, *non-optimal* features sets, and sketches obtained with different parameter values for bandwidth and number of reference features.

The results of the discrimination task are reported in **Figure 10E**. All observers were able to identify the original images from which the *optimal* sketches were extracted with extremely high accuracy (green bars). Even more important, performance was comparable to measurements obtained in the control experiment using the fully detailed original images in place of their sketches (red bars). Participants reported that they could not tell whether originals or sketches had been shown to them in these fast presentations. Similar results were obtained by using *optimal* sketches obtained with different model parameters (**Figure 10A** and **Figure 10C**; green vs blue bars).

Instead, the ability of observers to identify original images based on alternative sketches, built based on *non-optimal* features sets, was much worse than with *optimal* sketches even if both have the same information (yellow bars vs. blue bars). However, distributions of the number of points found in the two sets of sketches (**Figure 10C and Figure 10D**), taken over the whole image database have different average values: ~14000 for the alternative set and ~24000 for the set predicted by the model (**Figure 10F**). Therefore, an additional test has been performed to exclude that the observed difference in average performance might be due to the difference in the average number of visible points. For each experimental trial, authors reweighted in the final average the data taken with the pattern set predicted by the model by a factor equal to the ratio of the probability distributions of the two sets. In this way, the density distribution of the *optimal* features is forced to match that of the *non-optimal* rarer features, and any possible dependence of the result on the density of the image gets equalized between the two sets. The results show that the reweighting procedure has no significant effect, only shifting the results by less than one standard deviation (Figure 10E, striped-blu bars). To further investigate the issue, authors replotted the data splitting the trials into different sets, according to classes defined by the number of points in the sketches (**Figure 10G**). The difference in discrimination performance between the two sets is present over the whole range: even densely-populated sketches made *of non-optimal* features are less visible than those from the standard set confirming that the number of displayed points plays no measurable role in the measurements.

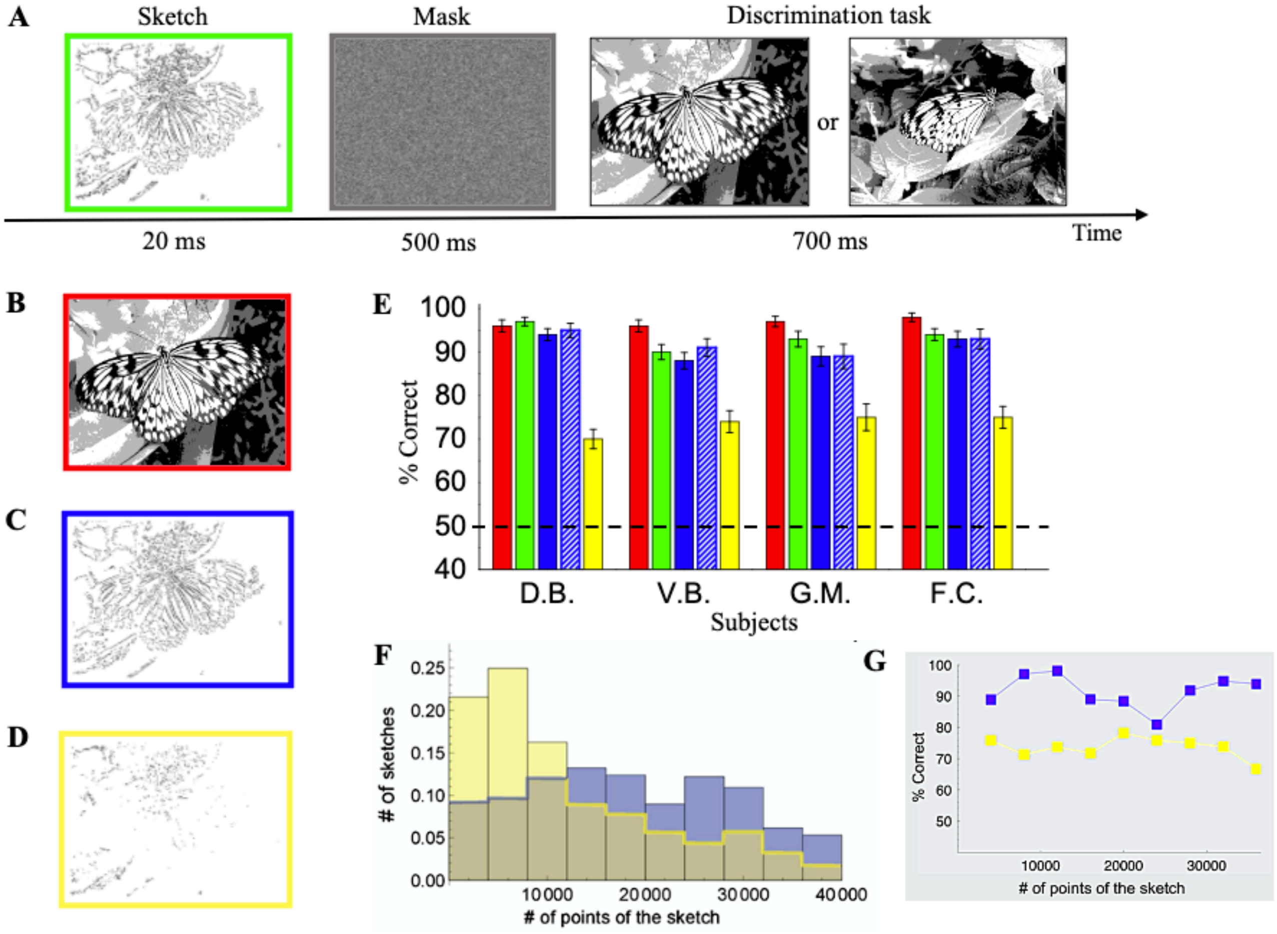

Figure 10. Images discrimination based on sketches obtained from different features sets. (A) Representation of the experimental procedure (2AFC). The sketch in (A) is obtained from the *optimal* features set of Figure 8C. The corresponding compression factor is 40, and its information content is 9.8% of the original. **(B)** Original 256 grey-levels image. **(C)** Sketch obtained from the *optimal* features set of Figure 8D. The corresponding compression factor is 67 and its information content is 5.5% of the original. **(D)** Sketch obtained from the 244 low-probability patterns set (Figure 8E shows a sub-sample); information (5.5 %) and compression (factor 90) are similar to (C). **(E)** Results of the discrimination task. Percentage of correct discrimination for sketches obtained as in (A), (C), (D) (green, blue, yellow bars respectively) and 256 grey-levels images as controls (red bars), for four observers. The striped-blue bars represent results obtained from the same dataset shown in blue, after reweighting the data to match the distribution of the number of patterns of the yellow dataset. Each data point represents 300 trials. The black dashed line indicates chance performance. Error bars are SD. **(F)** Distributions of the number of points found in the two sets in (C) and (D). **(G)** Percentages of correct discrimination plotted as a function of the number of matched patterns, for the same data as in Figure 10C and 10D. Figure adapted from (Del Viva et al., 2013).

Overall, the results of this psychophysics experiment support the prediction of the model, for which the features identified in natural images through constrained-maximum entropy criteria carry most of the information that the visual system needs for image discrimination under fast viewing conditions.

1.4.6 Effect of local features

The results discussed so far show that in fast vision it is sufficient to see very few details to discriminate images, provided that these few features are "the right ones". In the sketches, *optimal* features turn out to be arranged along objects' contours (**Figure 11A**) rather than being scattered throughout the image, and the spatial structure of the features belonging to a particular contour corresponds to the nature and orientation of the contour. The discrimination power provided by the sketches could be due either to the specific local features used in the sketch or to their global spatial arrangement in the images; this issue has been investigated in a further experiment (Del Viva et al., 2016). The contribution of individual local *optimal* features, located along with objects' contours (global features), has been studied by replacing them with *non-optimal* carriers of information, keeping their localization along the contours unchanged. That is, for example, small local vertical edges in a vertical contour were replaced with different *non-optimal* local features (**Figure 11B**). The results of a discrimination task showed that the disruption of these *optimal* local cues causes a decrease in image recognizability, despite its global structure was preserved (**Figure 11C**).

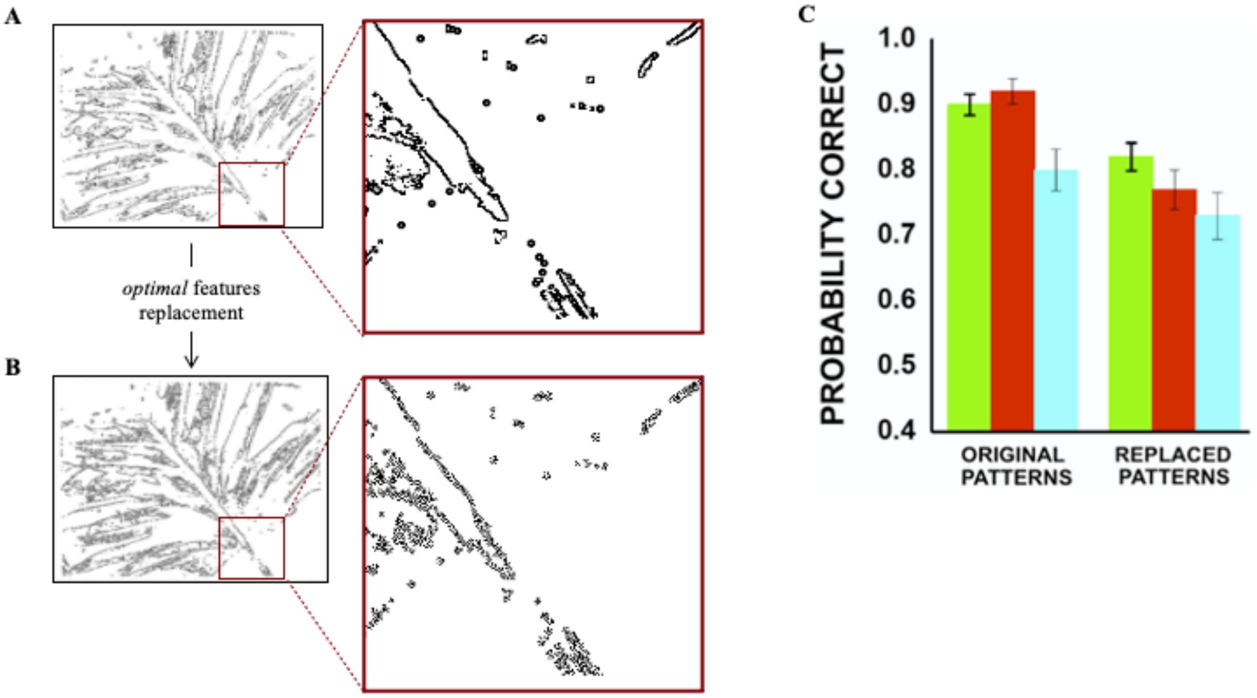

Figure 11. Contribution of local *optimal* features. (A) Sketch obtained by filtering the image with *optimal* features in Figure 8C. All the *optimal* features are positioned along image contours, as shown in the inset. **(B)** Sketch obtained by replacing the *optimal* features with *non-optimal* features. The spatial structure of the features along the contours does not correspond to the orientation of the contour, as shown in the inset. **(C)** Discrimination of sketches obtained with *optimal* features (left bars) and of sketches where *optimal* features were replaced by randomly chosen *non-optimal* features (right bars). Colors represent different observers. Figure adapted from (Del Viva et al., 2016).

1.4.7 Chromatic information and feature detection in fast visual analysis

As discussed so far, the visual system is able to recognize a scene based on a sketch made of very simple features; an open question is the role played by color in this process. Indeed, sketches have been most often represented and discussed only with monochromatic information (Del Viva et al., 2013; Marr, 1982), whereas natural scenes have extensive chromatic content that is a rich source of potentially useful information (Párraga et al., 1998). In a further study, Del Viva and colleagues (Del Viva et al., 2016) approached this question from the perspective of optimal information processing by a system endowed with limited computational resources.

Several past studies have explored the mechanisms of fast vision at different scales and stimulus durations, finding that both coarse and fine spatial information are simultaneously used in fast categorization of images (Oliva & Schyns, 1997b; Schyns & Oliva, 1999). Some of these models build a bottom-up saliency map, based on concurrent simultaneous processing of color with other modalities at multiple spatial scales, that is used to drive visual attention to potentially interesting image locations (Itti et al., 1998; Itti & Koch, 2001; Parkhurst et al., 2002; Torralba, 2003). In the first step of these models, visual input is decomposed into sets of different topographic feature maps (color, motion, orientation, etc.) at various scales. Within each map, spatial locations compete for saliency, and subsequently, these conspicuity single-modality maps are summed into a single master saliency map. Each of these parallel processes requires a certain amount of computing power; however, the required amount varies greatly amongst scales and modalities, and implementation details might not be necessarily the same for each of them. Computational limitations are expected to play the most important role in determining the features analyzed at the finest visible spatial scales, even more so for color. This is a direct consequence of the properties of the Fourier transform, where the information content is proportional to the square of spatial frequency.

While other studies have explored the role of color in fast vision either in complete images (Delorme et al., 2000; Gegenfurtner & Rieger, 2000) or at coarse spatial scale (Oliva & Schyns, 2000), it is not clear that there is any role of color at the finer scales. For this reason, the authors use their model, which is focused on computational cost, as a tool to explore the question of the potential role of fine-scale, information-rich color features in a context of competition among various types of available information for use by a limited capacity resource. Particularly, they evaluated the relative merits of luminance and color information as carriers of information in sketches of natural images, prepared under equal computational constraints, and measured the corresponding discrimination performance on human observers (Del Viva et al., 2016).

1.4.7.1 Discrimination of images based on luminance and color sketches

Natural images were selected from the same public database used in the previous experiment (Olmos & Kingdom, 2004; **Figure 12A**) and digitized to 1 or 2 bits. As already stated before, the need for such strong reduction of levels is a corollary of the central idea of compression by pattern filtering. This issue is even more severe for color information, which carries three times more bits than an achromatic luminance image. Therefore, using a larger number of levels in the model is not only unpractical but defeats its very purpose of saving computational resources. Also, it must be considered that this does not amount to an important limitation for applications within the field of fast vision: the frequency of neuronal discharge is indeed limited to ~500Hz, which means that in 20 ms only a very small number of spikes (~3-4) can physically be transmitted over each individual axon. Considering that the authors use pretty small features (6'X6', close to the resolution limit), this means that a very limited number of spikes are available to encode the intensity level of each signal. Even under ideal conditions, this very fact already limits the available information to very few bits.

The extraction of visual filters and preparation of sketches followed a similar procedure as in the previous experiment (see **Figure 12** for the case of 1-bit digitization; see **Figure 13** for the case of 2-bit digitization). Then sketches containing just one bit of luminance (**Figure 12B** and **12D**) were compared to 1-bit color information sketches (**Figure 12C** and **12E**). Computations revealed that the average information preserved in the image set by the pure-luminance filters is greater than that preserved by pure-color filters (**Figure 12F**). Since the same output capacity constraint was imposed in both conditions, color features turn out to be less effective in conveying information than luminance features when a strong compression is imposed.

To test how this difference in information content reflects in the visual discriminability of the images, a psychophysical study of image discrimination (2AFC) was performed based on these sketches. The results showed that 1-bit gray-scale sketches yielded very good discriminability of the original images. Color-only (equiluminant) sketches, on the other hand, yielded far worse discrimination than luminance-only sketches, consistent with chance performance, hence failing to show any evidence that the observers were able to use color-only information for a fast discrimination task (**Figure 12G**).

Note that these results cannot be explained only by the lower average information content of color sketches. Figure 12F indeed shows that the distribution of information content of the two types of sketches, although different on average, is largely overlapping. By separating into classes images having the same information content, clearly emerged that the response of the human visual system to the two types of sketches is completely different, even when they are compared on the basis of equivalent information content. Indeed, for color sketches, the discriminability is compatible with chance level, independently of the information content; for luminance-based sketches, the discriminability is constantly above chance over the whole range (**Figure 12H**).

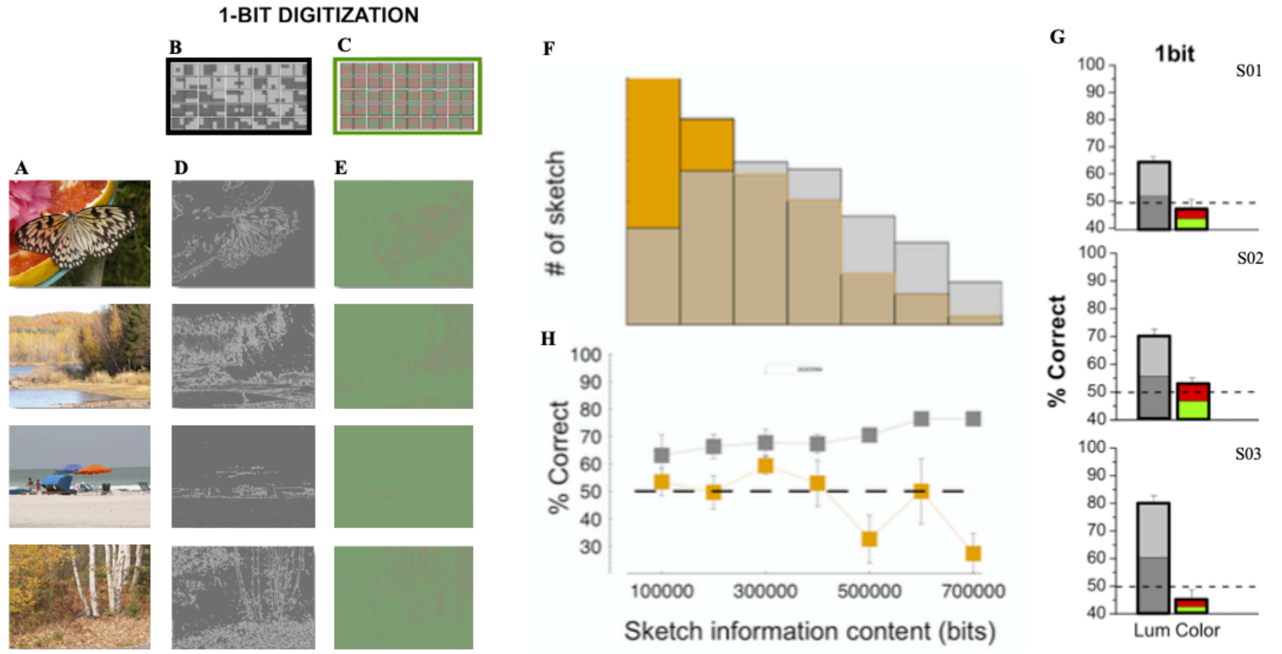

Figure 12. 1-bit digitization: visual filters, sketches, and image discrimination performance. (A) Examples of RGB images. (B) 1-bit luminance features. Features obtained after digitizing to 1 bit the *luminance* (L+M) coordinate. **(C) 1-bit color features.** Features obtained after digitizing to one bit the *l* (L/(L+M)) coordinate. **(D) Two gray-levels luminance sketches.** Sketches obtained with the features in (b). **(E) Color-only sketches.** Sketches obtained with the features in (c). **(F) Information content of 1-bit sketches.** Distributions of the information content of the 1-bit color (orange) and luminance (gray) sketches. **(G) Discrimination of images based on luminance and color sketches.** Percentage of correct discrimination with 1-bit luminance-only sketches (2 gray-levels bars) vs. 1-bit equiluminant sketches (2 red/green-levels bars). **(H) Discrimination as a function of sketches information content.** Percentage of correct discrimination plotted as a function of the information content of the color only and luminance sketches, for the same data as in (G). Figure adapted from (Del Viva et al., 2016).

All of the above observations are compatible with the visual system having made a well-defined choice in favor of using luminance-based features and ignoring color-based features, when considering fine spatial scales. These behavioral results are consistent with a human visual system that, under the pressure for optimization to use limited resources, follows the maximum-entropy principle. Maximum entropy, together with natural image statistics, dictates that luminance information is the privileged vehicle for quick image discrimination, at the expense of other potential sources of information.

This, however, does not exclude the possibility of color information being an important complementary source in addition to luminance information. To test for this possibility, a further experiment compared images constructed with 2 bits (4 levels) of luminance-only information (**Figure 13A** and **13C**) with images constructed with 1 luminance bit and an additional color bit (2 bits in all) (**Figure 13B** and **13D**). In spite of the fact that the compression requirements of the bottleneck were the same in both of these conditions, the entropy in the output sketches was, again, very different for the two conditions: luminance-only sketches contained, on average, 2.3 times more information than color plus luminance sketches. Corresponding psychophysical results showed that the addition of 1 color bit to the luminance bit did not lead to a reliable increase in performance over the use of 1 luminance bit alone (**Figure 13E)**.

Overall, these results suggest a strong preference for luminance-based features over color-based features under fast visual analysis. In sum, this study suggests that the computational limitations of the visual system have led to a system that at the finest spatial scales relies mostly on luminance, rather than color, for fast visual discrimination.

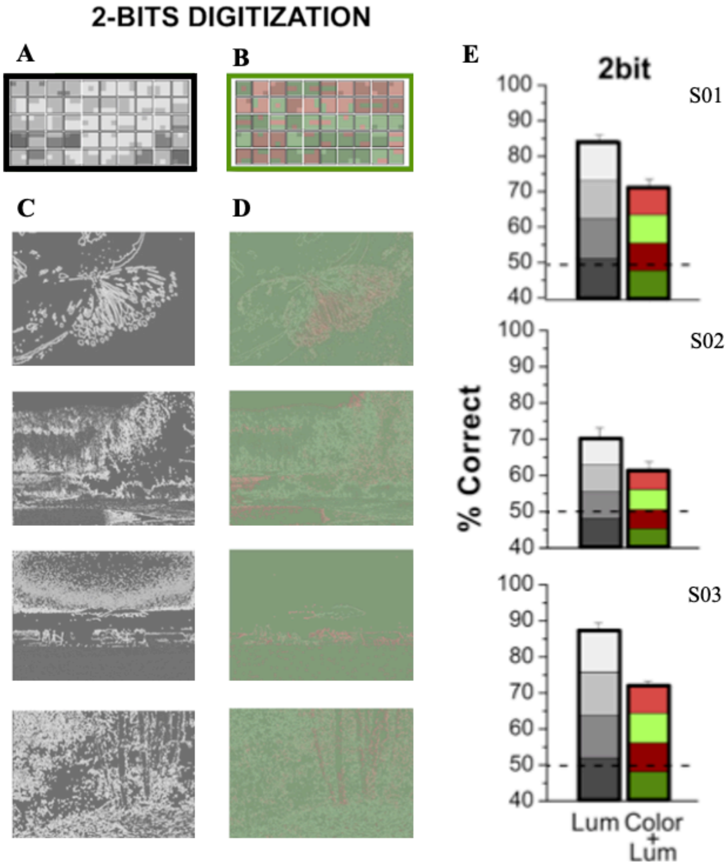

Figure 13. 2-bit digitization: visual filters, sketches, and image discrimination performance. (A) 2-bit luminance features. Features obtained after digitizing to 2 bit the luminance coordinate. **(B) 1-bit luminance + 1-bit color features.** Features obtained after digitizing to 1 bit both l (L/(L+M)) and

luminance (L+M) coordinates. **(C) Four gray-levels luminance sketches.** Sketches obtained with the features in (a). **(D) Color-only sketches.** Sketches obtained with the features in (b). **(E) Discrimination of images based on luminance and color sketches.** Percentage of correct discrimination with 2-bit luminance-only sketches (4 gray-levels bars) vs. 1-bit luminance + 1 bit color sketches (4 red/green-levels bars). Figure adapted from (Del Viva et al., 2016).

1.4.8 Considerations of model implementation

A model of visual compression with the ambition of a realistic description of human vision must allow for practical updating following changes in external conditions. Indeed, considering the plasticity of the visual systems, one might expect that the algorithms employed by the visual system should not only be economical to execute but also reasonably economical to set up and update when adapting to varying external or internal conditions. Most of the algorithms in this area have instead no concerns about computing power needed, which is related to how much time does it take to get to the solution, about when and how the solution is calculated, which is related to implementation issues, and how and when it can be updated. The authors thus wondered whether the visual system has made a choice that is maybe suboptimal from the point of view of the run-time performance but leads to easier and more efficient updates and improvement.

The optimality condition that is imposed by their model to the set of features, can be indeed formulated as a case of a class of well-known problems that go under the name of "knapsack problems". These problems admit exact numerical solutions, that in the general case are rather expensive to compute, as well as simpler approximate solutions, that are slightly less optimal, as the one that has been heuristically adopted their model. For this reason, in a preliminary study (Del Viva et al., 2017) they compared an exact solution to this class of optimization problems with the heuristic solution consisting in their model. First, they applied these two approaches to the extraction of optimal visual features obtaining similar but nonetheless clearly distinguishable solutions, raising the interesting question of which of the two better describes the actual performance of fast vision in human subjects. To this purpose, they compared human performance in image discrimination after fast presentation of sketches based on the two solutions (exact and heuristic) with psychophysical experiments, similar to those described before (see paragraph 1.4.5. and figure 10A). They found evidence that the actual performance of human vision agrees with the simpler approximate solution rather than the mathematical optimum. While the latter is slightly better from the point of view of computational efficiency of the image analysis, its evaluation requires a complex algorithm, that is not well suitable for calculation within the brain, and it needs to be re-calculated in case of changing in external conditions. On the other end, the heuristic approximate algorithm is computable and much easier to evaluate and update even if it provides a slightly less optimal, solution.

Interestingly, this experimental result thus suggests that plasticity is correlated to the existence of specific filters and receptive fields in the visual system.

1.4.9 Final considerations of the predictive power of the reference model

Overall, the constrained-maximum entropy model presented here, when applied to visual images is able to extract basic primitives such as edges and bars. These elements constitute the basis of edge-detection, allowing to capture important changes in properties of the world, which is one of the main goals of vision. Many computational models of early vision have been proposed, aimed at attaining the best performance in detecting objects contours by using biologically plausible elements, based on a-posteriori knowledge of physiological details (Marr & Hildreth, 1980; Morrone & Burr, 1988; Watt & Morgan, 1983). Because this function in the model was obtained without assuming any known biological detail, the first relevant finding of this work is that the edge-detection functionality in itself follows directly from very general principles, as the optimal solution for fast processing, when dealing with an information bottleneck and limited computational resources (Del Viva et al., 2013).

As discussed above, Olshausen and Field succeeded in deriving biologically plausible basis functions for visual representations, based on considerations of information efficiency (Olshausen & Field, 1996). Their work aims explicitly at reproducing the original luminance map as closely as possible, based on fitting with a minimum chi-square criterion, and utilizes a number of free parameters to achieve the best results. It can be described as aiming for an "almost-lossless" compression. In the reference model, instead, the authors aim at strong information compression and entropy maximization with no regards to the fidelity of reproduction: the limitations of the system lead to the selection of a very restricted number of salient features, all the other features available in the imput are discarded (lossy compression). The predictive power of the model however is very strong and is confirmed by the fact that the model output, even with extreme lossy compression (sketches could reach 10% of the originals, compressing data by a factor of 40) and in fast viewing conditions is easily recognizable by human observers, although the accuracy of the raw luminance map is worse than that obtained in Olshausen and Field (1996).

Past computational studies of features extraction argued that the visual system devotes resources to the detection of features of natural images in proportion to the probability of feature occurrence (Geisler et al., 2001; Simoncelli & Olshausen, 2001). Results here, on the other hand, show that principles of computational efficiency led to a somewhat different algorithm: resources are devoted to features with an intermediate probability of occurrence (see Figure 8B). Discarding the most probable input configurations is necessary to fit within the bandwidth limitations of the next processing stage. As an example, the most common visual patterns, uniform luminance patches, which are inefficient to encode, are automatically rejected by the model.

The application of the model to color images evidenced that when the system's limitations are taken into account, the most effective strategy is to ignore fine-scale color features and devote most of the bandwidth to gray-scale information. Confirmation of these predictions comes from psychophysics measurements of fast-viewing discrimination of natural scenes. These behavioral results are consistent with a human visual system that, under the pressure for optimization to use limited resources, follows the maximum-entropy principle and uses only some of the information available in the input.

Although it has not been discussed in this thesis, it is worth mentioning that there exists at least one concrete example of a successful implementation of this general model to data compression in electronic devices. Given that recognizing the presence of a certain set of discrete input patterns is within the typical capability of a neural network (Hopfield, 1982), it is also not hard to conceive that the present model can be implemented in a neural network as well.

1.5 Objective of the thesis: assessing visual saliency of optimally informative visual features

To summarize, this thesis is based on a recent constrained maximum-entropy model of early vision which, applied to natural images, allows the extraction of a limited number of *optimal* features, considered to be "salient" in fast visual analysis. The current work aims at adding further evidence that confirms the predictions of the reference model.

Our objective is to assess the visual saliency of these optimally informative visual features with a specific interest in testing their role in guiding eye movements. To this purpose, we conducted different studies on human observers, using psychophysical and eye movements paradigms, to answer the following questions:

1. Are *optimal* features significantly more salient than others even in the lack of any clues coming from global structure? To answer this question, in Study 1, *optimal* and *non-optimal* features were presented in isolation, in the absence of any meaningful global arrangement and semantic context,

and observers were explicitly asked to choose the most salient stimulus with perceptual and eye movement tasks.

2. Are *optimal* features able to automatically attract covert and overt attention? We address this question in Study 2, in which we implicitly tested the relative saliency of *optimal* and *non-optimal* features by using them as attentional cues in perceptual and oculomotor tasks, without explicitly requiring observers to pay attention to stimulus saliency.

3. Can *optimal* features influence the path of a saccade toward a target? To assess this, in Study 3, we compared the effects on saccades trajectories produced by *optimal* vs. *non-optimal* features used as distractors in a saccadic task, considering the magnitude of curvature as a measure of feature saliency.

4. *Optimal* local information contributes to successful image discrimination? To test this, Study 4 compared the relative contribution of global elements and optimal local features embedded in a few visible image fragments to image discrimination in fast vision, designing a more ecological task to explore whether *optimal* features also play a significant role in natural settings.

A key point of this work is also the comparison between the saliency determined by information maximization criteria to luminance-based saliency, achieved by comparing the effects induced by *optimal* versus *non-optimal* features to those obtained with features of different luminance.

In the following chapters, for each study, the relevant literature and the rationale of the experimental hypothesis are presented, then the methods and the results are described, followed by a discussion about the contribution of each study to the final objective of this thesis.

Chapter 2

Study 1: Early visual saliency based on isolated *optimal* features

2. STUDY 1: EARLY VISUAL SALIENCY BASED ON ISOLATED *OPTIMAL* FEATURES

2.1 Theoretical background and rationale

As broadly discussed in the introduction, a considerable amount of energy is indeed required to create an accurate representation of the visual scene in the shortest possible time (Attwell & Laughlin, 2001; Echeverri, 2006; Lennie, 2003), and, for this reason, the visual system is likely to operate a strong data reduction at an early stage of processing (Attneave, 1954; Barlow, 1961; Olshausen & Field, 1996), by creating a compact summary of the *relevant* features (Marr, 1982; Morgan, 2011).

While the existence of this early visual summary is rarely put into question, the principles driving the saliency of features, and the relative weight of local (Zhaoping, 2002; Xilin Zhang et al., 2012) and global cues in this process (Itti et al., 1998; Oliva, 2005; Oliva & Schyns, 1997b), are still subject to intense debate. The saliency of a visual stimulus depends on several physical properties (typically luminance, color, orientation of isoluminant contours - edges) and it scales with the degree of dissimilarity of each property (e.g., luminance) with regard to the statistics of that property in the surround (e.g., the stimulus luminance vs. the background luminance, or the stimulus orientation compared to the orientation of the neighboring elements (see for instance, Nothdurft, 1993; Treisman, 1985). However, a stimulus' saliency can often also be appreciated with isolated stimuli. Furthermore, the saliency related to each individual visual property of a single stimulus is typically combined into a global percept of stimulus saliency and different stimuli, defined by different conspicuous properties (e.g., a red square among green square and a tilted line among horizontal lines) can be compared and eventually empirically matched in terms of saliency (Nothdurft, 2000).

Several models have been proposed to quantitatively estimate the two-dimensional saliency distribution in a visual scene (the bottom-up saliency map). When considering more ecological conditions for vision, like during visual search with complex natural scenes, estimating the saliency of each part of the scene becomes much more difficult. Higher-level factors, such as object segmentation, semantic processing, and behavioral goals, do actually contribute, together with the physical properties, to define the relative conspicuity of the scene's regions (for a review, see Fecteau & Munoz, 2006; Itti & Borji, 2013). Models of eye guidance have tried to predict where people fixate in visual scenes and to relate these locations to visual saliency. Some studies have suggested that eye movements are mainly driven by regions with maximal feature contrast (Garcia-Diaz et al., 2012; Itti & Baldi, 2009; Itti & Koch, 2000, 2001). Concurrently, in presence of multiple features, objects, and information cues, the pattern of ocular fixations in a complex natural scene is often used as the operational definition of the saliency map of the scene (Itti & Borji, 2013). Finally, the specific task at hand does also play an important role and for instance, it has been shown that eye movements statistics in humans are consistent with an optimal search strategy that gather maximal information across the scene to successfully achieve the task (Bruce & Tsotsos, 2005; Garcia-Diaz et al., 2012; Najemnik & Geisler, 2005, 2008).

The idea that the saliency of visual features is based on the amount of information (Shannon, 1948) they carry about the visual scene has been proposed by the reference model described in Chapter 1 (Del Viva et al., 2013).

Here we ask whether the *optimal* features identified in past experiments (Del Viva et al., 2013) are perceived as salient outside the context of the global image structure to which they belong (*sketches*). We address this question through a saliency discrimination task between *optimal* and *non-optimal* features, by explicitly asking the participants to choose the stimulus which stands out or automatically grabs their attention, through either a hand-button press or a saccadic orienting response (Castellotti et al., 2021). This way of measuring saliency is not based on an automatic response, unlike in the majority of studies (e.g., Donk & van Zoest, 2008; Zhaoping & May, 2007), but it requires an explicit behavioral choice, as previously used, for example when preference does not depend on the intensity of a single low-level property of the stimulus – e.g., contrast, luminance, color (e.g., Nothdurft, 2000; Nothdurft, 1993a). This is the case of our stimuli, which do not differ for the low-level properties usually defining visual saliency, but for the internal spatial arrangement of black-and-white pixels (see Figures 8C and 8E). These differences derive from the process of constrained-entropy maximization of the statistics of visual scenes, required by early input data reduction (Del Viva et al., 2013). Thus, the saliency preference for isolated *optimal* features, even though asked explicitly, is not obvious.

Specifically, we conducted four psychophysics and one eye-movement experiment to determine the degree of saliency given by *optimal* features, compared to *non-optimal* features. In Experiment 1, to assess the minimal number of *optimal* features able to trigger a saliency discrimination, the preference for *optimal* vs. *non-optimal* features was measured as a function of their number. Experiment 2 was designed to assess how many *optimal* features surrounded by a group of *non-optimal* features ("signal-to-noise ratio", SNR) are necessary to consider them more salient. This is a more ecological condition than that in Experiment 1 because in natural images *optimal* features (edges and lines) are always surrounded by others that do not define object contours and are therefore considered as noise, according to our model. Visual saliency is strongly dependent on luminance contrast (Treisman, 1985), whose analysis involves early visual processing starting from the retina. It is therefore particularly important to study its effect in determining saliency in fast vision. We studied the effect of contrast in Experiments 3 and 4. In Experiment 3, the preference for *optimal* vs. *non-optimal* features was measured as a function of the contrast of both, to measure the lowest contrast needed to still choose the *optimal* features as the more salient. In Experiment 4, the preference for the *optimal* features was measured as a function of their contrast relative to *non-optimal* features, to measure the minimal contrast *optimal* features must have to be considered as salient as *non-optimal* features. We can consider this value as the contrast equivalent to the saliency given by the spatial structure of *optimal* features. Finally, the preference for *optimal* features as a function of their number relative to *non-optimal* features (SNR), was also measured with saccadic eye movements. In this work, eye movements are not used as an operational definition of saliency (e.g., Itti & Borji, 2013), but as an alternative modality to the psychophysics response. We argue that the dynamic and metric properties of gaze-orienting responses might provide additional insight on the saliency-based capture exerted by optimal features.

2.2 Aim of the study

In Study 1 (Castellotti et al., 2021), for the first time, *optimal* features were presented in isolation, to investigate whether they are considered visually more salient than other *non-optimal* features, even in the absence of any meaningful global arrangement (contour, line, etc.) and semantic context (sketch).

2.3 Materials and methods

2.3.1 Psychophysics experiments

2.3.1.1 Participants

The condition with one feature in Experiment 1 was tested on 20 observers (13 women, mean age = 27 ± 2 years). Five other different observers (3 women, mean age = 23 ± 3 years) participated in the other conditions of Experiment 1, and Experiments 2, 3, and 4. All observers had normal or corrected to normal visual acuity and no history of visual or neurological disorders. Observers were unaware of the aim of the experiments and gave written informed consent before the experiments.

2.3.1.2 Apparatus and set-up

All stimuli were programmed on an ACER computer running Windows 7 with Matlab 2016b, using the Psychophysics Toolbox extensions (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997), and displayed on a gamma-corrected CRT Silicon Graphics monitor with 1280 x 960 pixels resolution at 120 Hz refresh rate. The whole display (38.5 x 29.5 cm) subtended 38.5° x 29.5° of visual angle at a viewing distance of 57 cm. All experiments were carried out in a dark room, with no lighting other than the display screen. Ad hoc software in Mathematica (Wolfram Inc.) was used for the extraction of stimuli, curve fitting, and statistical analysis. Participants' manual responses were provided on a standard Dell keyboard.

2.3.1.3 Stimuli

Stimuli were two compounds of a certain number of small features, subtending 1.5 deg of visual angle at 57 cm distance (1.56 x 1.56 cm) and located horizontally at 3 deg eccentricity, right and left of the center of the screen. Each compound comprised several 3x3 pixels features, subtending 0.12 deg at 57 cm distance (0.12 x 0.12 cm) each. They were randomly selected with replacement (at each trial) from a set of 50 black and white *optimal* features (Figure 8C) selected according to the constrained maximum-entropy model, already used to build the sketches in previous experiments (Del Viva et al., 2013), and from a set of 50 black and white *non-optimal* features (Figure 8D), with the lowest probability of occurrence in the statistical distribution of all possible 3x3 pixel black and white features. We chose these *non-optimal* features as a control for saliency because the difference between *optimal* and *non-optimal* features is given only by their internal black-and-white pixel arrangement, and they do not differ, on average, in luminance and spatial frequency content (see **Supplementary Material**). The positions of the features within each compound were assigned randomly at each trial and were set such that the distance between neighboring features had to be about 3 pixels in each direction. Random selection and random position of features in the stimulus ensured that saliency was provided only by individual features rather than by their global arrangement. The left/right position of each compound was also varied randomly from trial to trial. Luminance white: 35 cd/m^2; luminance black: 1 cd/m^2; luminance grey background: 12 cd/m^2.

2.3.1.4 Procedure

In all experiments, participants were asked to choose which of the two compounds presented on each side of the screen was the most salient, in a 2AFC procedure. Participants were sitting in a dark room at 57 cm distance from the monitor. Each trial started with the presentation of a grey display for 800 ms, during which observers were asked to fixate a cross in the center of the screen. The compound stimuli were then shown for 26 ms on a grey background. After the stimulus presentation, participants indicated the more salient compound by pressing a computer key. There was no time limit for the response (**Figure 14A**). All data for each participant were collected in one single session of about one hour, divided into four blocks (one block/experiment).

In Experiment 1 the preference for a compound of *optimal* features (target) with respect to a compound of *non-optimal* features (distractor) was measured as a function of the number of features presented. The luminance contrast was 100% in all trials. Considering these features are very small and are presented for a very short time, the minimal number of optimal features that triggers a consistent preference based on saliency becomes very important. For this reason, in a preliminary phase, a single feature was presented on each side, to check for the possible presence of an effect even in this limit condition. A total of 200 trials/observers were run. 20 observers participated just in this measurement. Then, five different observers completed the experiment to assess the number of features that produce maximal saliency discrimination. Three, five, seven, and ten features in each compound were presented to these five observers. Target and distractor always had the same number of features, varying from trial to trial according to a constant-stimuli procedure. A total of 1200 trials per observer were run (**Figure 14B**).

In Experiment 2 the saliency-based preferential choice was measured as a function of the relative number of *optimal* vs. *non-optimal* features in the same compound. The target included a total of 10 *optimal* and *non-optimal* features in variable proportions (variable signal to noise ratio, SNR). The distractor included 10 *non-optimal* features. The luminance contrast (100%) and the total number of features in each compound (10) were kept constant in all trials. The SNR was either 0.1, or 0.4, or 0.6 or 1 (corresponding to 1, 4, 6, or 10 *optimal* features in the target compound), and this number was set randomly from trial to trial according to a constant stimuli procedure. A total of 1200 trials per observer were run (**Figure 14C**).

In Experiment 3, the strength of the saliency-based preferential choice was measured as a function of the contrast of both *optimal* features (target compound) and *non-optimal* features (distractor compound). In this experiment, the number of features in the two compounds was kept constant at 10. Contrast for both target and distractor was set at 0.15, 0.2, 0.25, 0.3, 0.5 and the value was set randomly from trial to trial according to a constant stimuli procedure. Participants were asked to press a computer key to indicate the more salient compound. A total of 500 trials per observer were run (**Figure 14D**).

In Experiment 4, the preference for *optimal* features (target compound) was measured as a function of their contrast relative to the contrast of *non-optimal* features (distractor compound). That is, in half of the trials, the contrast of the target was varied while the contrast of the distractor was kept constant at 100%. The contrast of the target for which the observers could not tell anymore which compound was more salient can be considered as 1-the contrast value equivalent to the saliency of our features. In the other half of the trials, the contrast of the target was kept constant at 100% while the contrast of the distractor was varied. These "catch trials" were used to avoid contrast cues that could bias the observers' choice. All these trials were randomized. The number of features in the two compounds was the same (10) in all trials. Contrast values in the varying compound were 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95, 1, set randomly from trial to trial according to a constant stimuli procedure. A total of 800 trials per observer were run (**Figure 14E**).

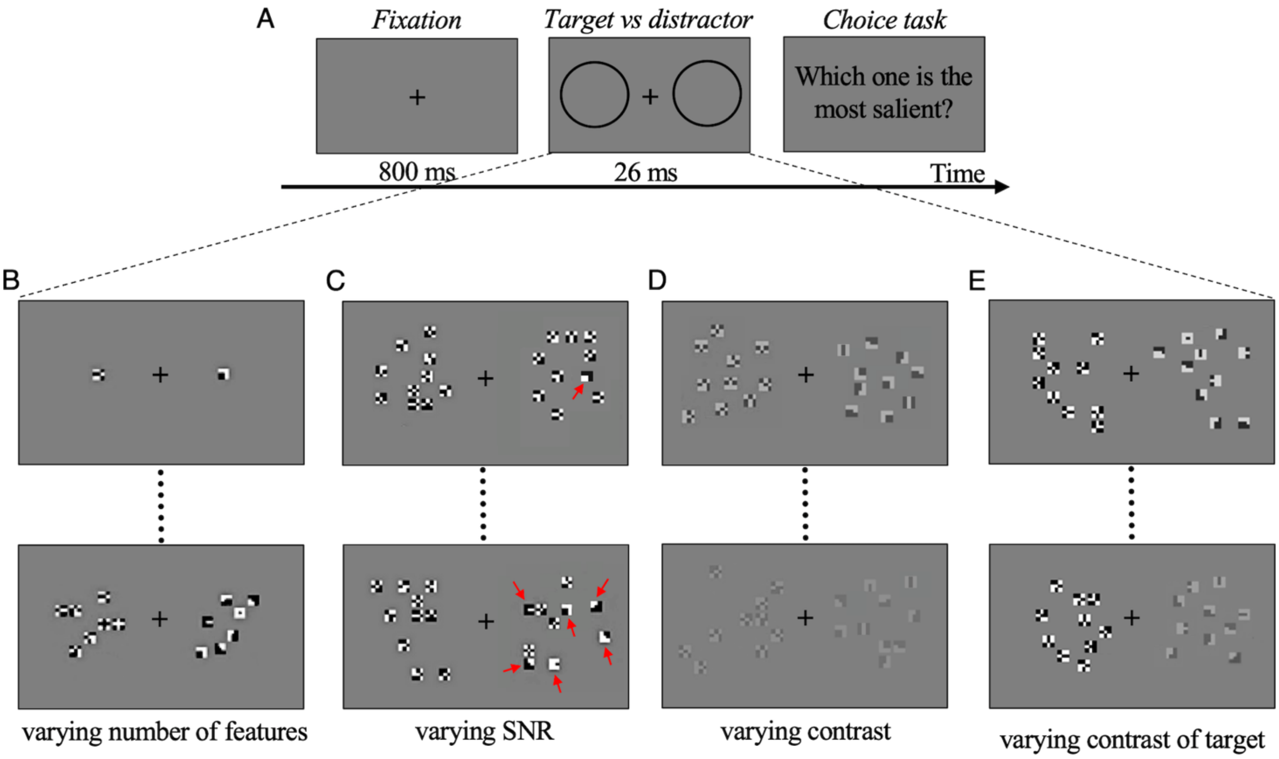

Figure 14. Study 1 – Psychophysics experiments: procedure and conditions. (A) Schematic representation of one trial. During stimulus presentation two compounds, target (deemed salient), and distractor (not salient), were presented randomly left/right. The two black circles (not visible in the real display) represent the location of target and distractor, shown below in (B), (C), (D), and (E) for each experimental condition. In (B), (C), (D), (E), the target is always on the right. **(B) Examples of stimuli for Experiment 1.** Upper panel: target with 1 *optimal* feature vs. distractor with 1 *non-optimal* feature. Lower panel: target with 7 *optimal* features vs. distractor with 7 *non-optimal* features. **(C) Examples of stimuli for Experiment 2.** Upper panel: target with 1 *optimal* feature plus 9 *non-optimal* features (SNR = 10%) vs. distractor with 10 *non-optimal* features. Lower panel: target with

6 *optimal* features plus 4 *non-optimal* features (SNR = 60%) vs distractor with 10 *non-optimal* features. Red arrows indicate *optimal* features. **(D) Examples of stimuli for Experiment 3.** Upper panel: target with 10 *optimal* features vs. distractor with 10 *non-optimal* features (contrast of both = 50%). Lower panel: target with 10 *optimal* features vs. distractor with 10 *non-optimal* features (contrast of both = 20%). **(E) Examples of stimuli for Experiment 4.** Upper panel: target with 10 *optimal* features (contrast = 80%) vs. distractor with 10 *non-optimal* features (contrast = 100%). Lower panel: target with 10 *optimal* features (contrast = 65%) vs. distractor with 10 *non-optimal* features (contrast = 100%). Compounds and features are oversized for illustration purposes. The target-distractor compounds position in the trials is randomized. Figure retrieved from (Castellotti et al., 2021).

2.3.1.5 Data processing

In Experiments 1 and 2, for each participant and condition, the probability of choosing the target (with binomial standard errors) has been calculated for each compound condition. In Experiments 3 and 4, for each participant and condition, a 2-parameters (position and slope) Maximum Likelihood fit was performed off-line with data obtained in all sessions, based on an ERF (sigmoid) psychometric function. Psychometric functions run from 0.5 to 1 in Experiment 3 and thresholds were defined as the target contrast yielding 75% correct discrimination. In Experiment 4, psychometric functions run from 0 to 1, and thresholds were defined as the target contrast yielding 50% correct discrimination. The goodness of fit was determined from the difference in log-Likelihood between the fit, and an ideal fit describing all points exactly. This is used to obtain a p-value under the chi-square approximation (Wilks' theorem).

2.3.2 Eye movements experiment

2.3.2.1 Participants

Seven observers (3 women, mean age = 30.1 ± 8 years) participated in the eye movements task and the psychophysical control experiment. All observers had normal or corrected to normal visual acuity and no history of visual or neurological disorders.

2.3.2.2 Apparatus and set-up

All stimuli were programmed on a MacPro computer running OS 10.6.8 with Matlab 2016b, using the Psychophysics Toolbox (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997), and the Eyelink Toolbox extensions (Cornelissen et al., 2002), and displayed on a Samsung SyncMaster 2233 LED-monitor with 1680 × 1050 pixels resolution at 120 Hz refresh rate. The whole display (47.2 x 29.5

cm) subtended about 47° x 29° of visual angle at a viewing distance of 57.3 cm. All experiments were carried out in a dark room, with no lighting other than the display screen. Eye movements were recorded using an Eyelink 1000 video-based eye tracker (sampling rate 1 kHz). The viewing was binocular, but only the right eye was tracked. A chin and forehead rest stabilized the head.

2.3.2.3 Stimuli

Stimuli were two compounds of 10 features each. The target compound comprised a variable number of *optimal* features (1, 4, 6, or 10) and *non-optimal* features; the distractor comprised only *non-optimal* features, analogously to the psychophysics Experiment 2. For the eye movements experiment, the target and distractor compound-pair could appear randomly at 5 different locations, with target and distractor arranged symmetrically with respect to the vertical meridian and their respective position (right or left) randomly switched across trials (**Figure 15**). If we consider the compound on the right-hand side, its position was defined by an angle of 0°, ±45°, or ±70° with respect to the horizontal midline (**Figure 15B**). In the following, we will refer to these angles to indicate the position of the compound pair. Angles were randomly alternated in the presentation sequence to maximally reduce motor preparation for the saccade and to assess possible spatial anisotropies of the *optimal* features-based saliency. Both compounds were displayed at a larger eccentricity (5°) than the one used in the psychophysical experiments (3°), in order to elicit goal-directed saccades, clearly aiming outside the perifoveal region (**Figure 15B**). To compensate for the larger eccentricity, all the stimuli were slightly larger than in Experiment 2. Compounds subtended 1.8 x 1.8 deg (nearly 1.8 x 1.8 cm) and individual features about 10 x 10 min of arc (0.17 x 0.17 cm) at 57.3 cm viewing distance. Each feature was defined by a 6 x 6 white and black pixels patch. Positions of features within each compound were randomly assigned at each trial, ensuring a distance of about 6 pixels in each direction between neighboring features. White pixels had a luminance of 82 cd/m2; black pixels: < 2 cd/m2; and the luminance of the grey background was about 42 cd/m^2.

2.3.2.4 Procedure

After a fixation period of random duration between 500 and 800 ms, the target-distractor pair was presented for 26 ms (three frames). Then, two placeholders were displayed for 800 ms at the compound-pair location. The placeholders ensured that observers could program a relatively accurate visually-guided saccade even once the compounds have disappeared. Observers were asked to move their gaze towards the location where they saw the “most salient stimulus”, in a 2AFC choice-saccade task (**Figure 15A**). 800 trials were collected for each observer.

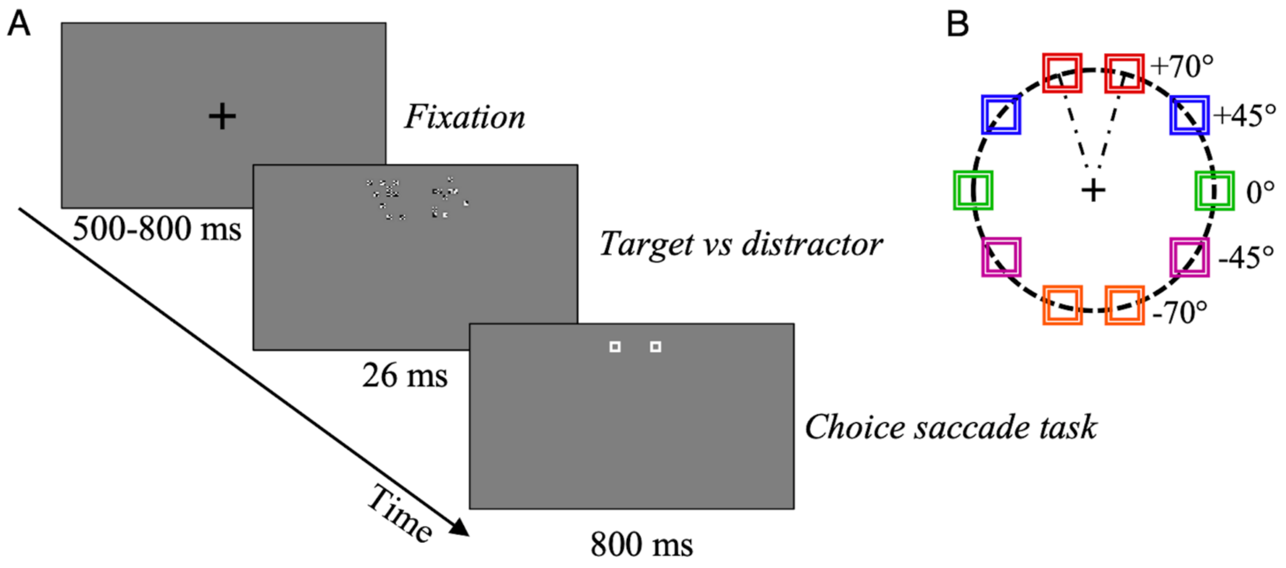

Figure 15. Study 1 – Eye movements experiment: procedure. (A) Schematic representation of one trial. After a random-duration fixation period, one compound comprising a variable proportion (SNR) of *optimal* features (target), and a compound with *non-optimal* features (distractor), were presented to the right or left, at different angles, for 26 ms. Then, two placeholders were shown at the target and distractor locations for 800 ms. In this example, target SNR = 60%. **(B) Target and distractor possible locations.** The two compounds could be presented randomly at one of 5 different locations (0°, ±45°, ±70°) as defined in the text, illustrated by different colors, at 5 deg eccentricity. Figure retrieved from (Castellotti et al., 2021).

Since experimental conditions are different from the psychophysics Experiment 2, as a control, we repeated the psychophysical measurements with these observers, stimuli, and setup. As in the psychophysics Experiment 2, the saliency-based preferential choice for the *optimal* features was measured as a function of their number relative to the total number of features (*optimal* and *non-optimal*, always equal to 10) in the same compound (SNR). In this control experiment, only the condition where the target and distractor were presented on the horizontal axis was tested and a total of 400 trials/observer were run.

2.3.2.5 Data processing

Ad hoc software in Matlab and Mathematica (Wolfram Inc.) was used for extraction of oculomotor parameters and statistical analysis. Recorded horizontal and vertical gaze positions were low-pass filtered using a Butterworth (acausal) filter of order 2 with a 30-Hz cutoff frequency and then numerically differentiated to obtain velocity measurements. We used an automatic conjoint acceleration and velocity threshold method to detect saccades (see for instance Damasse et al., 2018), and we visually inspected all oculomotor traces to exclude aberrant trials. We excluded from the analysis saccades with latencies below 140 ms, considered anticipatory and not guided by visual information in this type of choice-saccade tasks (e.g., Walker et al., 1997), and very late saccades, above 500 ms (less than 6% of the first detected saccades overall). Visual inspection of individual latency histograms confirmed that saccades with latency below 140 ms and above 500 ms did not belong to the principal mode of the distribution. When a small anticipatory saccade was detected (amplitude below 3 deg), the second saccade was used instead for the analysis (less than 2% of total).

For each saccade, we estimated latency, amplitude, endpoint position, and the distance between the eye position endpoint and the center of the target (or distractor) compound. Saccades ending within 1.5 deg of either target or distractor were classified as "valid", and respectively labeled "To-target" (*correct*) or "To-distractor" (*erroneous*). All the other saccades, landing farther than 1.5 deg from the compound, were considered as invalid, and labeled "Quasi-Target" or "Quasi-Distractor" when they brought the gaze in the same hemifield of the target or the distractor respectively. The choice of the 1.5° distance criterion was motivated, on one hand, by the requirement that the validity-surrounds would not overlap between the two compounds in the 70° (uppermost) and -70° (lowermost position) conditions. On the other hand, this criterion distance is reasonable for a target-compound with a side of approximately the same size.

2.4 Results

2.4.1 Psychophysics experiments

Results of Experiment 1 show that all observers found the target compound to be much more salient than the distractor. Even a single tiny *optimal* feature was chosen with probability = 0.71 ± 0.01 over its alternative by 20 observers (**Figure 16**). Probability of choosing the target as more salient increases with the number of features presented up to 10 features, where probability saturates for all observers. This number was used in all the following experiments.

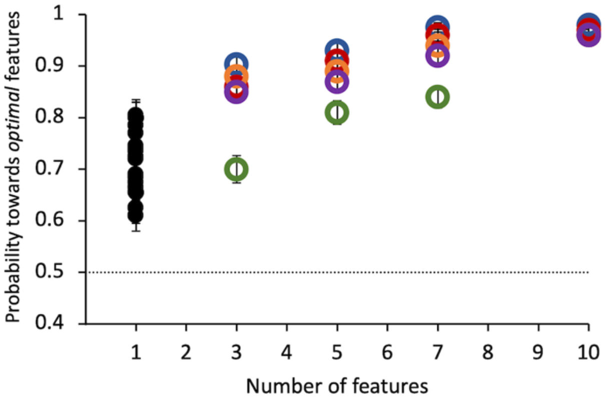

Figure 16. Study 1 – Psychophysics Experiment 1: probability of selecting the target as a function of the number of features. Colored symbols represent data from 5 individual observers. Black symbols represent data from 20 different individual observers, tested in this condition only. Errors of individual dots are binomial standard deviations. The dashed line indicates the guessing level for this task (0.5 probability). Figure adapted from (Castellotti et al., 2021).

Results of Experiment 2 show that even when *optimal* features are intermixed with *non-optimal* features in the same compound, observers still indicate this compound as more salient than the alternative. The probability increases with SNR. A compound with a single *optimal* feature surrounded by nine *non-optimal* features is sufficient to lead observers to consider this stimulus as more salient than the other with probability = 0.64 ± 0.02 ($z = 6$, $p < 0.001$) (**Figure 17**).

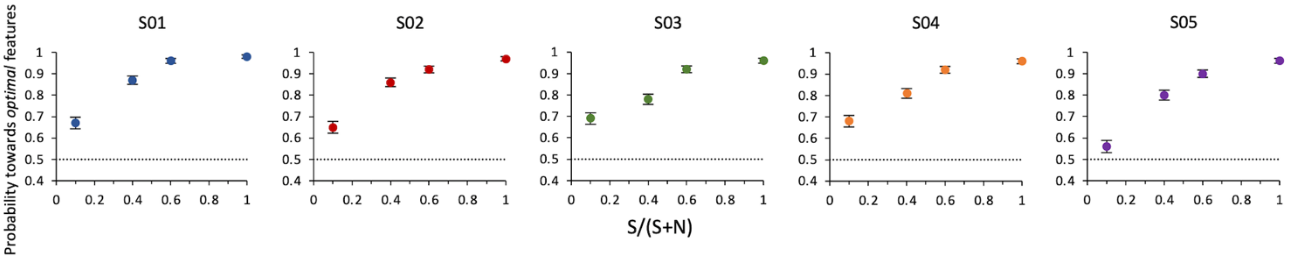

Figure 17. Study 1 – Psychophysics Experiment 2: probability of selecting the target as a function of SNR. Data from individual observers. The dashed line indicates the guessing level for this task (0.5 probability). Target SNR could be 0.1, 0.4, 0.6, or 1 (corresponding to 1, 4, 6, or 10 *optimal* features in the target compound out of 10 total features). Figure adapted from (Castellotti et al., 2021).

Results of Experiment 3 show that the lowest contrast needed to still choose the *optimal* features as the more salient is 0.23 ± 0.0006. This is the weighted average of the thresholds from maximum likelihood fits of individual data (**Figure 18**).

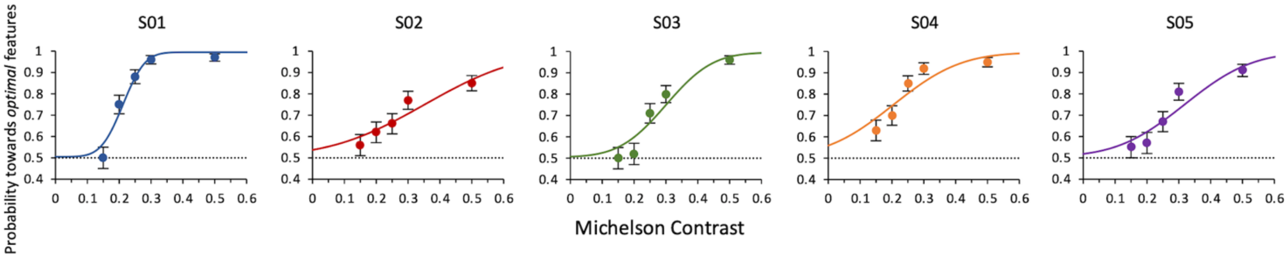

Figure 18. Study 1 – Psychophysics Experiment 3: probability of selecting the target as a function of contrast of both target and distractor. Data from individual observers. The line is the ML best fit. Individual thresholds are given by contrast values corresponding to 75% level performance and are respectively 0.21 ± 0.007, 0.35 ± 0.88, 0.30 ± 0.02, 0.20 ± 0.02, 0.32 ± 0.02. The dashed line indicates the guessing level for this task (0.5 probability). Figure adapted from (Castellotti et al., 2021).

Results Experiment 4 show that when the contrast of *non-optimal* features is lowered, all observers always deemed the compound of *optimal* features (kept at 100% contrast) the most salient one. Conversely, when the contrast of *non-optimal* features was kept at 100% and that of *optimal* features was lowered, they still considered them as more salient, but with a decreasing probability as the contrast decreased. The average contrast value for which the contrast of *optimal* features balances the saliency of the *non-optimal* features is 0.63 ± 0.004 (weighted average of individual thresholds) (**Figure 19**).

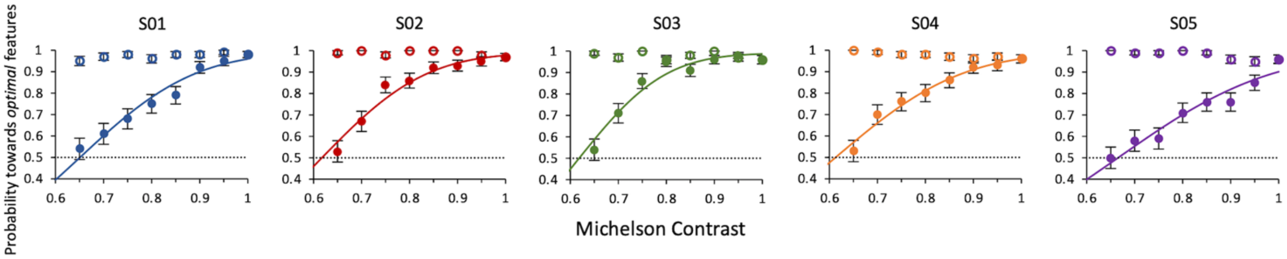

Figure 19. Study 1 – Psychophysics Experiment 4: probability of selecting the target as a function of relative contrast of target or distractor. Data from individual observers. Filled symbols: the contrast of the target is varied. Open symbols: the contrast of the distractor is varied. The line is the ML best fit. Individual thresholds are given by contrast values corresponding to 50% level performance and are respectively 0.65 ± 0.01, 0.62 ± 0.006, 0.62 ± 0.01, 0.61 ± 0.01, 0.67 ± 0.01. Figure adapted from (Castellotti et al., 2021).

2.4.2 Eye movement experiment

Figure 20 shows probabilities for the first *correct* saccade and psychophysical choice of the same observers, as a function of the relative number of *optimal* vs. *non-optimal* features in the compound

(SNR), when the target and distractor compounds were presented on the horizontal axis (0/180°). Both psychophysical and eye movements data confirm the results of Experiment 2, although with a smaller set of data and at a slightly larger eccentricity (5° instead of 3°).

That is, even when in the same compound *optimal* features are intermixed with *non-optimal* features, observers consider this compound as more salient than the other comprising only *non-optimal* features, and they do so with a probability that increases with SNR. A compound with just one *optimal* feature surrounded by nine *non-optimal* features is sufficient to lead observers to consider this stimulus as more salient than the other one with probability 0.65 ± 0.02 for psychophysics ($z = 3.66$, $p < 0.001$), and to orient the gaze toward it with probability 0.65 ± 0.03 for saccadic choice ($z = 1.75$, $p < 0.05$).

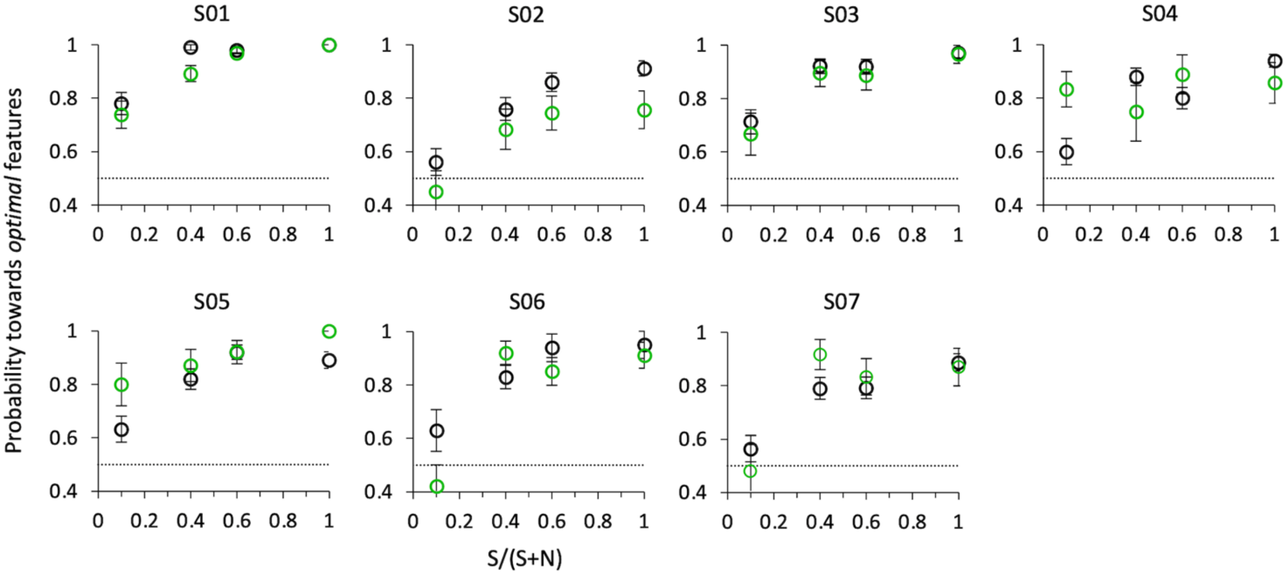

Figure 20. Study 1 – Eye movements experiment: probability of selecting the target as a function of SNR at angle 0°. Black circles: psychophysical results (button-press); green circles: eye movements results. Data from individual observers (weighted means with their errors). The dashed line indicates the guessing level for this task (0.5 probability). Figure retrieved from (Castellotti et al., 2021).

When all directions tested are considered (-70°, -45°, 0°, 45°, 70°), the average probability for the choice saccade to land in the vicinity of the target compound also increases with SNR (**Figure 21A**). The average performance depends on angles: compared to angle 0°, performance is lower for the upper quadrant, both for +45° ($z = -2.19$, $p < 0.05$) and +70° angles ($z = -5.20$, $p < 0.001$). The performance for the lower quadrant does not differ from 0°, either for -45° ($z = 0.09$, $p > 0.5$) and -70° angles ($z = -1.20$, $p > 0.05$).

We evaluated the mean latency of saccades that were correctly oriented toward the salient compound, as a function of the SNR and for each different angle of presentation. **Figure 21B** shows a strong difference of the saccadic latency across angles, with latencies being much shorter for eye movements directed toward the upper hemifield and in particular to the uppermost target-distractor compound location (angle +70°). A mixed-effects linear regression analysis of mean saccade latency (with SNR, angle, and choice-accuracy – to-Target vs to-Distractor saccades – as fixed-effects, and the same factors per observer as random-effects) revealed that only the angle but neither SNR nor choice-accuracy did significantly influence latency (mean regression slope = -0.46; standard error = 0.08; t = -5.63; $p < 0.01$).

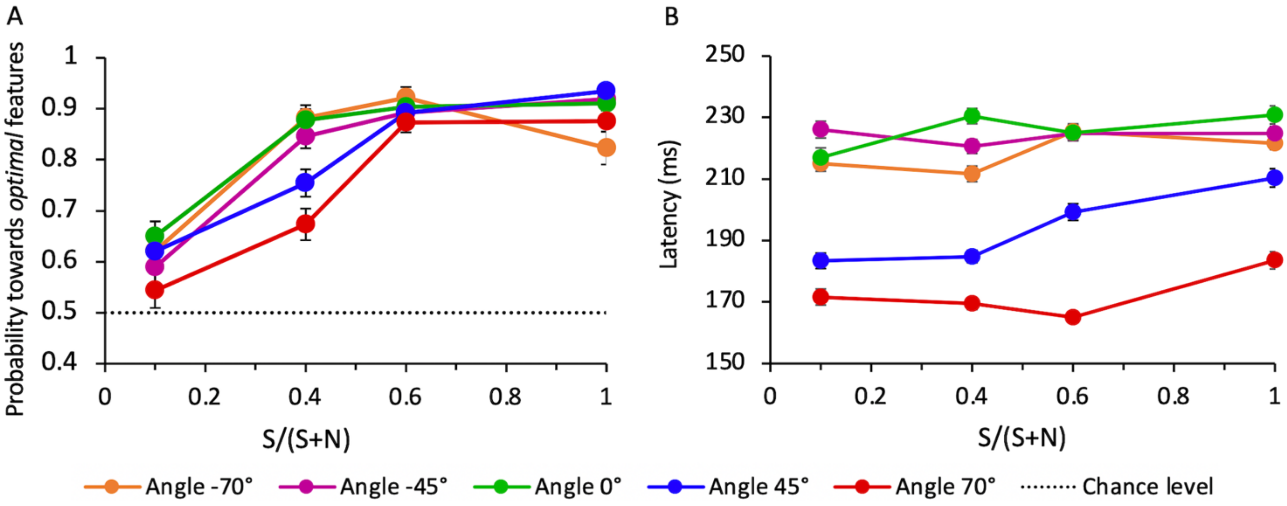

Figure 21. Study 1 – Eye movements experiment: correct saccades and their latency as a function of SNR at different angles. (A) Average probability of *correct* saccades. Data are pooled across observers (weighted means with their errors). The dashed line indicates the guessing level for this task (0.5 probability). (B) Average latency of *correct* saccades. Averages are taken over all correct trials (classified as "to-the-target") and all observers (weighted means with their errors). Figure retrieved from (Castellotti et al., 2021).

Figure 22 shows the landing point of all saccades of all observers for the lowest (**Figure 22A**) and highest (**Figure 22B**) stimulus saliency conditions. To better visualize saccadic accuracy and precision, data have been flipped and pooled as though the target compounds were always on the right hemifield and the distractor compounds were always on the left hemifield. Saccades are categorized as *valid*, "to-Target" (*correct*), or "to-Distractor" (*erroneous*), when they land at a distance lower than 1.5°, respectively from the target or distractor (filled circles in Figure 22).

To analyze the precision of landing positions for valid saccades, we calculated the absolute distance of landing position of *correct* and *erroneous* saccades from the compounds. At SNR 0.1, the mean absolute distance (± SEM) from the target compound of *correct* saccades was not significantly different from the distance of *erroneous* saccades from the distractor compound (respectively 0.74 ± 0.03° and 0.76 ± 0.03°; Paired Samples 1-tailed t-test, $t(6) = 0.93$, $p > 0.05$). At SNR 1, the distance from the target of *correct* saccades was instead significantly smaller than the distance of *erroneous* saccades from the distractor (respectively 0.59 ± 0.05° and 0.99° ± 0.09°; $t(6) = 4.19$, $p < 0.01$). At SNR 1 *correct* saccades landed closer to the target than at SNR 0.1 ($t(6) = -4.22$, $p < 0.01$), whereas *erroneous* saccades at SNR 1 landed further away from the distractor than at SNR 0.1 ($t(6) = 2.4$, $p < 0.05$). To investigate whether these differences could be explained by an "attraction" exerted by *optimal* features, we also analyzed the landing position along the horizontal axis, that is the presence of left-right biases in the saccade's directions. *Landing errors* of *correct* and *erroneous* saccades were computed as the difference between the horizontal component of the estimated eye position at the end of the saccade and the position of the center of the nearby target or distractor compound. According to our convention (see caption of Figure 22), for to-Target saccades, a *landing error* compatible with zero corresponds to a saccade landing precisely on the target center; a negative *landing error* corresponds to a saccade landing closer to the screen vertical midline with respect to the target center (thus in the direction of the distractor on the horizontal axis), whereas a positive *landing error* corresponds to a saccade landing further away from the screen vertical midline (beyond the target on the horizontal axis). The opposite relation holds for to-Distractor saccades. At SNR = 0.1 (Figure 22A) the mean *landing error* for saccades to-Target (-0.09° ± 0.002) is significantly different from 0 (One Sample 2-tailed t-test, $t(6) = -4.80$, $p < 0.01$). The mean *landing error* for saccades to-Distractor (0.19° ± 0.07) is significantly different from 0 as well ($t(6) = 5.06$, $p < 0.1$). Thus, with low-saliency compounds, both to-Target, and to-Distractor saccades land nearer to the screen vertical midline. However, the absolute value of the *landing error* of to-Distractor saccades is larger than that of to-Target saccades (Paired Samples 1-tailed t-test, $t(6) = 3.004$, $p < 0.05$), suggesting that to-Distractor saccades are less precise than to-Target saccades and that they tend to land shorter from the distractor and relatively closer to the salient compound on the opposite side. At SNR = 1 (Figure 22B) the mean *landing error* for to-Target saccades (-0.03° ± 0.1) is not significantly different from 0 ($t(6) = -0.67$, $p > 0.5$), whereas the mean *landing error* for to-Distractor saccades (0.28° ± 0.1) is significantly different from 0 ($t(6) = 3.89$, $p < 0.01$), again significantly greater than that of to-Target saccades ($t(6) = 3.63$, $p < 0.01$). Thus, when the target compound is very salient, to-Target saccades are precise and land very close to the compound center, whereas to-Distractor saccades are less precise and tend to fall short of the distractor, revealing a bias for the saccade landing position toward the salient compound. In addition, when SNR increases from 0.1 to 1, the *landing* *error* for saccades to-Distractor move further away from the distractor compound and relatively closer to the target compound ($t(6) = -2.39$, $p < 0.05$), whereas to-Target saccades land closer and closer to the center of the target compound ($t(6) = -2.21$, $p < 0.05$). When analyzed independently for different angles, the precision of valid to-Target saccades does also provide different results. The mean *landing error* (± SEM) for the 70° angle is quite large: -0.48° ± 0.05 at SNR =0.1 and -0.46° ± 0.04 at SNR = 1. In contrast, saccades are much more precise at 0° angle, with a *landing error* compatible with 0° within our uncertainty (-0.03° ± 0.04 at SNR 0.1; 0.01° ± 0.04 at SNR 1).

To assess the attraction of *optimal* features independently on the criterion we choose for validity, we also analyzed the behavior of invalid saccades landing farther than 1.5° from either target ("Quasi-Target") or distractor ("Quasi-Distractor"), for the two extreme SNR values, 0.1 and 1 (empty circles in Figure 22). First, in order to measure saccade accuracy, the ratios of "Quasi-Target"/"to-Target" and "Quasi-Distractor"/"to-Distractor" saccades were compared. When considering all saccades independently of the angle, the "Quasi-Distractor"/"to-Distractor" ratio is larger than "Quasi-Target"/"to-Target", both for the lowest (Binomial test, 18% vs. 11%, $p = 0.004$) and highest (18% vs. 9%, $p = 0.0004$) SNR values. This result suggests that when saccades are directed on the side of the distractor, the probability to meet the 1.5° criterion from the goal is lower, compared to saccades directed on the side of the salient compound. When different angles are considered separately, the landing position at 0° is the most accurate, with a low ratio of "Quasi-Target"/"to-Target" (3% at SNR = 0.1 and 1% at SNR = 1), becoming progressively less accurate moving further away from 0° (see Table 1). Then the horizontal landing position with respect to the vertical midline of the screen of these invalid saccades was analyzed, to detect possible biases due to saliency. At SNR = 0.1 (Figure 22A), the absolute horizontal landing position of "Quasi-Target" and "Quasi-Distractor" saccades are statistically compatible ($t(6) = 0.32$, $p > 0.5$). In contrast, at SNR = 1 (Figure 22B) Quasi-Distractor saccades land away from their goal and closer to the center of the screen compared to Quasi-Target saccades ($t(6) = 2.06$, $p < 0.05$). In addition, when SNR increases from 0.1 to 1, there is a significant shift of the mean landing position of Quasi-Distractor saccades ($t(6) = -2.55$, $p < 0.05$) away from the Distractor in the direction of the Target, and a significant shift of Quasi-Target saccades in the direction of the target ($t(6) = -3.62$, $p < 0.05$). Therefore, similarly to valid saccades, invalid saccades tend also to be relatively biased away from the distractor and further toward the salient compound when saliency increases, pointing to the general validity of these effects, regardless of the specific criterion for saccade validity.

Table 1. Study 1 – Eye movements experiment. Ratios of "Quasi-Target"/"to-Target" and "Quasi-Distractor"/"to-Distractor" saccades.

	SNR 0.1					SNR 1				
Angle	**-70°**	**-45°**	**0°**	**+45°**	**+70°**	**-70°**	**-45°**	**0°**	**+45°**	**+70°**
Quasi-Target	12.9%	10.5%	3.3%	7.8%	22.6%	11.9%	6.8%	1.5%	11.0%	15.8%
Quasi-Distractor	26.7%	19.6%	4.3%	18.0%	22.2%	23.5%	17.4%	5.0%	22.5%	23.0%

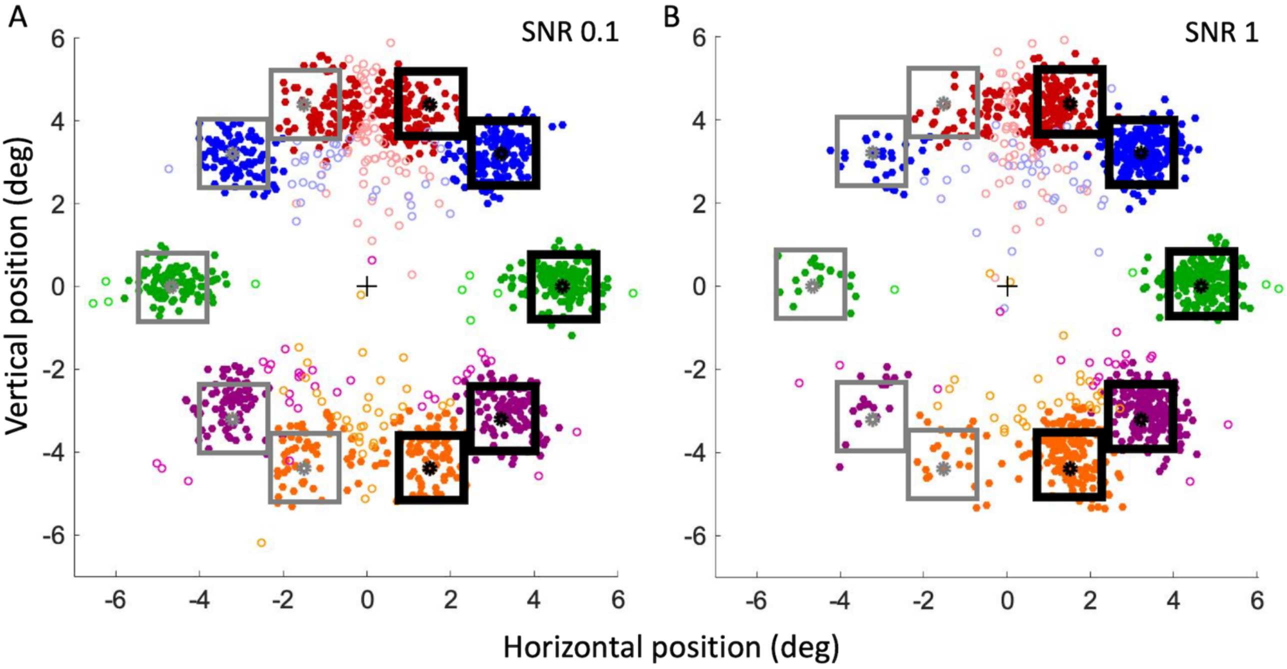

Figure 22. Study 1 – Eye movements experiment: landing positions of correct and erroneous saccades for SNR 0.1 and SNR 1. (A) SNR 0.1. (B) SNR 1. For our convention, the center of the screen corresponds to a horizontal component of 0; the target compound (black squares) is always on the right hemifield, and the distractor is always on the left hemifield (grey squares). Filled circles: valid saccades (within 1.5°); empty circles: invalid saccades. Red: angle +70°; blue: angle +45°; green: angle 0°; violet: angle -45°; orange: angle -70°. Figure retrieved from (Castellotti et al., 2021).

Finally, we analyzed whether the integration of visual information across time influences the selection of salient features for saccade orientation. If this were true, we would expect the choice performance to vary as a function of saccade latency. A general principle of perceptual decision-making models is that the percentage of correct choices is an increasing function of the response reaction time (Ratcliff & McKoon, 2008).

Figure 23 shows, for the two angles that most differ for performance and latency (0° and 70°) and the two extreme SNR values (0.1 and 1), the pooled probability for a saccade to land at the target compound depending on its latency. Latencies were divided into "fast" and "slow" depending on whether they were below or above the individual median latency respectively. Our results highlight some variability across angles and SNR values. When the target-distractor compound pair is hardly discriminable (SNR = 0.1) and is displayed in the upper hemifield (70° angle), longer-latency saccades lead to better performance compared to short-latency ones (z = -2.4 p = 0.0081). The opposite is true at 0° angle, with a significant decrease of performance for longer-latency saccades (z = 2.12, p = 0.017), pointing in this case to a disadvantage for target selection performance with prolonged integration of visual information in time. See the Discussion section for a possible explanation for this surprising result. With highly salient target compounds (SNR = 1) saccade latency does not have a systematic effect on the choice performance at either 0° or 70° angle, in agreement with the idea that feature-based selection is a fast mechanism that does not benefit from a long temporal integration.

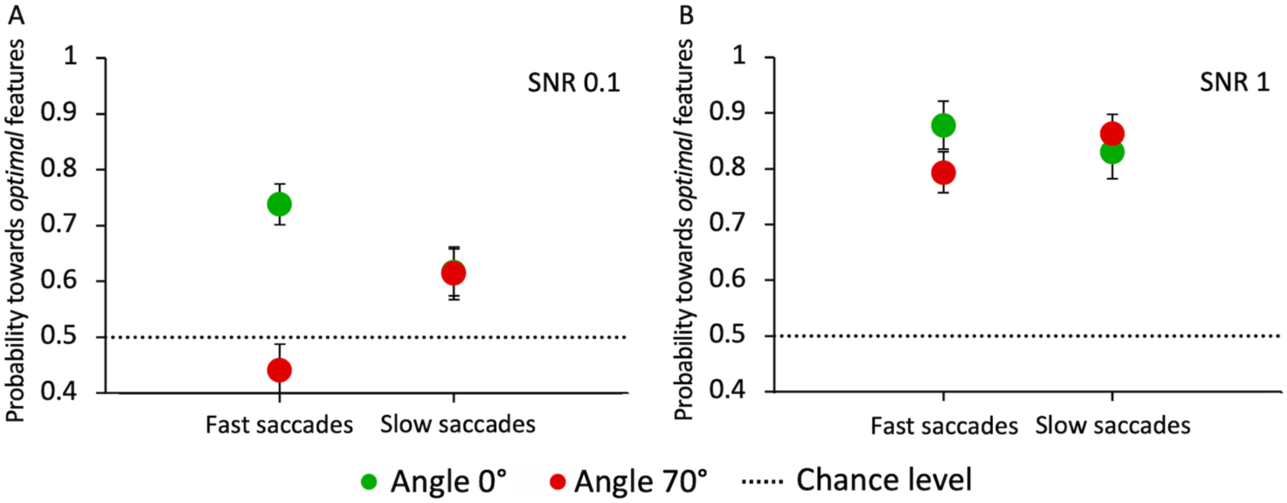

Figure 23. Study 1 – Eye movements experiment: probability of saccades towards the target as a function of latency, at angles 0° and 70°. (A) Probability of saccades towards the target for SNR 0.1. (B) Probability of saccades towards the target for SNR 1. Data are pooled across observers (weighted means with errors). "Fast" and "slow" saccades are determined, participant by participant, based on whether they were below or above the individual median latency respectively. Green circles: angle 0°; red circles: angle 70°. The dashed line indicates the guessing level (0.5 probability). Figure retrieved from (Castellotti et al., 2021).

2.5 Discussion

In this work (Castellotti et al., 2021), we found that a specific set of local features, originally identified based on constrained-entropy maximization criteria (Del Viva et al., 2013), are selected as more *salient* than others even in the absence of any global arrangement, both in psychophysical and oculomotor tasks. In past works, the role of those features in early vision had already been shown, but their involvement in saliency determination is evidenced here for the first time.

Psychophysical results show that few *optimal* isolated features are perceived as more salient than the *non-optimal* features by all participants. Their saliency scales with their luminance-contrast and number when presented alone, and with SNR when surrounded by *non-optimal* features. *Optimal* features are so prominent that just one of them can trigger a preferential choice, after having been seen for only 26 ms., both when it is presented alone and when is surrounded by 9 *non-optimal* features. Luminance contrast values are often considered as a reference for saliency comparisons between stimulus dimensions (Nothdurft, 2000; Nothdurft, 1993b, 1993a). Ten *optimal* features are still preferred when their luminance-contrast is 65% than that of *non-optimal* features. That is, the saliency instantiated by these specific features, is equivalent to the saliency instantiated by *non-optimal* features with a luminance contrast increased by 35%.

The same pattern of results was obtained in the eye movements experiment. Observers preferentially direct their saccades to the compound including *optimal* features (target), with a probability that increases with the proportion of *optimal* features. We did not find evidence, instead, of a systematic reduction of saccade latency with increasing SNR. This is somewhat unexpected considering that the most widespread models of perceptual decisions assume that response latency is inversely proportional to the rate of accumulation of noisy sensory information (Ratcliff & McKoon, 2008), which in turn is directly proportional to the sensory SNR.

When analyzing saliency discrimination performance as a function of saccadic latency, we first observed different effects depending on the angle of presentation of the target-distractor compounds pair and on the SNR. The clear up-down anisotropy found in oculomotor data is arguably connected with stimulus saliency, but still deserves a brief discussion and future investigations. Saccades oriented to the upper visual field had a dramatically reduced latency with respect to the lower visual field, even more pronounced than in previous studies (e.g., Honda & Findlay, 1992). The increased latency for horizontal compared to upward vertical saccades found here might be due to the bilateral presentation of the target-distractor pair, which is known to maximize the Remote Distractor Effect on saccades latency (Benson, 2008; Walker et al., 1997). These phenomena have been attributed to purely oculomotor properties, rather than to visual processing mechanisms (Honda & Findlay, 1992; Walker et al., 1997), coherently with the relative independence, in our data, of the latency on SNR. Finally, the lower performance in the upper hemifield probably reflects the superiority of perceptual discrimination in the lower visual field (see for example Talgar & Carrasco, 2002).

Overall, under reasonable conditions of visibility (high SNR), postponing the response execution (i.e., increasing the time for integration of sensory evidence) does not seem to help to further improve the selectivity for salient features. Previous studies have reported that short-latency saccades were more strongly affected by salient distractors than slower saccades, suggesting that target selection based on saliency (instantiated by luminance or orientation-contrast) could be facilitated for early saccades (Donk & van Zoest, 2008). A similar fast capture exerted by salient features could explain, in our study, the relative independence of discrimination accuracy upon saccadic latency. More generally, the independence of saccadic latency on SNR is consistent with a fast bottom-up mechanism for saliency extraction, like the one proposed by Del Viva and colleagues (Del Viva et al., 2013), rather than a slower and detailed processing of sensory information.

Saccadic precision instead depends strongly on SNR: the higher the SNR the more precise are the saccades directed to the target. Saccades directed to the distractor are instead less precise and further biased in the direction of the target with increasing SNR. This attraction bias toward the salient compound is independent on the validity criterion of saccades chosen in this study.

All together, these results point to a rapid orientation of saccades towards the salient information provided by *optimal* features.

The visual system is capable of detecting very quickly potentially dangerous or very interesting stimuli to activate emotive or fight-or-flight autonomic responses essential for survival (Morris et al., 1999). This analysis does not need, and probably does not use, detailed visual information but needs fast and reliable processing of relevant elements (LeDoux, 1996; Öhman et al., 2001; Perrinet & Bednar, 2015). This processing could take advantage of a quick inspection of different small regions distributed over the image, each providing enough information about the whole scene. For this reason, it could use a constrained maximum-entropy approach to extract a saliency map, that the oculomotor system could use to drive eye movements toward potentially relevant locations (Garcia-Diaz et al., 2012; Itti & Baldi, 2009; Itti & Koch, 2000, 2001; Najemnik & Geisler, 2005, 2008; Schütz et al., 2012; Tatler & Melcher, 2007).

Such rapid and optimal selection of information, devoid of detailed fine-scale color or luminance information (Del Viva et al., 2016), could be sufficient per se to provide salient locations in first viewed scenes that could be followed, only at those locations, by a more detailed analysis. This would require a much larger computational power and may be only possible if performed more slowly and/or on a reduced part of the image. Our hypothesis does not exclude other rapid simultaneous processing of large-scale visual properties, that do not need such compression (e.g., Gegenfurtner & Rieger, 2000).

To conclude, the results presented in this study confirmed that local visual saliency can be determined by the amount of information that local features carry about the visual scene weighed with their processing costs for the system, as predicted by the reference model (Del Viva et al., 2013). They also suggest that these salient features participate early in the visual reconstruction process that must be, at least partly, initiated at the local level.

In this study, saliency has been tested by explicitly asking participants to choose between two different stimuli. At this point, we wondered whether *optimal* features can rapidly and automatically attract the subject's attention in attentional tasks, in which "saliency" is implicitly manipulated rather than explicitly cued.

Chapter 3

Study 2: Information-optimal local features automatically attract covert and overt attention

3. STUDY 2: INFORMATION-OPTIMAL LOCAL FEATURES AUTOMATICALLY ATTRACT COVERT AND OVERT ATTENTION

3.1 Theoretical background and rationale

Visual attention is used to prioritize significant objects in a complex visual environment (Bergen & Julesz, 1983; Hikosaka et al., 1996; Nakayama & Mackeben, 1989; Treisman & Gelade, 1980). Selective visual processes can be set in place automatically and very quickly, as proven by the intrinsic saliency of some visual stimuli which obtain priority processing (*exogenous attention*). A salient stimulus automatically pops out of a visual scene, suggesting that saliency is computed pre-attentively across the entire visual field. This bottom-up saliency is largely independent of the nature of the specific task at hand, it operates very rapidly and is primarily driven by the nature of the stimuli, although it can be influenced by contextual effects of the visual surroundings (e.g., figure–ground). Other types of attentional selective processes can be driven in a top-down manner and influenced by task-dependent cues requiring a voluntary 'effort' and cognitive strategic processes (*endogenous attention*); for example, an instruction like 'look for the red horizontal target'. Both mechanisms can operate in parallel, although their characteristic time courses are different (for a review, see Carrasco, 2011).

As already discussed in the introduction of the Study 1, the principles driving the bottom-up saliency of visual features are still subject of intense debate. Most models proposed for the estimation of the bottom-up saliency map rely on the empirical observation of neurons' biological properties and generally define a-priori the dimensions along which a stimulus' saliency can vary (luminance, edge orientation, etc.). Some approaches compute the local maxima of the contrast of these dimensions with respect to the surround to identify salient features (Nothdurft, 1993b; Treisman, 1985). Eye tracking models, on the other end, have tried to relate visual saliency to the locations where fixations occur (Garcia-Diaz et al., 2012; Itti & Koch, 2000, 2001). In fact, in the presence of multiple information cues in a complex natural scene, the pattern of ocular fixations is often used as an operational definition of the saliency map of the scene (Itti & Borji, 2013).

Here, instead of considering the saliency map as consisting of known elements as addressed in previous studies (luminance, color, etc.), we adopt a different point of view following the approach for efficient extraction of visual features (Del Viva et al., 2013).

In Study 1, we have already shown that participants, when explicitly asked to choose the most salient stimulus, preferred *optimal* features, even when their number or contrast was lower than *non-optimal* features (Castellotti et al., 2021).

In the present work, we implicitly tested the relative saliency of *optimal* and *non-optimal* features by using them as cues in perceptual and oculomotor attentional tasks (Castellotti et al., 2022).

Although the vast majority of spatially-cued attention-orienting tasks (Posner paradigm; Posner, 1980) have used a single cue (for a review, see Carrasco, 2011), in our work, we designed a novel spatial-cueing task, in which two brief peripheral bilateral cues are presented before the target. Few studies have previously proven the efficacy of dual cues of different saliency (in terms of luminance contrast) in the automatic capture of attention toward the most salient one (Kean & Lambert, 2003; Zhao et al., 2012). Here we use one *optimal* feature (deemed salient cue) and one non-*optimal* feature (deemed non-salient cue), which may preferentially attract the observer's attention and eye movements to one location instead of another. Contrast-based saliency of cues was tested as a control for the saliency determined by the specific spatial structure of *optimal* features. That is, in the control condition, the saliency of the cues was manipulated through their relative contrast, presenting one high-luminance (deemed salient) and one low-luminance feature (deemed non-salient) as attentional cues (Kean & Lambert, 2003).

We measured covert attention and gaze-orienting performance with two different tasks in two experiments. The covert-attention task required to identify the orientation of a gabor presented with different contrasts in positions that were cued by an *optimal* or *non-optimal* feature. Exploiting the fact that attention automatically shifts to salient stimuli (Nothdurft, 2002; Theeuwes, 2010; Theeuwes & van der Burg, 2008), and that contrast sensitivity of stimuli presented in attended locations improves (Carrasco, 2006), if *optimal* features are actually salient and able to automatically capture our attention, we expect lower contrast thresholds for targets presented in the position cued by one of them ("saliency cueing effect" in valid trials; Carrasco, 2011; Posner, 1980). On the contrary, if the participants' covert attention is captured on the opposite side of the target (invalid trials), the contrast threshold for the target discrimination should increase.

The gaze-orienting task only required making a saccade toward a visual target. Given that attentional capture to a specific location precede (Deubel & Schneider, 1996; Montagnini & Castet, 2007) and facilitate a subsequent saccadic gaze shift to that location (Kowler et al., 1995; Montagnini & Castet, 2007; Theeuwes & Godijn, 2001), if *optimal* features are actually salient and capture our attention, we expect that saccades latencies will decrease towards targets presented in the position cued by one of them (valid trials). On the contrary, if the participants' attention is captured on the opposite side of the target (invalid trials), the target-directed saccadic latencies will increase. In addition, since salient stimuli can elicit automatic short-latency saccades (Ludwig et al., 2004), we might observe an automatic fast attraction of gaze (overt attention) exerted by our salient cues, irrespective to the target shown.

In both experiments, cueing features are presented for few milliseconds to probe early visual stages, implicated by the reference model.

In both tasks, the cue validity (i.e., the percentage of cases in which the target is presented in the position cued by the salient feature) could be 80% or 50%. The comparison between these two validity conditions, may help disentangling the nature of the attentional processes at play, since a facilitation effect based purely on exogenous attention, hence guided only by the stimulus properties and not willfully monitored (Theeuwes, 1991), should not increase with cue validity (Giordano et al., 2009; Jonides, 1998; Posner et al., 1982). On the other hand, if some cognitive strategic process is at play, we might expect an increased facilitation in valid trials when cue validity is 80%, compared to when the salient cue is uninformative about the target position (cue validity 50%; for a review, see Carrasco, 2011).

3.2 Aim of the study

In Study 2 (Castellotti et al., 2022), we aim to study the bottom-up saliency driven by *optimal* features without explicitly requiring the participants to pay attention to stimulus saliency. We implicitly tested the relative saliency of *optimal* and *non-optimal* features by engaging participants in carrying out covert attentional and gaze-orienting tasks, whose performance might be influenced by the saliency of the task-irrelevant presented features.

3.3 Materials and methods

3.3.1 Covert-attention experiment

3.3.1.1 Participants

Sixteen young naïve adults (10 women, mean age = 27.8 ± 2.3 years) took part in the experiment. Eye movements were monitored on five of them in a separate control session. Written informed consent was obtained from all participants.

3.3.1.2 Apparatus and set-up

Participants were tested individually in a dark room. Stimuli were presented through an ACER computer (Windows 7) using the Psychophysics Toolbox extensions for Matlab (Brainard, 1997). The experiment was displayed on a 120-Hz gamma-corrected CRT Silicon Graphics monitor (1024x768 pixel), subtending 38.5° x 29.5° of visual angle at a viewing distance of 57 cm. To control correct fixation, the left eye position was recorded with an EyeLink 1000 system (SR research - 500 Hz).

3.3.1.3 Stimuli and conditions

In both experimental and control conditions, stimuli preceding the target consisted of two small cues (arrays of 9 x 9 pixels, 0.3 degrees), which could be equally salient (neutral trials) or one more salient than the other (non-neutral trials).

In the experimental condition, cues saliency was based on the spatial arrangement of their black (luminance 4 cd/m^2) and white (44 cd/m^2) pixels. Two types of cues were used: *optimal* features and *non-optimal* features*,* as identified by the reference constrained maximum-entropy model (Del Viva et al., 2013). At each trial, the *optimal* feature, used as "salient cue", was randomly extracted from the set of 50 features selected as the best information-carriers (Figure 8C); whereas the *non-optimal* feature, used as "non-salient cue", was extracted from a set of 50 features selected amongst those with the lowest probability of occurrence in the statistical distribution of all possible features (Figure 8E). Neutral trials (**Figure 24A – left panel**) consisted of the presentation of two *non-optimal* features, deemed as two equally non-salient cues. In non-neutral trials (**Figure 24A – right panel**), one *optimal* and one *non-optimal* feature were presented, that is one cue deemed as more salient than the other according to the reference model. A comparison between the saliency of these two types of features was already assessed in the previous study where observers preferred one *optimal* feature to a *non-optimal* in 70% of cases and ten *optimal* features were still preferred when their luminance-contrast was only 65% than that of ten *non-optimal* features (Castellotti et al., 2021).

In the control condition, the saliency of the cues was exploited through their relative luminance contrast. Neutral trials (**Figure 24B – left panel**) consisted of the presentation of two identical grey features with equal luminance (20 cd/m^2), slightly higher than that of the background, and therefore equally non-salient cues. In non-neutral trials (**Figure 24B – right panel**), one high-luminance and one low-luminance feature were presented, that is one salient and one non-salient cue based on their different luminance. To compare the effect of different cue types on equal grounds, in the control condition the luminance values were set to 20 cd/m^2 and 23 cd/m^2. These values produce a contrast corresponding to the "equivalent contrast" found to match the saliency difference between *optimal* and *non-optimal* features (Castellotti et al., 2021).

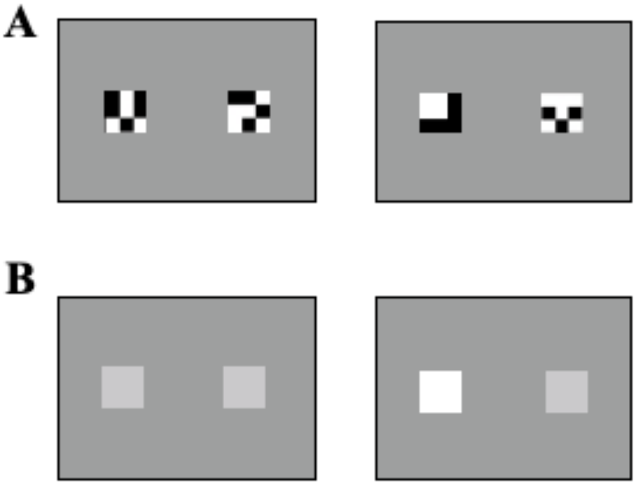

Figure 24. Study 2 – Stimuli and conditions. (A) Experimental condition. Left panel: - Example of neutral trial. Two *non-optimal* features, considered equally non-salient by the reference model, used as baseline for the experimental condition. Right panel – Example of non-neutral trial. One *optimal* (left) and one *non-optimal* (right) feature, one more salient than the other according to the reference model. **(B) Control condition.** Left panel: - Neutral trial. Two cues with the same luminance contrast, considered equally salient, used as baseline for the control condition. Right panel – Example of non-neutral trial. Two cues with different luminance contrasts, one being considered more salient than the other based on their luminance. Features are shown oversized for illustration purposes. The cue stimuli and conditions used in the covert attention and gaze-orienting tasks were the same. Figure adapted from (Castellotti et al., 2022).

3.3.1.4 Procedure

Each trial (**Figure 25**) started with the presentation of a grey display (16 cd/m^2) with a central fixation point, followed by two peripheral cues, bilaterally presented at 5° of eccentricity. After 150 ms of SOA, a tilted gabor appeared at the same location of one of the two cues. The task instructions, given to the participants at the beginning of the experimental session by the experimenter, required them to discriminate the orientation of the gabor (i.e., clockwise or anticlockwise, communicated by a button press) while maintaining fixation. Eight different gabor contrasts, in the range between 0.01 and 0.09, were tested, presented in random order according to a constant stimuli procedure. The contrast values were slightly different across observers based on preliminary rough estimates of individual thresholds. Reaction times were also measured.

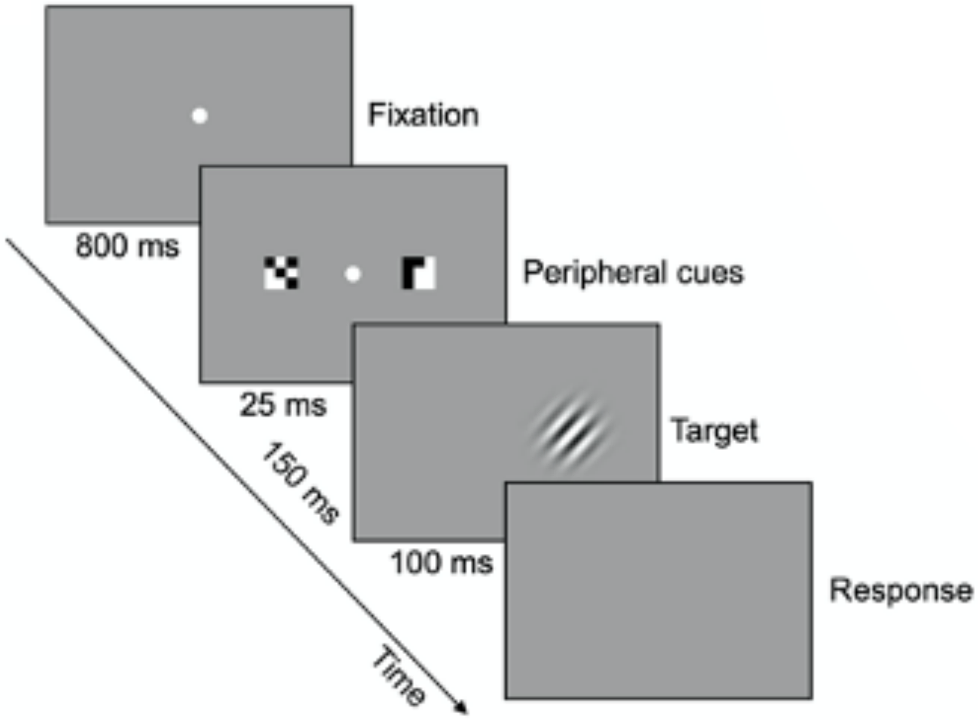

Figure 25. Study 2 – Covert-attention task: procedure. Example of non-neutral trial of the experimental condition, in which the *optimal* feature (the one on the right) precedes the target presentation (valid trial). Target is a ±20°-tilted gabor (40 pixels large, 1.5 cm diameter, spatial frequency of 2 cycles per degree, sigma of 5.7, phase of 0.5π). Cues and targets are shown oversized for illustration purposes. Figure adapted from (Castellotti et al., 2022).

In both experimental and control conditions, 224 neutral and 500 non-neutral trials were presented. In neutral trials, the two cues were equally salient and uninformative about the position of the following target. Non-neutral trials could be “valid trials” or “invalid trials”, based on the position of the (deemed) salient cue with respect to the following target. In valid trials, the salient cue (*optimal* feature or high-luminance cue) was presented at the same location of the target, whereas, in the invalid trials, the salient cue was presented on the opposite side of the gabor, which therefore appeared after the non-salient cue (*non-optimal* feature or low-luminance cue).

The percentage of cases in which the target was presented in the position cued by the salient feature was defined as “cue validity”. 250 non-neutral trials have 50% cue validity (125 valid trials and 125 invalid) and the other 250 trials have 80% cue validity (200 valid trials and 50 invalid).

Data for each participant were collected in two sessions, one for each cue validity, performed in random order across participants. In each session, participants performed one block of neutral trials and two blocks of non-neutral trials, one for the experimental and one for the control condition, presented in random order. Each participant performed 1448 trials in total.

A subset of participants performed an additional separate session with one block of neutral trials and one block of non-neutral trials with 80% cue validity while their eye movements were recorded. This session allowed us to control that the results obtained in the main covert-attention task were not due to uncontrolled saccades toward the salient cue, which could potentially reduce the perceptual threshold by reducing the distance of the gabor from the fovea.

3.3.1.5 Data processing

Percent correct data were fitted (MLE) with a cumulative Gaussian error function. For each participant, condition (experimental and control), cue validity (50% and 80%), and trial type (neutral, valid, and invalid) thresholds were calculated as the target contrasts yielding 80% correct responses. Thresholds for neutral trials of each condition were used as baseline for non-neutral trials. That is, they were subtracted from those obtained in valid and invalid trials, and the result divided by them to provide a measure of the percentage increase or decrease of contrast thresholds in non-neutral trials.

In the control session, we considered a saccade execution any shift of gaze position further than 2 degrees from the fixation point.

3.3.2 Gaze-orienting experiment

3.3.2.1 Participants

Sixteen naïve young adults (11 women, mean age = 27.6 ± 1.9 years) participated in the experiment. Written informed consent was obtained from all participants (all different from those of the covert-attention experiment). The local ethics committee (*Comité d'éthique d'Aix-Marseille Université, ref: 2014-12-3-05*) approved the experimental paradigm, which complied with the Declaration of Helsinki.

3.3.2.2 Apparatus and set-up

Each participant was tested individually in a dark room. Stimuli were presented through a MacPro computer (OS 10.6.8), using the Psychophysics Toolbox (Brainard, 1997) and the Eyelink Toolbox extensions (Cornelissen et al., 2002) for Matlab. The experiment was displayed on a 120-Hz CRS Display++ LCD monitor (1920×1080 pixel), subtending 70x40 degrees at a viewing distance of 57 cm. Participants' viewing was binocular, but only the right eye was recorded by an Eyelink 1000 video-based eye tracker (1 kHz). The observers' head was stabilized with a chin- and forehead rest.

3.3.2.3 Stimuli and conditions

Cue stimuli and conditions were the same as in the covert-attention experiment (Figure 24).

3.3.2.4 Procedure

Each trial (**Figure 26**) started with the presentation of a grey display (16 cd/m^2), with a central fixation point, followed by two peripheral cues, bilaterally presented at 5° of eccentricity. After 150 ms of SOA, the target appeared on the left or on the right of the center, in the same location of one of the two cues. At the beginning of the experimental session, participants were instructed to make a saccade towards the target as quickly as possible. The following trial started only after participants resumed fixation.

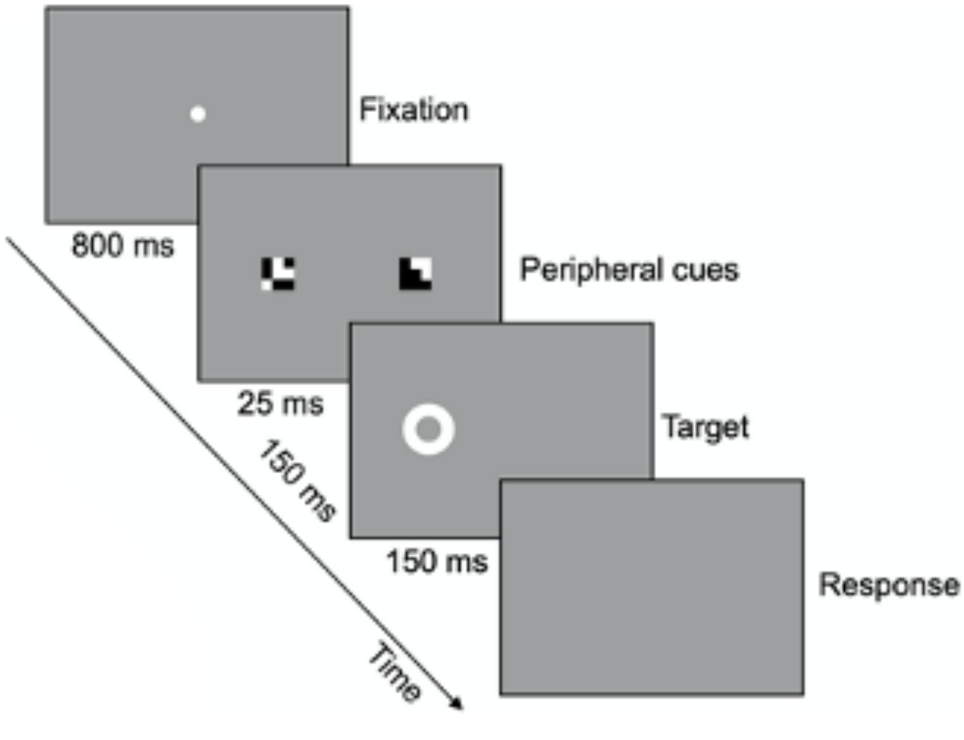

Figure 26. Study 2 – Gaze-orienting task: procedure. Example of non-neutral trial of the experimental condition, in which the *optimal* feature (the one on the right) is presented on the opposite side of the saccadic target (invalid trial). Target is a circular white placeholder, 100% contrast, 9 pixels large (0.3 cm diameter). Cues and targets are shown oversized for illustration purposes. Figure adapted from (Castellotti et al., 2022).

In both experimental and control conditions, 200 neutral and 500 non-neutral trials were presented. 250 non-neutral trials have 50% cue validity, the other 250 trials have 80% cue validity.

Data were collected in two sessions, one for each cue validity. In each session, 100 neutral trials and 250 non-neutral trials for each condition were tested. Each participant performed 1400 trials in total.

3.3.2.5 Data processing

Oculomotor parameters were extracted by using ad hoc software in Matlab. Recorded horizontal and vertical gaze positions were low-pass filtered with a Butterworth (acausal) filter of order 2 with a 30-Hz cutoff frequency and then numerically differentiated to obtain velocity measurements. An automatic conjoint acceleration and velocity threshold method was used to detect saccades (Damasse et al., 2018). Aberrant trials, without recorded saccades (e.g., due to a long blink), were excluded (less than 3% of all saccades).

In each trial, we considered a "*regular saccade*" the first detected saccade with a latency (with respect to target onset) longer than 80 ms (Carpenter, 1988; Fischer & Ramsperger, 1984) and shorter than 500 ms (~95% of the first detected saccades overall), and an amplitude larger than 2 degrees (40% of the entire eccentricity). For each regular saccade, we estimated latency and direction. Each regular saccade was labeled as "*correct*" if directed to the target or as "*erroneous*" if directed towards the opposite side of the target. The mean latencies of *correct* saccades calculated in neutral trials of each condition were used as baseline for non-neutral trials. That is, they were subtracted by the latency values obtained in valid and invalid trials, and the result divided by them, yielding to a measure of the percentage increase or decrease of saccadic latencies. The percentage of saccade direction errors relative to the total number of saccades in each condition was also measured.

Trials' inspection revealed a consistent number of saccades faster than 80 ms. These values are considered too fast to be due to the onset of the target (Carpenter, 1988) and are probably generated by the presentation of the cues. Therefore, saccades with latency shorter than 80 ms and longer than -70 ms (with respect to the target onset, corresponding to a latency of 80 ms with respect to cue onset), and amplitude of at least 1 degree (a widely used arbitrary threshold-amplitude to exclude fixational microsaccades (Otero-Millan et al., 2008), were categorized as "*anticipatory saccades*", potentially elicited by the cue. In most cases, an anticipatory saccade preceded a regular saccade, which continued in the same direction or reversed direction to reach the target. The percentage of anticipatory saccades over the total number of saccades in each condition was measured. Since not all participants performed anticipatory saccades, weighted averages were computed, taking into account their number in each condition, and percentages over the number of anticipatory saccades in non-neutral trials were calculated to estimate their preferential direction.

3.4 Results

3.4.1 Covert-attention experiment

Contrast thresholds

The performance of one participant for valid, invalid, and neutral trials in the experimental condition is shown in **Figure 27A**, as an example. At the lowest contrast, performance is at chance-level and then increases with target contrast in all trial types. However, performance is higher in valid trials than in neutral trials. Invalid trials have the lowest performance.

Contrast thresholds averaged over 16 participants are reported in **Figure 27B** (experimental condition) and **26C** (control condition). In the experimental condition, average gabor' contrast thresholds for blocks with 50% cue validity are 0.045 ± 0.004 (SEM), 0.042 ± 0.004, and 0.048 ± 0.004 for neutral, valid, and invalid trials, respectively. For blocks with 80% cue validity average thresholds are 0.047 ± 0.004, 0.044 ± 0.004, and 0.051 ± 0.004 for neutral, valid, and invalid trials, respectively. Average percentage threshold changes in valid and invalid trials compared to baseline values for the experimental condition are reported in **Figure 27D**. Contrast thresholds with respect to baseline decrease in valid trials and increase in invalid trials, both for 50% (-6.19 ± 2% and +5.24 ± 2.5%, respectively) and 80% cue validity (-4.81 ± 2.2% and +10.72 ± 2.8%, respectively).

Results for the luminance control condition are very similar: averaged target contrast thresholds for 50% cue validity are 0.044 ± 0.004, 0.041 ± 0.005, and 0.048 ± 0.005 in neutral, valid and invalid trials, respectively. Average contrast thresholds for 80% cue validity are 0.045 ± 0.004, 0.042 ± 0.004 and 0.049 ± 0.004 in neutral, valid and invalid trials, respectively. Average relative threshold changes in the control condition are reported in **Figure 27E**. As in the experimental condition, percentage contrast thresholds with respect to baseline decrease in valid trials and increase in invalid trials, both for 50% (-6.41 ± 2.3% and +8.69 ± 2.1%, respectively) and 80% cue validity (-6.80 ± 2.4% and +10.39 ± 2.2%, respectively). Although the differences between means are small, three-ways ANOVA analysis – with factors: condition (two levels: experimental vs. control), trial type (three levels: neutral, vs. valid, vs. invalid), and cue validity (two levels: 50% vs. 80%) – evidences a significant main effect of type of trial (F(2, 30) = 67.53, p < 0.001, $\eta^2 = 0.03$ – *small* effect size) but no effect of either condition (F(1,15) = 1.03, p = 0.3, $\eta^2 = 0.002$) or cue validity (F(1,15) = 1.84, p = 0.19, $\eta^2 = 0.001$) on the perceptual threshold. No interactions between the three factors have been found. Pairwise comparisons *t*-tests (with Bonferroni corrections), performed to assess significant differences between the means of different trial types are reported in the caption of Figure 27.

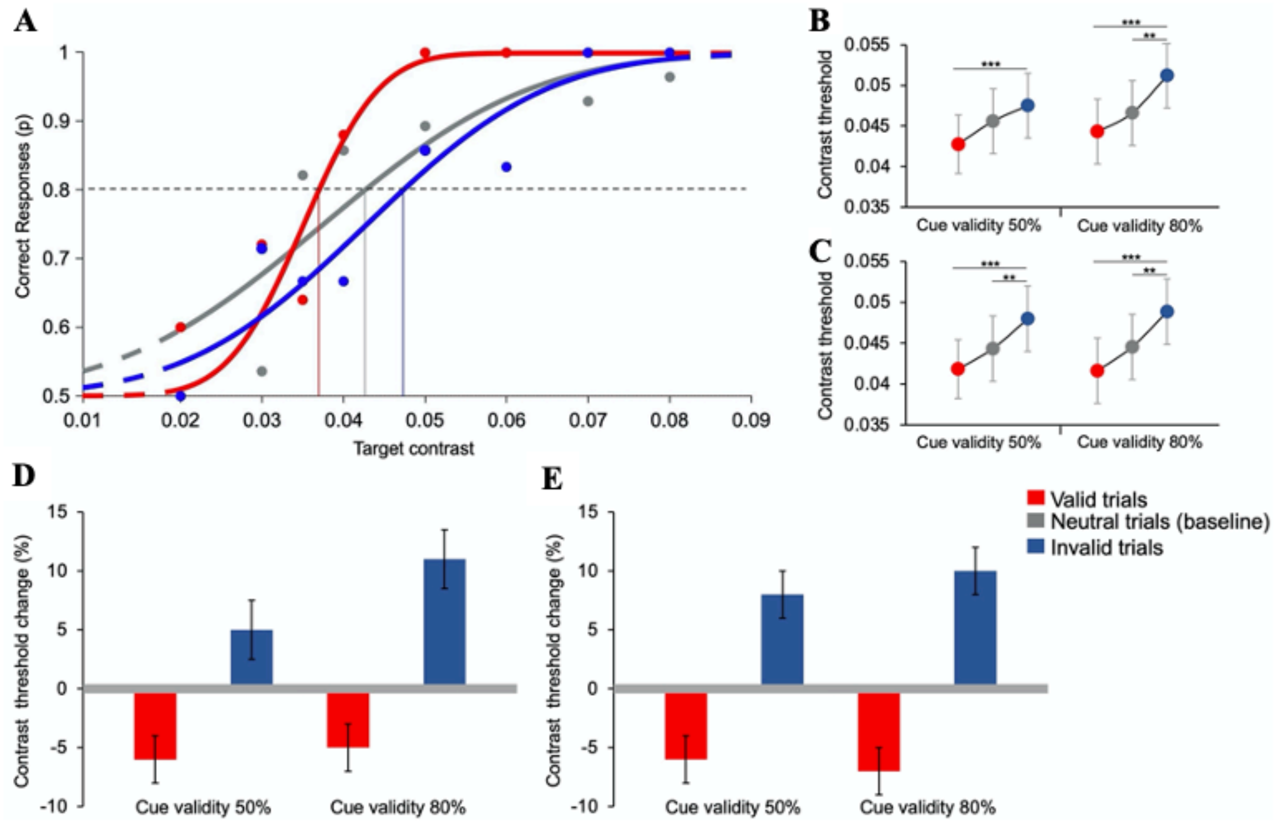

Figure 27. Study 2 – Covert-attention experiment: contrast thresholds in valid, invalid, and neutral trials. (A) Example of correct responses as a function of target contrast for one participant. Performance obtained in the experimental condition with 80% cue validity. Data are obtained from 224 neutral trials, 200 valid trials, and 50 invalid trials. Filled circles represent the proportion of times in which the participant correctly discriminated the gabor orientation presented with a specific contrast. The curves represent cumulative Gaussian error fits of the data. The vertical lines represent the contrast values yielding 80% correct responses (dashed line). **(B) Experimental condition.** Group average contrast thresholds in neutral (grey), valid (red), and invalid (blue) trials. Post-hoc *t*-tests (Bonferroni correction) show a significant difference between valid vs. invalid trials for 50% cue validity (*t*(2) = 5.12, *p* < 0.001). They show also significant differences between valid vs. invalid trials (*t*(2) = 6.05, *p* < 0.001), and invalid vs. neutral trials (*t*(2) = -4.07, *p* = 0.002) for 80% cue validity. **(C) Control condition.** Group average contrast thresholds in neutral (grey), valid (red), and invalid (blue) trials. Post-hoc *t*-tests (Bonferroni correction) show a significant difference between valid vs. invalid trials (*t*(2) = 6.52, *p* < 0.001), and invalid vs neutral trials (*t*(2) = -3.73, *p* = 0.006) for 50% cue validity. They also show significant differences between valid vs. invalid trials (*t*(2) = 6.26, *p* < 0.001), and invalid vs. neutral trials (*t*(2) = -3.64, *p* = 0.009) for 80% cue validity. Asterisks mark statistically significant pairwise comparisons across trial types: ***p* < 0.01, ****p* < 0.001. Error bars are SEM. **(D) Experimental condition.** Group average threshold changes in valid (red) and invalid (blue) trials compared to the baseline (grey line, neutral trials with two *non-optimal* features). **(E) Control condition.** Group average threshold changes in valid and invalid trials compared to the baseline (neutral trials with two identical low-luminance grey features). Figure retrieved from (Castellotti et al., 2022).

Average contrast thresholds changes of participants in the 50% cue validity condition were not statistically different depending on whether this condition was performed before or after the 80% cue validity condition (Independent *t*-test – experimental condition: valid: $t(14) = -0.32$, $p = 0.7$, invalid: $t(14) = 0.47$, $p = 0.6$; control condition: valid: $t(14) = -0.67$, $p = 0.5$, invalid: $t(14) = -0.60$, $p = 0.6$), thus arguing against a prominent role for sessions' order.

Reaction times analysis showed no differences between valid and invalid trials; indeed, in all types of trials, conditions and cue validities, average reaction times are very long (~600ms).

In the additional session of the experimental condition with 80% cue validity, in which observers' fixation was monitored, contrast thresholds change was -5.46 ± 2.7% in valid trials and +10.7 ± 2.7% in invalid trials, comparable to those obtained in the first participation without fixation control (respectively: -5.6 ± 2.8%, +13.4 ± 4.6%). On average, only one or two saccades over 474 trials were detected in these observers.

3.4.2 Gaze-orienting experiment

Saccadic latencies

Latencies of *regular* saccades directed to the target (*correct*) differ across different types of trials. Saccadic latencies averaged over 16 participants are reported in Figure **28A** (experimental condition) and **28B** (control condition).

In the experimental condition, for 50% cue validity, average latencies are 167 ± 5.9 ms (SEM), 161 ± 5.7 ms, and 174 ± 6.7 ms in neutral, valid and invalid trials, respectively. Average latencies for 80% cue validity are 162 ± 5.3 ms, 155 ± 5.6 ms, and 171 ± 6.1 ms in neutral, valid and invalid trials, respectively. The average latency changes in valid and invalid trials relative to baseline values are reported in **Figure 28C**. Percentage latencies changes relative to baseline decrease in valid trials and increase in invalid trials, both for 50% (-3.5 ± 1% and +3.7 ± 2%, respectively) and 80% cue validity (-4.1 ± 1% and +5.5 ± 2%, relatively).

Results of the luminance control condition are comparable to those of the experimental condition. The mean saccadic latency is 163 ± 6.6ms in neutral trials, 155 ± 6.4ms in valid trials, and 176 ± 7.5ms in invalid trials, for blocks with 50% cue validity. For blocks with 80% cue validity mean saccadic latency is 158 ± 5.5ms in neutral trials, 150 ± 5.4ms in valid trials, and 168 ± 7.2ms in invalid trials. Average percentage latency changes are reported in **Figure 28D.** Percentage latencies with respect to baseline decrease in valid trials and increase in invalid trials (50% cue validity: -4.9 ± 1%

and +7.4 ± 1%, respectively; 80% cue validity: -5.5 ± 1% and +6.20 ± 1%, respectively). Also in this case, results for blocks with 50% cue validity are similar to those obtained for 80% cue validity.

Three-ways ANOVA analysis – with factors: condition (two levels: experimental vs. control), trial types (three levels: neutral, vs. valid, vs. invalid), and cue validity (two levels: 50% vs. 80%) – evidences a significant main effect of types of trial (F(2,30) = 34.11, p < 0.001, η^2 = 0.08 – *small* effect size) but no effect of either condition (F(1,15) = 2.3, p = 0.15, η^2 = 0.005) or cue validity (F(1,15) = 2.69, p = 0.12, η^2 = 0.01) on saccadic latency. No interactions between the three factors have been found. Pairwise comparisons t-tests (with Bonferroni corrections), performed to assess significant differences between the means of the different trial types, are reported in the caption of Figure 27.

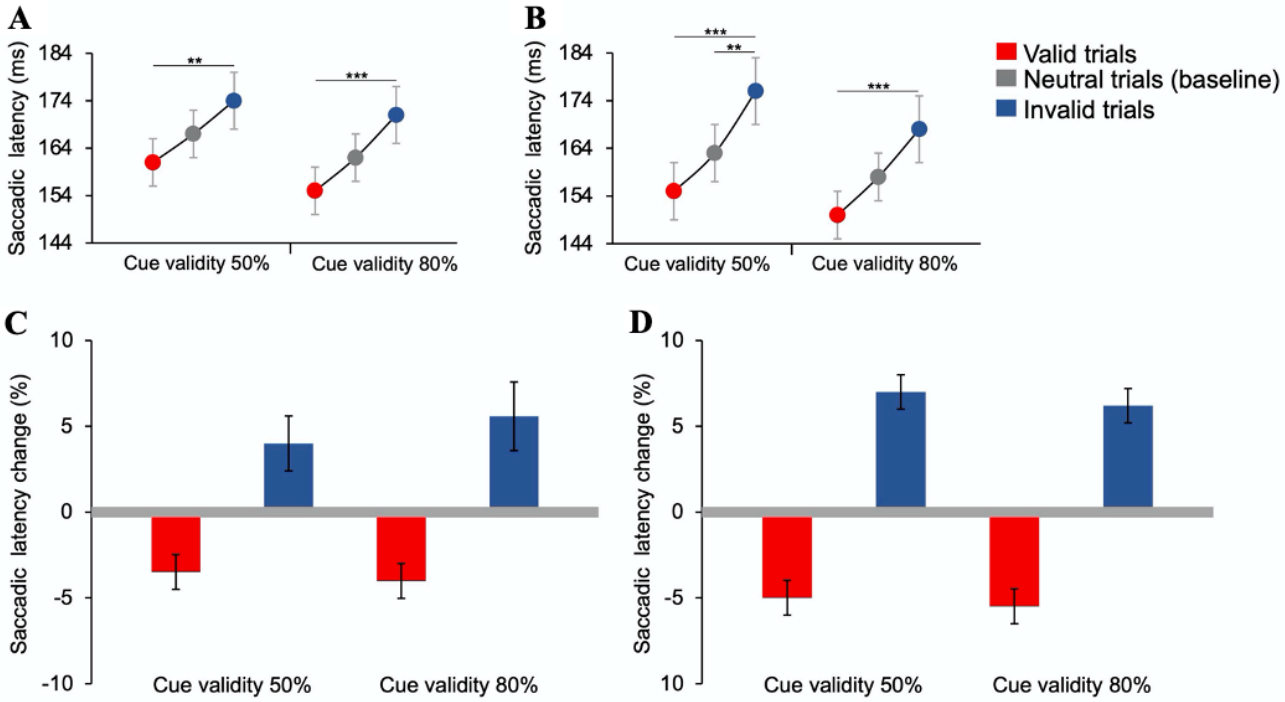

Figure 28. Study 2 – Gaze-orienting experiment: saccadic latencies in neutral, valid, and invalid trials. (A) Experimental condition. Group average latencies in neutral (grey), valid (red), and invalid (blue) trials. Post-hoc t-*t*ests (Bonferroni correction) show a significant difference between valid vs. invalid trials for 50% (t(2) = 3.99, p = 0.008) and 80% cue validity (t(2) = 5.21, p < 0.001). **(B) Control condition.** Group average latencies in neutral (grey), valid (red), and invalid (blue) trials. Post-hoc t-tests (Bonferroni correction) show significant differences between valid vs. invalid trials (t(2) = 6.72, p < 0.001), and invalid vs. neutral trials (shown with a line near the baseline; t(2) = -4.05, p = 0.007) for 50% cue validity. They also show a significant difference between valid vs. invalid trials (t(2) = 6.16, p < 0.001) for 80% cue validity. Asterisks mark statistically significant pairwise comparisons across trial types: **p < 0.01, ***p < 0.001. Error bars are SEM. **(C) Experimental condition.** Group average latencies changes in valid (red) and invalid (blue) trials

compared to the baseline (grey line, neutral trials with two *non-optimal* features). **(D) Control condition.** Group average latencies changes in valid and invalid trials compared to the baseline (neutral trials with two identical low-luminance grey features). Figure retrieved from (Castellotti et al., 2022).

A possible effect of sessions' order does not seem very likely. Indeed, averaged saccadic latency changes in the 50% cue validity condition did not change significantly if the participants had first performed the session with 80% cue validity or the other way around (Independent *t*-test – experimental condition: valid: $t(14) = 1.07$, $p = 0.3$, invalid: $t(14) = 0.76$, $p = 0.4$; control condition: valid: $t(14) = 0.50$, $p = 0.6$, invalid: $t(14) = 0.37$, $p = 0.7$).

Saccadic direction errors

In the gaze-orienting task, although participants were instructed to make an accurate saccade towards the visual target, there was a small proportion of trials in which the participants moved their eyes towards the opposite side of the target (*erroneous* saccades). Independently on the condition and the cue validity, a small percentage of direction errors relative to the total number of saccades is present in all trial types, even in neutral trials, in which the cues preceding the target were equally salient. Interestingly, the proportion of erroneous saccades decreases in valid trials and increases in the invalid ones with respect to neutrals (baseline) (**Figure 29**). Specifically, in the experimental condition with 50% cue validity (**Figure 29A, left**), there are, on average, 2.2 ± 0.8% (SEM) erroneous saccades in neutral trials, and only 0.4 ± 0.1% in valid trials. Instead, in invalid trials, there are 4.3 ± 0.7% direction errors. Similarly, in trials with 80% cue validity (**Figure 29A, right**), there are 2.9 ± 1.1%, 1.1 ± 0.5%, and 7.1 ± 0.5% errors in neutral, valid and invalid trials, respectively.

The same pattern of results holds for the control condition. In trials with 50% cue validity (**Figure 29B, left**), there are 2 ± 0.9%, 0.2 ± 0.1% and 5.3 ± 1% direction errors in neutral, valid and invalid trials, respectively. In trials with 80% cue validity **(Figure 29B, right**), there are 2.7 ± 0.8%, 0.8 ± 0.3%, and 6 ± 1.3 %, direction errors in neutral, valid and invalid trials, respectively.

Friedman non-parametric test (for binomial distributed data) confirms that there is an effect of trial type and that the proportion of direction errors is significantly higher in invalid trials compared to valid ones ($\chi2(11) = 78.08$, $p < 0.001$, W = 0.3). Pairwise comparisons with Conover post-hoc tests (with Bonferroni corrections), performed to assess significant differences between the means of the different trial types, are reported in the caption of Figure 29.

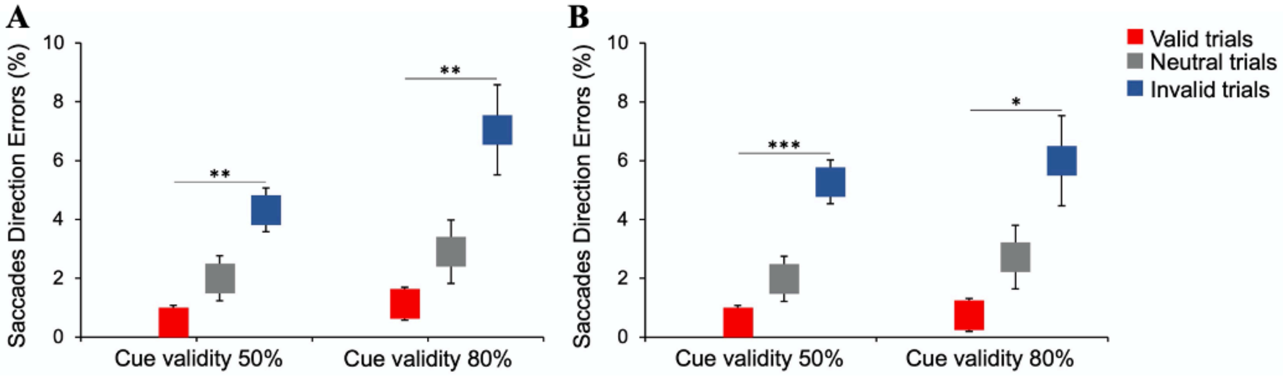

Figure 29. Study 2 – Gaze-orienting experiment: percentage of saccadic direction errors in neutral, valid, and invalid trials. (A) Experimental condition. Post-hoc Conover tests (Bonferroni corrections) show that the percentage of saccades direction errors in valid trials (red) is lower than that in invalid trials (blue) for 50% ($t(11) = 4.05$, $p < 0.01$) and 80% cue validity ($t(11) = 4.02$, $p < 0.01$). **(B) Control condition.** Post-hoc Conover tests (Bonferroni correction) show that the percentage of saccades direction errors in valid trials is lower than that in invalid trials for 50% ($t(11) = 5.02$, $p < 0.001$) and 80% cue validity ($t(11) = 3.56$, $p < 0.05$). Asterisks mark statistically significant pairwise comparisons across trial types: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$. Error bars are SEM. Figure retrieved from (Castellotti et al., 2022).

Anticipatory saccades

As shown in **Figure 30**, in each condition (experimental and control) and cue validity (50% and 80%), there is a high percentage of anticipatory saccades with respect to the total number of saccades, that is very early saccades with respect to stimulus onset. Statistical analyses, reported in the caption of Figure 30, show that the percentage of anticipatory saccades changes across trial types. Not surprisingly, the number of anticipatory saccades is always higher in non-neutral trials (i.e., valid and invalid trials) compared to neutral trials (percentages reported over each vertical bar in Figure 30). No differences across experimental and control conditions emerge. In the experimental condition there are slightly more anticipatory saccades for 80% than 50% cue validity, whereas there are no differences across cue validities for the control condition.

Anticipatory saccades in non-neutral trials clearly show a preferential direction. In the experimental condition, they are mainly directed to the *optimal* feature, with no difference between cue validities. A similar percentage of preferential direction is obtained in non-neutral trials of the control condition, where anticipatory saccades are mainly directed to the high-luminance feature, again not depending on cue validity.

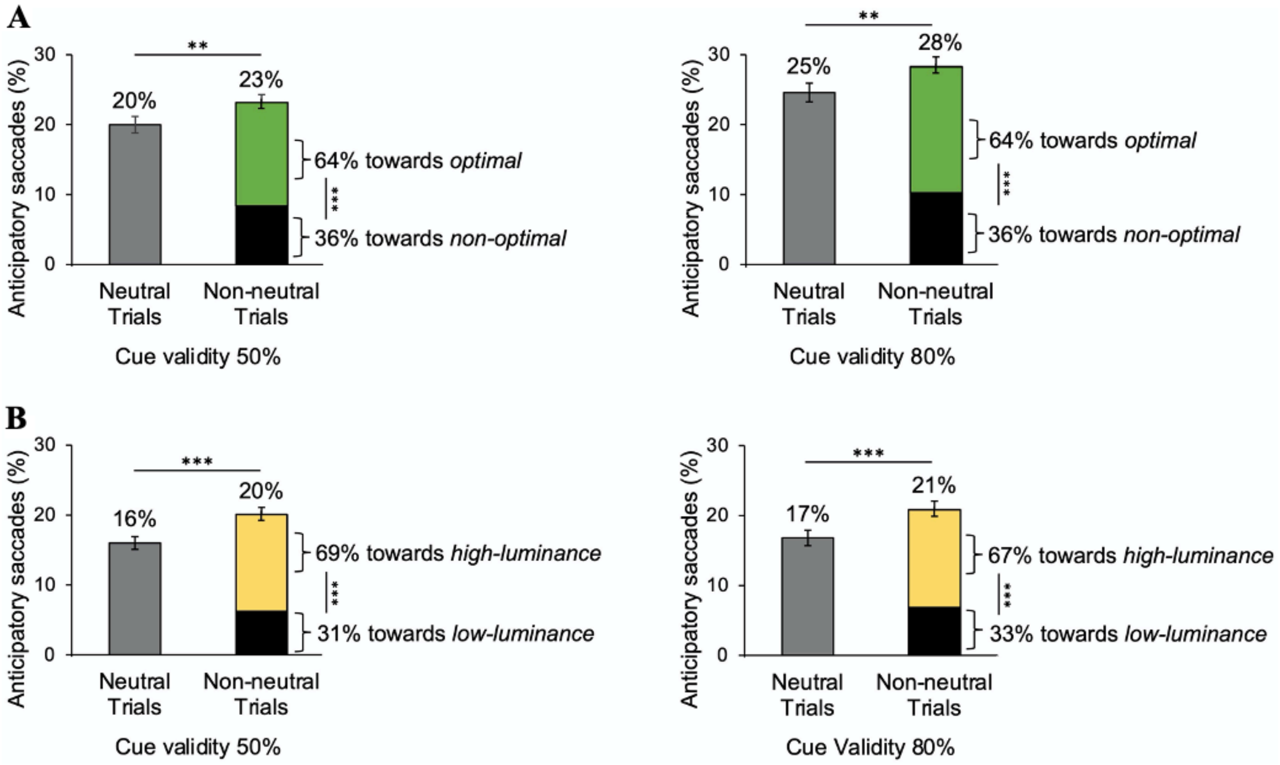

Figure 30. Study 2 – Gaze-orienting experiment: percentage and direction of anticipatory saccades in neutral and non-neutral trials. Percentages shown above bars are computed over the total number of saccades. Percentages shown on the sides of green/black and yellow/black bars are computed over the number of anticipatory saccades done in non-neutral trials and represent the preferential direction towards which these saccades are directed. Binomial data were considered as normally distributed due to the numerosity of observations for each trial type (>30). **(A) Experimental condition.** The percentage of anticipatory saccades in non-neutral trials (green/black bars) is higher than that in neutral (grey bars), for 50% (left panel; $z = 2.33$, $p = 0.009$) and 80% cue validity (right panel; $z = 2.70$, $p = 0.003$). Anticipatory saccades in non-neutral trials are preferentially directed to the *optimal* feature (green) compared to the *non-optimal* one (black) (50% cue validity – $z = 11.27$, $p < 0.001$; 80% cue validity – $z = 12.57$, $p < 0.001$). **(B) Control condition.** The percentage of anticipatory saccades in non-neutral trials (yellow-black bars) is higher than that in neutral (grey bars), for 50% (left panel; $z = 3.42$, $p < 0.001$) and 80% cue validity (right panel; $z = 3.30$, $p < 0.001$). Anticipatory saccades in non-neutral trials are preferentially directed to the high-luminance feature (yellow) compared to the low-luminance one (black) (50% cue validity: $z = 14.39$, $p < 0.001$; 80% cue validity: $z = 13.22$, $p < 0.001$). Asterisks mark statistically significant pairwise comparisons across trial types: $^{**}p < 0.01$, $^{***}p < 0.001$. Error bars are SEM. Figure retrieved from (Castellotti et al., 2022).

3.5 Discussion

In the present study, we tested the automatic capture exerted by a specific set of local features deemed salient, originally identified as optimal information carriers based on constrained-entropy maximization criteria (Del Viva et al., 2013). Differently to the Study 1 (Castellotti et al., 2021), here, the *optimal* features saliency is implicitly addressed by measuring participants' performance in perceptual and oculomotor dual-cueing attentional tasks where they are used as cues.

Results of the covert-attention task show that when the target is cued by an *optimal* feature its contrast threshold decreases, while when the target is presented on the opposite side of the *optimal* cue its contrast threshold increases. Since contrast sensitivity improves at attended locations (Carrasco, 2006), the effect found here could be attributed to the attentional capture of *optimal* features towards the location where they are shown, being more salient than the others. Since this effect is not due to eye movements to the target, it can be attributed to covert attention.

The saliency-based capture of *optimal* features is also seen in the overt response in the gaze-orienting task. First, latencies of *regular* saccades towards the target decrease when cued by *optimal* features. As the dynamics of ocular movements is known to be influenced by attentional factors (Kowler et al., 1995; Montagnini & Castet, 2007), this can be seen as indication that *optimal* features are perceived as a potentially salient stimulus to be analyzed. Saccades directed to the target cued by *non-optimal* feature are slower, probably because the system had already allocated attention and was ready to direct the gaze on the opposite side and has therefore to re-allocate its resources.

The attentional capture exerted by the *optimal* features is also reflected in the number of errors in saccade direction. We find that some *regular* saccades were not directed to the target, despite task requirements. When the target is cued by an *optimal* feature very few errors occur, compared to when the position of the *optimal* cue and the target do not match. This indicates that an *optimal* feature attracts the participant's attention overtly, triggering a saccade to a location, that in some trials does not allow a correction based on the target location, resulting in the gaze landing on the opposite side of the target.

Finally, anticipatory saccades, which are considered too fast to be due to the onset of the target and are probably generated by the presentation of the cues (Carpenter, 1988), are more numerous when the two cues differ in saliency (non-neutral trials with an *optimal* and a *non-optimal* cue) rather than when they are equally salient (neutral trials, two *non-optimal* cues). This fast oculomotor response might be due to an imbalance of mutual inhibition between neural populations representing the two locations, possibly occurring in the superior colliculus (Goffart et al., 2012). Moreover, in non-neutral trials, anticipatory saccades are mainly directed to the side where the *optimal* feature is presented, providing further support for the fast, automatic attraction exerted by the *optimal* features.

Overall, the results of our experiments reveal the presence of a "saliency-based cueing effect", in which the participants' covert and overt attention is attracted by the *optimal* features. That is, *optimal* features result to be used as salient attentional cues by our visual system.

In both tasks and conditions, there is no evidence of an increase of the attention-grabbing effect with cue validity. This is in agreement with most other studies, showing that, unlike endogenous attention, exogenous attention is automatic and unaffected by cue validity (Carrasco, 2011; Giordano et al., 2009); that is, attention capture by *optimal* features' seems to be automatic and guided by the exogenous properties of the features. However, the peripherical cue position and the brevity of the SOA may have precluded an emergence of endogenous effects, which are usually manipulated by central cues and need more time to occur compared to exogenous effects (Carrasco, 2011; Giordano et al., 2009; Müller & Rabbitt, 1989; Wright & Ward, 1994). Kean and colleagues (2003) have found a somewhat counterintuitive effect, whereby attention is captured by the most salient cue only when the cues are irrelevant to task (i.e., cue validity 50%), but not when they indicate the critical location for attentional allocation to the target (i.e., cue validity 80%) (Kean & Lambert, 2003). However, their participants were informed of the contingent relationship between the bright cue and the target location, and expectancy has been shown to attenuate the automatic attention capture (Lambert et al., 1987). Our participants were completely unaware of the cues' predictivity (or non-predictivity) relative to the goal location, so attention capture is not expected to decrease.

The vast majority of the studies investigating spatial-cueing effects on automatic capture of attention have used a single peripheral cue (for a review, see Carrasco, 2011; Chica et al., 2014). Here instead we chose to present two simultaneous peripheral cues, with either the same or different assumed saliency, to test the power of the model-predicted *optimal* features in an implicit competition to attract the participants' attention. To our knowledge, only one other study has used this dual-cue paradigm to study gaze-orienting task with luminance-based cues (Kean & Lambert, 2003). This dual cues saliency manipulation is also directly comparable with that used in Study 1 (Castellotti et al., 2021), where participants could discriminate between the saliency of two stimuli, very similar to those used here, but, unlike here, were explicitly asked to do so. Given our result, such a double spatial-cueing paradigm may be a useful general tool to test the saliency of two stimuli, also ontologically different from each other, by directly comparing their ability to capture attention. Note however that, the latency advantage found here is similar to that elicited by a single uninformative cue versus a non-cued target location (Briand et al., 2000; Danziger & Kingstone, 1999; Tepin & Dark, 1992), under comparable experimental conditions and at short SOAs. These findings seem to suggest that the more salient of two peripheral cues elicits an attention-capturing effect of a similar magnitude to that of a single peripheral stimulus (Kean & Lambert, 2003).

Very interestingly, in both tasks, all the effects found with *optimal* vs. *non-optimal* features are comparable to those obtained with cues of different luminance. This suggests that the saliency provided by *optimal* features is comparable to that of high-luminance cues, if compared on equal grounds (see paragraph 3.3.1.3). Had we used a larger luminance difference between the lighter and the darker cues, we might have obtained a more pronounced saliency effect than with our *optimal* features, but the saliency would have not been comparable. Note that the saccadic latency advantage for the locations cued by high-luminance with respect to low-luminance features is comparable to that obtained by Kean et al. (2003) (Kean & Lambert, 2003).

Our findings confirm that the set of *optimal* features identified by that reference model are indeed more salient than others also when used as implicit spatial cues in covert and overt attention tasks. We argue that the saliency map provided by these features seem thus to be used to automatically guide attention and eye movements towards informative locations.

To further analyzed *optimal* features saliency and eye movement properties, we followed the literature showing that salient features influence the path of saccades, usually inducing saccadic curvature toward or away from their location (for reviews see Walker & McSorley, 2008; Van der Stigchel, 2010). In the next study, we then tested whether the presentation of an *optimal* feature along with a target interferes in a quick and automatic manner with the ongoing oculomotor programming of the target-directed saccade.

Chapter 4

Study 3: Saccadic trajectories deviate toward or away from optimally informative visual features

4. STUDY 3: SACCADIC TRAJECTORIES DEVIATE TOWARD OR AWAY FROM OPTIMALLY INFORMATIVE VISUAL FEATURES

4.1 Theoretical background and rationale

As demonstrated by classic studies on eye movements, the trajectory taken by the eye to reach a target position does not follow a straight line between the saccade starting and ending point (Sheliga et al., 1994; Yarbus, 1967). In recent years, a consistent number of studies demonstrated that the magnitude and direction of this natural saccadic curvature can be modulated by the presence of a competing distractor stimulus presented along with the saccade target. Indeed, visual distractors may cause a deviation either *toward* (attraction) or *away* from (repulsion) their location (for reviews see Walker & McSorley, 2008; Van der Stigchel, 2010).

Factors determining the direction of the curvature are still under investigation. Some studies suggested that the spatial distance between target and distractor modulates the curvature. For example, saccade trajectories tended to deviate toward the distractor location when this was presented close to the target, whereas trajectories deviated away from the distractor when presented closer to fixation (Van der Stigchel et al., 2007a). Differences also emerged based on target and distractor location predictability: when their location was unpredictable, trajectories deviated toward distractors, whereas, with predictable locations, saccades deviated away from distractors (Walker et al., 2006).

Besides the role of spatial position, some studies have also shown that the temporal distance between distractor presentation and saccade onset influences its trajectory. When target and distractors were presented simultaneously, shorter-latency saccades (less than ~200 ms) deviated toward distractors, whereas longer-latency saccades deviated away from distractors (McSorley et al., 2006; Theeuwes & Godijn, 2004; Van Zoest et al., 2004; Walker et al., 2006; Walker & McSorley, 2008).

Jonikatis and Belopolsky (2014) induced oculomotor competition by briefly presenting a task-irrelevant distractor (50 ms) at different times during the peri-saccadic epoch (from -400 to +600 ms from saccade onset). They found that the distractor offset time relative to saccade onset (DSOA) influenced the amplitude of the curvature; the deviation *away* was maximal when the distractor-to-saccade onset asynchrony was long and decreased as DSOA became shorter (Jonikaitis & Belopolsky, 2014).

Distractor stimuli can also influence saccades' endpoint position. Indeed, when a distractor is presented in close spatial proximity to a target, saccades tend to land in between the two objects rather than on the target – the so-called "Global effect" (Coren and Hoenig, 1972; Findlay, 1982; Wollenberg et al., 2018).

Finally, the effects of a distractor are also reflected in the temporal properties of saccades. Indeed, distractors positioned at a remote location from the target evoke longer saccade latencies as compared to distractors close to a target (i.e., "Remote Distractor Effect," RDE; Godijn & Theeuwes, 2002; Walker et al., 1997, 1995). Further studies also found that a distractor presented before the target reduces the saccadic latency, contrary to a distractor presented after the target which delays saccades latency (Ross & Ross, 1980), differently depending on whether it is in the same hemifield as the target and near or far from the fovea (Casteau & Vitu-Thibault, 2012; Walker et al., 1997).

It has been speculated that all these effects may be attributed to inhibitory processes occurring in the oculomotor system in situations where observers have to make a fast and accurate eye movement to a target while ignoring a competing distractor. Hence, the directional deviation of the saccade trajectory away or toward the distractor location would reflect the outcome of this competition. If the distractor location is only weakly inhibited the saccade trajectory will deviate toward the distractor before heading to the target (Heeman et al., 2016; Van der Stigchel, 2010; Van der Stigchel et al., 2006, 2007b), whereas strong inhibition will cause deviation away from the distractor (Doyle & Walker, 2001; Tipper et al., 2001, 1997; Van der Stigchel et al., 2007b, 2006).

The fundamental point for our current study is that a larger saccade deviation either away or toward a distractor implies a stronger influence of the distractor (e.g., as obtained with larger, or brighter distractors – Deuble et al., 1984; Findlay, 1982). In other words, more salient distractors yield more pronounced competition, that in turn leads to stronger attraction and requires greater inhibition, inducing overall greater saccadic curvature. This interpretation has been broadly used to explain a variety of findings. For example, it has been speculated that distractors that share visual similarities with the saccade target produce greater trajectory deviation than dissimilar distractors because they are more behaviorally salient for the visuomotor system engaged in that particular saccadic task (Ludwig & Gilchrist, 2003). A similar speculation based on a broad definition of saliency has been proposed for distractors closer to fixation *vs.* distractors far from fixation (Van der Stigchel et al., 2007a), for bimodal distractors *vs.* unimodal ones (Heeman et al., 2016), and for abrupt onset *vs.* color singleton distractors (Godijn & Theeuwes, 2004).

Recent studies specifically tested distractor saliency effects on saccade curvature. Jonikaitis and Belopolsky (2014) used a double saccade task and manipulated the salience of the distractor presented before the first saccade by adjusting its luminance. They observed that the degree of curvature of the second saccade away from the distractor increased when the distractor's luminance increased, suggesting that information about the distractor's salience was also transferred across saccades (Jonikaitis & Belopolsky, 2014). This saliency effect was later reproduced across different sensory modalities (Szinte et al., 2020). Van Zoest et al. (2012) modified the distractor salience in terms of orientation contrast relative to the surrounding stimuli. Their results revealed that saccades deviated toward the irrelevant distractor and this deviation was stronger for more salient distractors, with a stronger effect on the shortest saccade latencies (Van Zoest et al., 2012). Finally, Tudge and colleagues (2018) exploited the assumption of the relation between distractor saliency and saccade deviation to estimate the saliency of stimuli defined by combinations of different features with respect to single-feature stimuli (Tudge et al., 2018). Many studies also found that distractors with task-relevant features produce deviations in saccade trajectories, showing that the visuo-oculomotor system is not just sensitive to low-level saliency, but also to high-level manipulations of distractors' saliency (Kehoe et al., 2021; Kehoe, Aybulut, et al., 2018; Kehoe, Rahimi, et al., 2018; Van Der Stigchel et al., 2011; Van der Stigchel et al., 2009). As an example, Hickey and van Zoest (2012) found that saccade trajectories are influenced by reward-associated distractors, demonstrating top-down, task-dependent influences on saccadic curvature (Hickey & Van Zoest, 2012).

Building-up on these findings, in the present study (Castellotti et al., 2023a), we compare the effects on saccades trajectories produced by *optimal* vs. *non-optimal* features used as distractors, considering the magnitude of curvature as a measure of feature saliency. We expected that, if *optimal* features are indeed more salient, their presence will interfere in a quick and automatic manner with the ongoing oculomotor programming, and they will induce a larger saccadic curvature. As a control for this saliency effect, we compared it with the saccadic curvature induced by high-luminous vs. low-luminous distractors. In order to characterize the time course of the saliency effect, we also investigated changes in saccade trajectory deviations at different DSOA and we looked at the possible effects of distractors' saliency on endpoint position and saccade latency.

4.2 Aim of the study

In Study 3 (Castellotti et al., 2023a), we further tested the saliency predictions of the constrained maximum-entropy model by using *optimal* and *non-optimal* features as distracting stimuli in a saccadic task and measuring the resulting saccadic curvature. This approach allowed us to study the low-level, automatic integration of these *optimal* features in fast visuo-oculomotor processes, along with its dynamics.

4.3 Materials and methods

4.3.1 Participants

Twenty-three healthy volunteers participated in the present study (aged 22–34 years, M = 26.48, SD =3.07, fourteen females and nine males).

4.3.2 Apparatus and set-up

Each participant was tested individually in a dark room. Stimuli presentation was controlled by a MacPro computer (OS 10.6.8), using the Psychophysics Toolbox (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) and the Eyelink Toolbox extensions (Cornelissen et al., 2002) for Matlab. The experiment was displayed on a 120 Hz CRS Display++ LCD monitor (1920 by 1080 pixel), subtending 70 by 40 degrees at a viewing distance of 57 cm. Participants' viewing was binocular, but only the right eye was recorded by an Eyelink 1000 video-based eye tracker (1 kHz). The observers' head was stabilized with a chin- and forehead rest.

4.3.3 Stimuli and conditions

Trials were distributed across three different conditions: baseline (200 trials), experimental (400 trials), and luminance-control (400 trials) condition. Trials' order was randomized. In the baseline condition, no distractor was presented. The experimental and luminance-control conditions differ for the visual distractor stimuli used in the task: in each condition, the distractor could be high salient (200 trials) or low salient (200 trials). In the experimental condition, distractors' saliency differs based on the spatial arrangement of their black (greyscale value = 50; luminance 4 cd/m^2) and white (greyscale value = 250; 44 cd/m^2) pixels, whereas, in the control condition, distractor saliency was based on their luminance contrast with respect to the background.

Specifically, in the experimental condition, two types of distractors were used: *optimal* features and *non-optimal* features, as identified by the reference constrained maximum-entropy model (Del Viva et al., 2013). At each trial, the *optimal* feature, assumed to be a "high salient distractor," was randomly extracted from a set of 50 features selected as the best information carriers (Figure 8C; average greyscale value = 151; Weber contrast = 0.19; RMS contrast: M = 1.2, SD = 0.1, range [0.98 1.3]); whereas the *non-optimal* feature, used as a “low salient distractor,” was extracted from the set of 50 features selected amongst those with the lowest probability of occurrence in the statistical distribution of all possible features (Figure 8E; average greyscale value = 149, Weber contrast = 0.18; RMS contrast: M = 1.2, SD = 0.05, range [1.0 1.3]), and thus classified as poorly informative by the model. On average, then optimal and non-optimal features do not differ in their contrast properties and spatial frequency content (see **Supplementary Material**). The theoretical range of the unidimensional spatial frequency for our 9 x 9 pixels features is comprised between 1.67 and 5 cycles/deg for both sets.

In the luminance-control condition, the distractor stimulus could be a grey uniform square with low luminance contrast with respect to the background (20 cd/m^2; greyscale value = 150; Weber/RMS contrast = 0.18; SD = 0), or a white uniform square with high luminance contrast (23 cd/m^2; greyscale value = 173; Weber/RMS contrast = 0.36, SD = 0). The luminance values of the control distractor stimuli were purposely chosen to compare their effect with the effect induced by the distractor used in the experimental condition. Indeed, these values produce a contrast corresponding to the “equivalent contrast” found to match the saliency difference between *optimal* and *non-optimal* features (Castellotti et al., 2021).

4.3.4 Procedure

At the beginning of the experimental session, a standard nine-point gaze-calibration routine was run and possibly repeated to ensure good quality of the eye movement recordings. The whole task (and the calibration) was presented on a grey display background (16 cd/m^2). See **Figure 31A** for an illustration of a trial. Each trial started with the presentation of a white central fixation point (44 cd/m^2) for a variable duration, uniformly distributed in the range between 400–800 ms, followed by the target (white ring, 18 pixels large, 0.6 °) shown for 700 ms at 7° eccentricity on the vertical meridian, above or below the initial fixation. A small distractor stimulus (arrays of 9 × 9 pixels, 0.3 °, see Stimuli section for details) could randomly appear for 25 ms on the right or on the left, halfway from the target (3.5° vertically and horizontally from fixation), with a variable delay with respect to target onset (Distractor-to-Target Onset Asynchrony - DTOA from -150 to +50 ms). See **Figure 31B** for an illustration of the display sequence. Observers were instructed to make a fast and accurate saccade toward the target. The following trial started only after participants resumed fixation. Each participant performed 1000 trials in total, in a single session with a small pause every 100 trials.

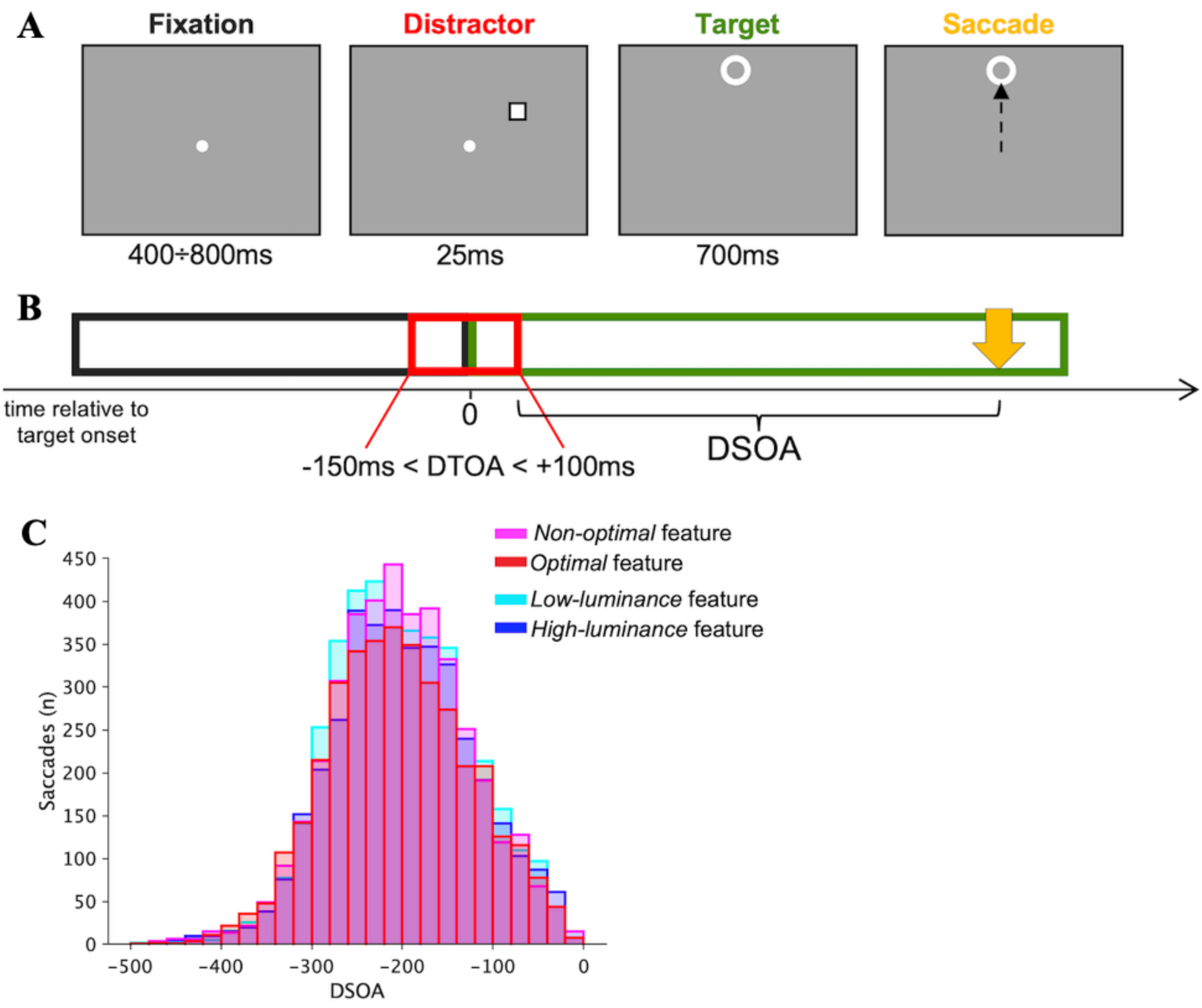

Figure 31. Study 3 – Procedure, stimuli, and DSOA distribution. (A) Example of a trial. Durations of the different stimuli are reported under each panel. In this example, the distractor is presented as an empty square (9x9 pixels) and had a negative DTOA (presented before the target), and the target is presented 7° above the fixation. **(B) Display sequence.** Fixation point disappears as soon as the target appears. The time 0 refers to the target onset. Relative to target onset, the distractor could appear from 150 ms before the target to 100 ms after the target (DTOA = Distractor to Target onset asynchrony). The time distance between the distractor offset and the saccade onset (yellow arrow) is the distractor offset-to-saccade onset asynchrony (DSOA). **(C) DSOA distribution for each distractor condition.** Data are binned every 20ms. Figure retrieved from (Castellotti et al., 2023a).

Task characteristics were chosen to maximize the possibility of observing curvature effects. First, we instructed only vertical saccadic eye movements because curvature effects are more pronounced for vertical than for horizontal saccades (Jonikaitis & Belopolsky, 2014; Walker & McSorley, 2008). Second, the target could only appear at two possible locations, because it has been found that predictable target locations make inhibitory mechanisms more pronounced in the target selection process (Van der Stigchel et al., 2006; Walker et al., 2006). Also, utilizing different DTOA produce saccades starting at different times to the distractor (distractor offset-to-saccade onset asynchrony - DSOA) and allows us to analyze the temporal dynamics of the distractor's effects on the saccade programming.

4.3.5 Data processing and statistical analysis

Data about saccades were stored and analyzed offline with MATLAB routines (The MathWorks, Inc.). Recorded horizontal and vertical gaze positions were low-pass filtered with a Butterworth (acausal) filter of order 2 with a 30 Hz cutoff frequency and then numerically differentiated to obtain velocity measurements. An automatic conjoint acceleration and velocity threshold method was used to detect saccades (Krauzlis & Miles, 1996).

For each trial, the first saccade after target onset was analyzed. Trials without recorded saccades (e.g., due to a long blink), saccades starting further than 2° from the fixation point, saccades landing further than 2° from the target, and saccades with latency lower than 80 ms or longer than 500 ms (with respect to target onset), were excluded (~15% of all saccades).

For each correct saccade, we extracted its latency, starting point, landing position, and curvature. To obtain the latter, we first smoothened the gaze position trajectory by computing the mean horizontal position for each 0.2° spatial bin on the vertical axis. By doing this, a unique pair ($x_{i,}$ y_i) was obtained for each bin between the y-position of the fixation and of the target location, removing a possible ambiguity in the definition of the saccadic curvature. Then, we normalized each trace by subtracting the mean trajectory (horizontal and vertical eye position) obtained in the trials without distractors (baseline condition), separately for upward and downward saccades. This normalization ensured that any deviation of the saccade trajectory in response to a distractor was not due to the idiosyncratic curvature of the saccade trajectory of individual participants. Normalized y-coordinates were rotated to direct all saccades upward and, finally, normalized x-coordinates were inverted for trials in which distractors were presented counterclockwise (i.e., to the left) relatively to the target direction. This way, positive and negative values represent coordinates and curvature angles that were directed either toward or away from the distractor's head-centered position, respectively. With these coordinates we then determined the saccade curvature angle for each trial, that is, the median of the angular deviations of each sample gaze-position point from a straight line connecting the starting and ending point of the saccade. The saccade endpoint deviation was defined as the angular distance between the saccade endpoint and the target in degrees.

For each saccade, we calculated the DSOA, which is the distractor offset-to-saccade onset asynchrony in milliseconds.

The time course of saccadic curvature, latency, landing point, and starting position as a function of the DSOA has been analyzed with the SMART procedure, a smoothing method for the analysis of response time courses (van Leeuwen et al., 2019). Data were analyzed at a 1-ms resolution and smoothed with a Gaussian kernel of 10 ms width. 1000 permutations were run for every test. We set the significance level as $p < 0.01$. Time courses for each distractor condition have been compared with each other, and each condition has been compared with the baseline condition. Data have been compared in the DSOA window between -340 ms and -20 ms. This interval has been chosen by looking at the distribution of DSOA across all trials (**Figure 31C**) and selecting the 20 ms bins which included at least 40 saccades for each distractor condition. For each comparison, we report the time window of the significant cluster, the average of smoothed Gaussian data and 95% confidence interval within each significant cluster.

To produce figures of average saccade trajectories for each distractor condition (Figure 32C and 32D), we averaged the normalized trajectories of all participants and estimated their standard error of the mean (SEM). Since each participant's trajectory may have different starting/ending point (therefore different lengths in the y-coordinates), only y-coordinates including at least fifteen participants (i.e., fifteen x-coordinates) were considered. This process led to final average trajectories having the central part populated by all participants, and the outermost parts containing only some participants (with a minimum of 15 participants).

4.4 Results

Here, our interest is measuring the effect of a distractor on the saccade trajectory assuming that its representation will be further inhibited with a longer delay between the distractor and the saccade. For this reason, we considered the temporal distance between the distractor presentation and the saccade onset (DSOA) as the relevant variable for the direction of the saccade's deviation (Figure 31B). Previous works on the effect of a distractor on saccadic trajectory have considered the saccadic latency as the independent variable because the target and the distractor were usually presented simultaneously, so latency and DSOA were the same (Theeuwes and Godijn, 2004; Van der Stigchel, 2010; Van Zoest et al., 2012; Walker et al., 2006; Walker & McSorley, 2008). Saccade distribution for each distractor type across the DSOA time course (bins of 20ms) is reported in Figure 31C. In the following analyses, we considered the DSOA values between -340ms and -20ms, the interval in which there are a reasonable number of saccades to reconstruct the time course (20-ms bins including at least 40 saccades for each distractor condition).

First, we analyzed saccadic curvature as a function of the DSOA with the SMART procedure (**Figure 32**), a smoothing method designed for the analysis of response time courses that retains high temporal resolution with no need to define time bins (van Leeuwen et al., 2019). All saccades' trajectories shown in the following graphs are flipped so that the target is always in the upper field and the distractor is always on the right; namely, negative curvature indicates deviation away from the distractor ("repulsion"), whereas positive curvature indicates deviation toward the distractor ("attraction").

Saccades time courses for *non-optimal* vs. *optimal* features, and for *low-luminance* vs *high-luminance* features, are reported in **Figure 32A** and **32B**, respectively. Common across all conditions is that, at large negative DSOA, the presentation of all distractors induced a negative deviation, significantly different from the natural deviation measured in the condition without a distractor (baseline condition). Around DSOA of -200 ms the deviation away decreases, until becoming significantly positive after -150 ms. After reaching the peak around -120 ms, the curvature begins to decrease. After -100 ms, the curvature becomes compatible with zero for the low-salient distractors (*non-optimal* and *low-luminance* features) and slightly negative for the high-salient distractors (*optimal* and *high-luminance* features). All significant DSOA time windows between all types of distractors conditions vs. the no-distractor baseline condition are reported in the caption of Figure 32.

Time courses for each distractor condition have been compared with each other. For each comparison, we report the time window of the significant cluster (ms), the average of smoothed Gaussian data and 95% confidence interval within each significant cluster ($p < 0.01$).

Comparing the curvature induced by the two types of distractors of the experimental condition (**Figure 32A**), we observed that, at large negative DSOA ([-267, -204] ms), *optimal* features induce a larger deviation *away* from their location than that induced by *non-optimal* features (-1.55° vs. -0.89° ± 0.2°). At intermediate DSOA ([-180, -161] ms*)* *optimal* features induce less deviation away than *non-optimal* (-0.39° vs. -0.97° ± 0.2°) and, finally, at short DSOA ([-111, -102] ms*)* they produce a larger deviation *toward* their position than that induced by *non-optimal* features (2.05° vs. 1.54° ± 0.2°).

Similar results emerged in the luminance-control condition (**Figure 32B**). Indeed, at long DSOA ([-283, -265] ms and [-248, -217] ms), *high-luminance* distractors induce a stronger repulsion than that induced by *low-luminance* distractors (-1.38° vs. -0.8° ± 0.1 and -1.19° vs. -0.49° ± 0.2°), whereas, at shorter DSOA ([-154, -125] ms), *high-luminance* distractors induce a stronger attraction than *low-luminance* distractors (1.10° vs. 0.48° ± 0.2°).

Overall, these results indicate that high-salient features, such as very bright features as well as optimally informative features, induce a larger curvature in the saccade trajectory than that induced by lower salient distractors (less bright or less optimal), and this holds both when the curvature is toward and away from the distractor.

Some differences also emerged by comparing the curvature induced by distractors of the experimental vs. control conditions (not reported in the figure). Indeed, at long DSOA, the deviation *away* is larger for *non-optimal* vs. *low-luminance* features ([-324, -303] ms: -1.4° vs. -0.33° ± 0.3°), as well as for *optimal* vs. *high-luminance* features ([-263, -245] ms: -1.8° vs. -1.1° ± 0.23°). Also, at short DSOA, the deviation *toward* is larger for *non-optimal* vs. *low-luminance* features ([-138, -121] ms: -2° vs. 2° ± 0.3°), as for *optimal* vs. *high-luminance* features ([-127, -118] ms: -2.3° vs. 1.8° ± 0.3°). Therefore, both repulsion and attraction effects are more pronounced in the presence of experimental distractors than uniform-luminance distractors.

Examples of average saccades trajectories in the trials with *non-optimal* vs. *optimal* features, and *low-luminance* vs. *high-luminance* features, within two significant time windows found with the SMART procedure, are reported in **Figure 32C** and **32D**, respectively. Data shown in panels C and D are saccades trajectory averaged across participants with shaded areas representing the standard error of the mean (SE).

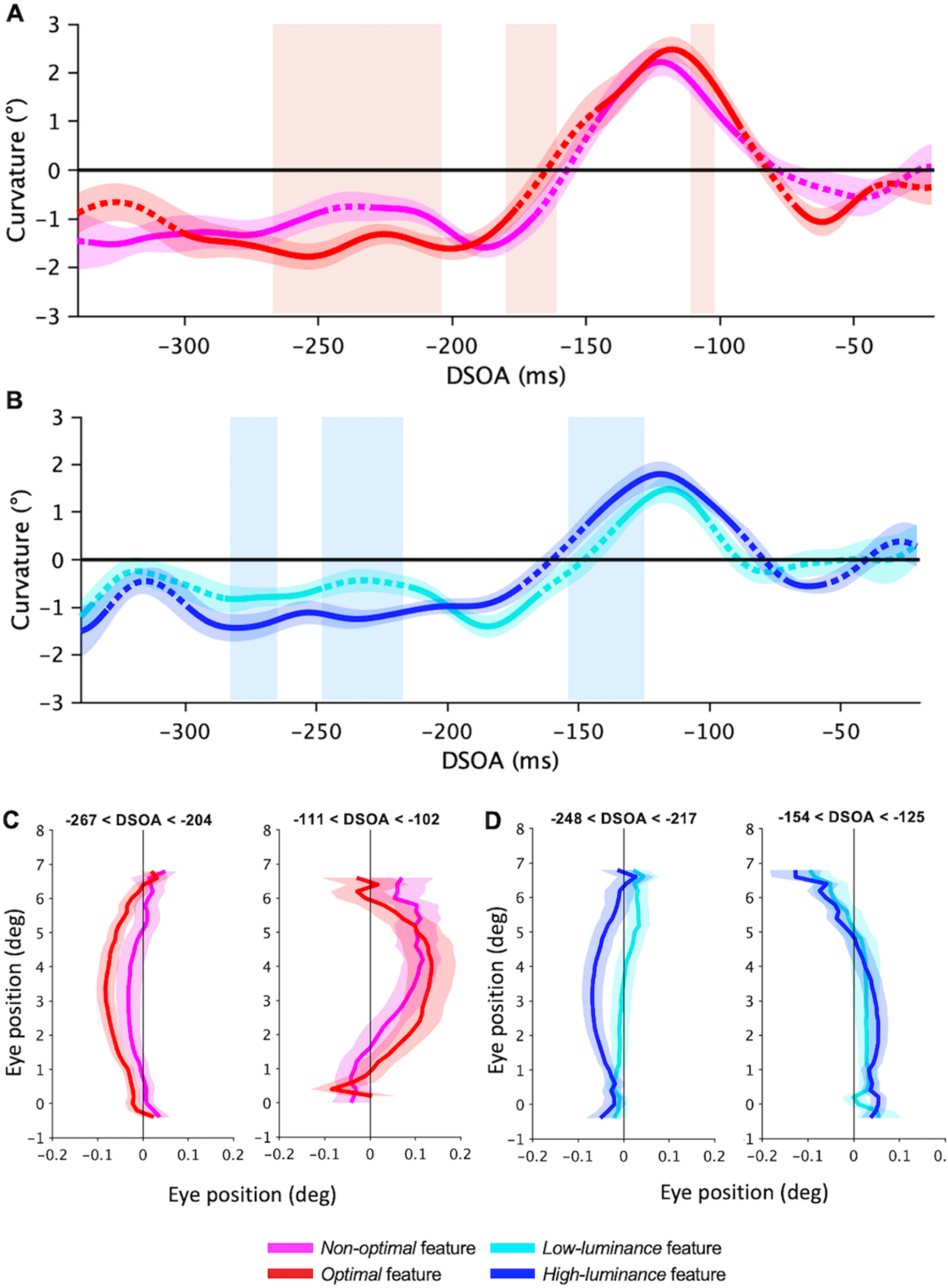
A
Curvature (°)
3
2
1
0
−1
−2
−3
−300
−250
−200
−150
−100
−50
DSOA (ms)
B
Curvature (°)
DSOA (ms)
C
-267 < DSOA < -204
-111 < DSOA < -102
D
-248 < DSOA < -217
-154 < DSOA < -125
Eye position (deg)
−0.2
−0.1
0
0.1
0.2
Eye position (deg)
Non-optimal feature
Optimal feature
Low-luminance feature
High-luminance feature

Figure 32. Study 3 – Saccadic curvature. (A-B) Curvature time course as a function of DSOA (SMART analysis) for *optimal* (red) vs. *non-optimal* (pink) features (A), and for high-luminance (blue) vs. low-luminance (cyan) features (B). Shaded areas around the curves: 95% confidence interval (van Leeuwen et al., 2019). Colorful shaded rectangles: time intervals with significantly different curvature induced by high-salient vs. low-salient distractors ($p < 0.01$; light-red: *optimal* vs. *non-optimal* features; light-blue: high vs. low-luminance features). Solid parts of lines: time windows (ms) with curvature significantly different from zero ($p < 0.01$; baseline vs. *non-optimal* features: [-336, -247] ms, [-229, -169] ms, [-146, -93] ms; baseline vs. *optimal* features: [-302, -176] ms, [-145, -92] ms, [-72, -43] ms; baseline vs. *low-luminance* features: [-340, -336] ms, [-285, -248] ms, [-214, -166] ms, [-134, -102] ms; baseline vs. *high-luminance* features: [-340, -33] ms, [-298, -174] ms, [-147, -90] ms, [-70, -57] ms). Dashed parts of lines: curvature compatible with zero. **(C) Experimental condition - Average trajectories.** Left panel: saccades with DSOA between -267 ms and -204 ms showing deviation away from the distractor; right panel: saccades with DSOA between -111 ms and -102 ms showing deviation toward the distractor. **(D) Luminance-control condition - Average trajectories.** Left panel: saccades with DSOA between -248 ms and -217 ms showing deviation away from the distractor; right panel: saccades with DSOA between -154 ms and -125 ms showing deviation toward the distractor. Errors are SE across participants. Figure retrieved from (Castellotti et al., 2023a).

The latency of saccades was also analyzed as a function of DSOA (from -340 ms to -20 ms) with the SMART procedure (**Figure 33A**). The average latency of saccades in the trials with no distractor is 184.5 ± 2.8 ms (SEM across participants) (horizontal black line in Figure 33A). Common across all trials with a distractor is that the saccadic latency is longer at the longest DSOA (< -250 ms), and shorter at the shortest DSOA (> -250 ms) compared to the baseline latency. All significant DSOA time windows between all types of distractors conditions vs. the no-distractor baseline condition are reported in the caption of Figure 33A.

No significant differences were detected when comparing the latency time courses obtained with high-salient vs. low-salient distractors in both experimental and control conditions. Instead, some differences emerged between the two conditions. Indeed, there are some time windows (ms) where the latency (ms) for low-luminance distractors is shorter than for *non-optimal* features ([-313, -280] ms: 197 vs. 205 ± 2 ms; [-238, -214] ms: 181 vs. 188 ± 2 ms) and *optimal* features ([-308, -284] ms: 196 vs. 203 ± 2 ms; [-98, -91] ms: 167 vs. 171 ± 1 ms). In the same way, with saccade preparation for high-luminance distractors is more delayed than for *non-optimal* features ([-300, -294] ms: 197 vs. 205 ± 2 ms; [-236, -210] ms: 181 vs. 188 ± 1 ms) and *optimal* features ([-306, -292] ms: 199 vs. 205 ± 1 ms; [-223, -216] ms: 181 vs. 186 ± 1 ms). This result suggests that saccadic latency is not sensitive to the degree of saliency (either when considering differences in luminance or in amount of information), rather it seems to be influenced by other distractors characteristics. See the Discussion section for possible explanations of this result.

Finally, the landing point of saccades was also analyzed as a function of DSOA (from -340 ms to -20 ms) with the SMART procedure (**Figure 33B and 33C**). Saccades in trials with no distractor, and a target positioned at 7° upward, on the vertical meridian, land on average at 0.2° ± 0.1° on the x-axis (black horizontal solid line in **Figure 33B**), and at 6.8° ± 0.13° on the y-axis (black horizontal solid line in **Figure 33C**). Thus, when performing a saccade toward the target, the baseline landing point is in a position slightly lower right than the actual target position (dashed horizontal black lines in Figure 33B and 33C). Landing x- position of saccades for any type of distractor is compatible with the baseline position until around -200 ms. Then, we observed a significant shift in the opposite direction of the distractors (negative landing X-point) for about 100 ms, and finally a shift toward the distractor (positive landing X-point) between -100 ms and -50 ms (Figure 33B). Landing y-position for all types of distractors is compatible with the baseline position across almost all DSOA times, with a small tendency of shifting downward (below the target) at the shortest DSOA (significant only for the high-luminance distractor in a very small time-window; Figure 33C). All significant DSOA time windows between all types of distractors conditions vs. the no-distractor baseline condition are reported in the caption of Figure 33.

Comparing the landing point time courses obtained with high-salient vs. low-salient distractors in both experimental and control conditions, we could observe that the landing position does not change with distractor saliency, neither in the horizontal (Figure 33B) nor in the vertical dimension (Figure 33C). Also, no differences emerged between experimental and control distractors.

We finally analyzed starting X- and Y-positions of saccades (not shown in the figure), finding that they do not differ across distractor conditions nor with DSOA. They are always compatible with the average baseline starting positions (x = -0.01° ± 0.1°, y = 0.1° ± 0.02°) found in the no-distractor condition.

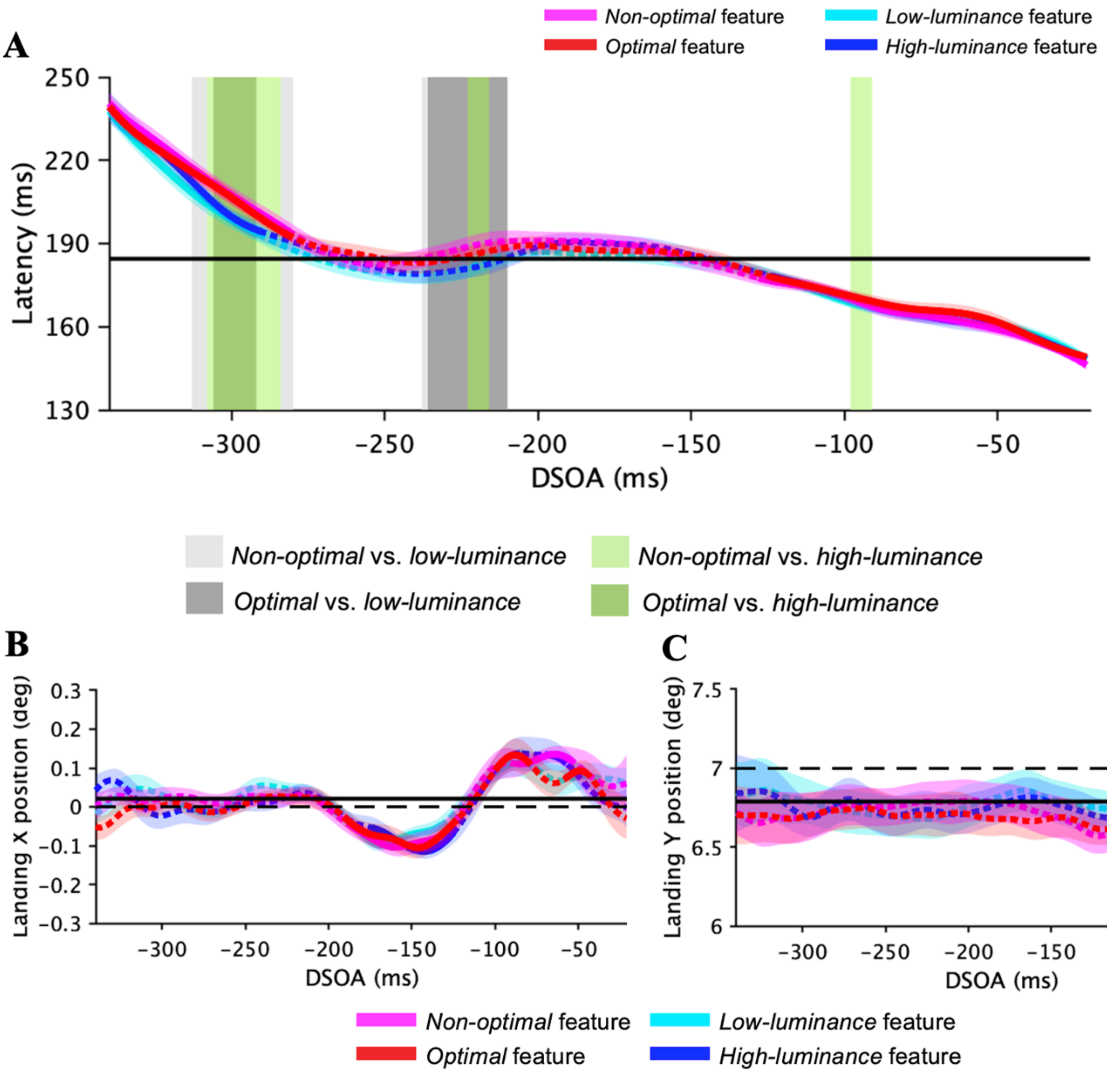

Figure 33. Study 3 – Latency and landing point. (A) Latency. Latency time course as a function of DSOA (SMART analysis) for *optimal* features (red), *non-optimal* features (pink), high-luminance features (blue), and low-luminance features (cyan). Shaded areas around the curves: 95% confidence intervals (van Leeuwen et al., 2019). Solid parts of lines: time windows (ms) with latency significantly different from baseline (p < 0.01; baseline vs. *non-optimal* features: [-340, -281] ms, [-128, -21] ms; baseline vs. *optimal* features: [-340, -280] ms, [-124, -21] ms; baseline vs. *low-luminance* features: [-340, -291] ms, [-122, -21] ms; baseline vs. *high-luminance* features: [-340, -290] ms, [-124, -21] ms). Dashed parts of lines: latency compatible with baseline. Colorful shaded rectangles: time intervals with significantly different latency with experimental vs. luminance-control distractors (color legend in the figure). **(B-C) Landing point.** Landing X- (A) and Y-position (B) time course as a function of DSOA (SMART analysis) for *optimal* features (red), *non-optimal* features (pink), high-luminance features (blue), and low-luminance features (cyan). Shaded areas: 95% confidence interval (van Leeuwen et al., 2019). Solid parts of lines: time windows (ms) with landing X- and Y-positions significantly different from baseline (p < 0.01). (A) baseline vs. *non-optimal* features: [-87, -132] ms, [-86, -54] ms; baseline vs. *optimal* features: [-191, -123] ms, [-100, -79] ms, [-58, -45] ms; baseline vs. *low-luminance* features: [-187, -132] ms, [-101, -89] ms; baseline

vs. *high-luminance* features: [-174, -125] ms, [-93, -85] ms, [-70, -57] ms. (B) baseline vs. *high-luminance* features: [-91, -72] ms. Dashed parts of lines: landing X- and Y-positions compatible with baseline. Figure retrieved from (Castellotti et al., 2023a).

4.5 Discussion

In the present study, we compared the saliency of some local features, originally identified as *optimal* or *non-optimal* information carriers based on constrained-entropy maximization criteria (Del Viva et al., 2013). To do this, we presented these features as distractors in a simple saccadic task and used the saccadic curvature as a measure of their relative saliency. Our main objective was to investigate whether the saliency provided by features' optimality, which is already known to induce an automatic attentional attraction (Castellotti et al., 2021, 2022), also occurs so rapidly that it can influence the trajectory of a planned saccade. For our main goal, the saccadic curvature has been merely used as a mean to quantify the model-predicted visual saliency. However, this study may also serve to better characterize the temporal factors modulating the saccadic curvature and shed some light on the saccadic programming dynamics in presence of distracting stimuli.

Our main result is that the saccadic curvature induced by *optimal* features is larger than that produced by *non-optimal* ones, suggesting that *optimal* features act as highly salient distractors which strongly compete with the target location in the oculomotor centers. The results obtained in the luminance-control condition, in which we compared the deviation induced by distractors with high or low luminance contrast, also support our main hypothesis. Indeed, high-luminance features induce a larger deviation than low-luminance features, and the amplitude of the deviation is comparable to that found with *optimal* features. This suggests that the interference on saccade programming produced by differences in features' optimality is as powerful as that produced by luminance-based saliency. Note that all types of distractors, also *non-optimal* and *low-luminance* features, evoke some degree of curvature compared to the condition where no distractor is presented. This confirms the general finding that the path of target-oriented saccades is influenced by visual competing distractors even if they are task-irrelevant and low visually salient (for reviews see Walker & McSorley, 2008; Van der Stigchel, 2010).

Interesting results also emerged by the inspection of the time courses of saccade trajectory deviations. Indeed, we found that the saccadic curvature direction changes as a function of the temporal distance between the saccade onset and the distractor offset. In both experimental and luminance-control conditions, saccades starting at least 200/150 ms after the distractor tend to deviate away from it, whereas when the temporal interval between distractor offset and saccade onset is small saccades curve toward the distractor location. When comparing the saliency effect within the experimental and within the control condition, one can observe a small relative shift of the time intervals with a significantly different curvature for the more salient stimuli as compared to the less salient ones. For instance, optimal features induce a significantly larger curvature toward the distractor, compared to non-optimal features already for DSOAs around -100ms, whereas the difference between high-luminance and low-luminance distractors becomes significant only if the distractor is presented more than 125ms before saccade onset. Yet, the qualitative agreement of the time course of saliency effect across experimental and control conditions is apparent. Further studies will be needed to precisely investigate possible quantitative differences between optimal feature-based and luminance-based saliency in more detail.

When considering the relationship between curvature and saccadic latency, our results confirm previous findings, showing a trend for long latency saccades to have larger deviations away from distractors than fast saccades (McSorley et al., 2006; Theeuwes & Godijn, 2004; Walker & McSorley, 2008).

Results also show that our experimental distractors, both *optimal* and *non-optimal*, evoke larger deviations than those induced by our control uniform-luminance distractors. Similarly, saccadic latency increases more with optimal and non-optimal features than with uniform-luminance distractors. There are some possible explanations for these results, which would deserve further investigation. First, our experimental stimuli have an internal structure, not present in the control stimuli. Thus we could speculate that these results are related to recent findings about saccade-contingent neuronal activity in the superior colliculus being tuned to the spatial frequencies of a visual stimulus (Buonocore & Hafed, 2021). In addition, our findings could also be explained by experimental features' internal contrast. Indeed, optimal and non-optimal features, although having the same average luminance contrast with respect to the background as the control features (Weber contrast), they have a higher internal luminance contrast compared to them (RMS contrast). Finally, since optimal and non-optimal distractors are more complex than the simple uniform-luminance distractors, the perceptual load needed for their analysis could be higher than that required by the control stimuli (Lavie et al., 2014), thus influencing saccadic curvature and latency.

More in general, in our study, we found an increase of saccadic latency with all distractors at long DSOA. This is in line with previous studies, showing that the presentation of a competing stimulus along with the target delays saccade onset (Remote distractor effects; Walker et al., 2006). However, we did not find an increase nor a decrease of saccadic latency in the presence of a high salient distractor compared to low salient distractors. The absence of any interference in the temporal domain may be the result of the low uncertainty about target locations: the target was always presented directly above or below fixation and so participants were potentially able to prepare an eye movement in advance to these two relevant target locations (Heeman et al., 2016).

We did not find any differences in saccades' endpoint positions across different distractor types and saliency levels. This result could be somehow unexpected, since prior research found distractor-saliency effects on endpoint deviations (e.g., Kehoe et al., 2021; Van Der Stigchel et al., 2011). The discrepancy between our results and the previous ones may be related to the fact that shifts in saccades ending points toward the distractor are usually found in specific conditions of spatial proximity between target and distractor (Global effect; Coren and Hoenig, 1972; Findlay, 1982; Wollenberg et al., 2018), and high uncertainty about their position (Coëffé & O'regan, 1987; Walker et al., 2006). In our paradigm, these factors were not optimized for observing the Global effect, and this might explain why the saccadic landing point does not change with the distractor condition. When analyzing the landing point time course as a function of DSOA, we found that the landing point of saccades starting way after the distractor is compatible with the natural landing point of saccades when no distractor is presented. Instead, saccades starting shortly after the distractor offset landed in a position shifted from the baseline. We can speculate that the effect on the saccadic landing point could be simply a sort of automatic "over correction" after the initial deviation caused by the distractor.

On a more general ground, distractor effects have been explained in terms of the programming of an oculomotor vector toward the target and distractor locations. The oculomotor system has to resolve the competition between these two vectors in order to determine the goal of the next eye movement (Meeter et al., 2010; Trappenberg et al., 2001). In this process, the vector programmed toward the distractor needs to be inhibited in order to avoid making an eye movement towards it rather than toward the instructed target (Rizzolatti et al., 1987; Sheliga et al., 1995). The intermediate layers of the superior colliculus (SC) are often implicated in oculomotor competition (Calvert, 2001; Chalupa & Rhoades, 1977; Finlay et al., 1978; Jay & Sparks, 1987, 1984) because they contain a large population of neurons with large overlapping *motor fields* that encode the saccadic displacement (Lee et al., 1988; Mcilwain, 1991) which can be regarded as forming a "motor map". When the population of neurons encoding the target overlaps with a second population encoding the distractor, an error in the computation of the initial saccade direction may occur. The initial saccade deviation, either towards or away from a competing distractor, is thought to reflect the level of neural activity at the distractor site around the time of saccade onset (McPeek et al., 2003). Therefore, when the competition is unresolved and there is strong activity at the distractor site, the saccade will deviate towards its location (Walker et al., 2006), coherently with our results obtained when saccades started too close in time to the distractor. In contrast, later in time, when the inhibition of distractor-related activity is achieved, saccades can even deviate away from it (Godijn & Theeuwes, 2004), as we found for saccades starting way after the distractor offset. It has been assumed that the stronger the inhibition required by the distractor, the larger the deviation away (Rizzolatti et al., 1987; Sheliga et al., 1995). This would explain why our high salient distractors induce a larger deviation away.

Whereas spatiotemporal factors that relate local excitation at the distractor site in the SC to saccades curved toward distractors are well understood (McPeek, 2006; McPeek et al., 2003), the mechanisms responsible for saccades curved away from distractors are more controversial (Aizawa & Wurtz, 1998; White et al., 2012; see also Wang et al., 2012; Wang & Theeuwes, 2014 for related models).

Recently, based on behavioral measurements, Kehoe and Fallah (2017) proposed a model of activation and inhibition explaining both towards and away curvature. In particular, they modeled DSOA functions separately for saccades curved toward and away from distractors and suggested that a similar temporal process determined the magnitude of saccade curvatures in both contexts. Their behavioral and theoretical results are in line with our findings (Kehoe & Fallah, 2017).

To summarize, our work extends and generalizes the notion that the saccadic curvature towards or away from a distractor increases with the saliency of the distractor presented in competition with the saccade target. In fact, we show that the relative saliency of *optimal* vs. *non-optimal* features predicted by the constrained maximum-entropy model is reflected in the magnitude of the saccadic curvature. Our findings also indicate that the saliency based on information maximization has a comparable effect to luminance-based saliency, as well as a similar time course.

We conclude that the automatic capture exerted by optimally informative visual features results from early processes in the visual stream, such that it can interfere with the correct programming of fast, visually guided saccades. Finally, our study confirms that saccadic curvature can be used as an objective measure to quantify the relative saliency of a visual stimulus.

All the studies presented so far involved non-ecological paradigms presenting single *optimal* features as stimuli to investigate their visual saliency. Our next objective was to design an experiment that tested the role of these features within more naturalistic stimuli and using an ecological task. We were also interested in assessing the strength of automatic attention capturing of these local features compared to global visual elements. To these purposes, in the last study describe below, we considered a condition in which observers have to quickly explore an image based on a few fragments in order to discriminate it later, and we compared the relative contribution to image reconstruction of global and local information, given by *optimal* features, contained in those fragments.

Chapter 5

Study 4: Fast discrimination of fragmentary images: the role of local optimal information

5. STUDY 4: FAST DISCRIMINATION OF FRAGMENTARY IMAGES: THE ROLE OF LOCAL OPTIMAL INFORMATION

5.1 Theoretical background and rationale

In the real world, humans are constantly exposed to partially occluded objects, which the visual system must analyze and recognize very quickly for survival purposes. Thus, in real scenes, the visual system copes with the recognition of incomplete images, whose mechanisms are still not completely understood. Many studies have demonstrated that humans can successfully recognize fragmented images (Brown & Koch, 2000; Johnson & Olshausen, 2005; Murray et al., 2001; Tang et al., 2018; Ullman et al., 2016), but most of them focus on the rules to solve the occlusion and on how the system fills the missing information. Instead, here we are not interested in understanding the mechanisms through which the visual system binds the fragments into a whole image. We rather focus on the identification of the most relevant fragments to be analyzed and on the extraction of salient local features within these fragments. Hence, we focus on the low-level stages of this process.

As already discussed in the previous chapters, to explain the mechanisms of information selection, several models of visual search employ the concept of *saliency map*, a two-dimensional map that encodes the saliency of the objects in the visual scene (Itti et al., 1998). For this study, it is particularly relevant the open question about the principles driving visual salience and the relative contribution of local (Zhaoping, 2002; Xukun Zhang et al., 2020) and global cues (Itti et al., 1998; Oliva and Schyns, 1997a). Global and local information are related to spatial frequency: low spatial frequencies carry information about the global contrast distribution whereas high spatial frequencies mainly provide fine information about local details (Blakemore & Campbell, 1969; Boeschoten et al., 2005; Kauffmann et al., 2014; Webster & de Valois, 1985). Nevertheless, several past studies have explored the mechanisms of fast vision at different scales and stimulus durations, finding that both coarse and fine spatial information are simultaneously used in fast image categorization (Oliva & Schyns, 1997b; Schyns & Oliva, 1999).

In the present study (Castellotti et al., 2023b), we hypothesize that the perception of incomplete images in fast vision partly starts from the extraction of local high-frequency salient features contained in the visible image fragments. Salient features are the *optimal* features predicted by the reference model (Del Viva et al., 2013).

We explore whether these specific local features still play an important role in more natural settings, where all existing features are kept (*optimal* and *non-optimal*), but the overall available information is drastically reduced. For this purpose, we created images where only a few fragments are shown, and the remaining parts are covered by a grey mask. In this way, we obtain visual stimuli with the same properties as the original images, in which the features are spatially and structurally unaltered, but the overall available information is reduced. To find the essential information needed in order to discriminate a visual scene, we pushed the visual system to its limits: the stimuli had very few visible parts and short durations. Specifically, participants had to covertly attend to a few briefly presented small fragments (or just one fragment) of binarized images and then use them to discriminate the underlying image (target) from another (distractor).

Observers could solve this task mainly by matching the position of black and white parts of the fragmented image and the target (global information), without the need to analyze the internal content of the fragments. If this were the case, we would expect the performance to depend on fragments contrast. On the other hand, performance could be related to the *optimal* information contained in the fragments, as predicted by the reference model. In this case, we would expect performance to depend on the number of local *optimal* features contained in the fragments. With multiple fragments covert attention could potentially be directed toward one of them; for this reason, we also measured discrimination by showing just a single fragment. This allowed us to correlate correct responses to the specific local information and contrast.

We then repeated the same discrimination task randomly inverting the contrast of the target and/or the distractor image. The purpose of this manipulation is to reduce the contribution of global information, given by the position of black/white large areas, and bring out the contribution of high-frequency components that could be masked by the prevalence of positional cues in original-contrast images. The response of complex cortical cells is more or less independent of contrast inversion, while it depends on the local spatial distribution of luminances (Baylis & Driver, 2001; Niell & Stryker, 2008).

Before testing our main experimental hypothesis, we conducted two preliminary experiments to test the limits for discrimination of our fragmented digitized images, presented for a very short time, to probe the size and number of the fragments to be used in the main experiment.

5.2 Aim of the study

In Study 4 (Castellotti et al., 2023b), we aim to show that local information, when derived from maximum-entropy optimization criteria coupled with strict computational limitations, contributes to fast image discrimination. To study this, we compare how much observers' performance relies on local high-frequency components and on low-frequency global cues in very challenging conditions for our visual system.

5.3 Materials and methods

5.3.1 Participants

Twenty young volunteers took part in this study. Ten observers (mean age = 25.3 ± 1.8 years) participated in Preliminary experiment 1, and five of them (mean age = 25.2 ± 1.8 years) also participated in Preliminary experiment 2. Ten other observers (mean age = 26.5 ± 2.9 years), all different from those of the preliminary experiments, participated in the Main experiment. All observers had normal or corrected to normal vision and no history of visual or neurological disorders.

5.3.2 Apparatus and set-up

The apparatus and set-up were the same for the Preliminary and the Main experiments. All stimuli were programmed on an ACER computer running Windows 10 with Matlab 2018b, using the Psychophysics Toolbox extensions (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997). The experiment was displayed on a gamma-corrected CRT Silicon Graphics monitor (1152x864 pixels resolution, 38.5x29.5 cm, 120 Hz refresh rate), subtending 38.5x29.5 degree of visual angle at a 57 cm viewing distance. All experiments were carried out in a completely dark room.

5.3.3 Procedure and stimuli

5.3.3.1 Preliminary experiment 1

The experimental procedure is represented in **Figure 34A**. Each trial started with the presentation of a white fixation point (300 ms) on gray background (14 cd/m^2) followed by the brief presentation (25 ms) of one stimulus in the center of the screen. Stimuli were composed of a certain number of image fragments of different sizes, resulting in a kind of “covered” image, revealing only small visible parts to the observer (see the paragraphs below for stimuli details). Immediately after, a mask appeared for 500ms, followed by two black-white images sequentially presented for 350ms each. One of the two images corresponded to the fragmented "covered" image (*target*), while the other (*distractor*) was randomly extracted from the set of images used (see the paragraphs below for image details). At each trial, the target was randomly presented in the first or the second interval. Images in the task were randomly displaced diagonally by 10 pixels, either to the top-left, top-right, bottom-left, or bottom-right, with respect to the position of the fragmented "covered" image. This spatial shift was introduced to avoid exact spatial matching between stimulus and target image. Observers were required to discriminate the target in a two-interval forced choice task (2IFC), by pressing a computer key.

Stimuli were prepared starting from 327 1-bit black and white renditions of naturalistic images, extracted from the same public database (Olmos & Kingdom, 2004) used in the original work (Del Viva et al., 2013). Images' size was 918x672 pixels, subtending 32.4x23.7° of visual angle at 57 cm. The luminance of white, black, and medium gray was 35 cd/m^2, 1 cd/m^{2,} and 12 cd/m^2, respectively.

In Preliminary experiment 1, we measured discrimination as a function of the image's visible area. Some examples of stimuli are reported in **Figure 34B**. We used the following stimulus configurations: the whole image as a control (100% visible area, see **Figure 34B – first panels**); a squared "frame" comprised between 4.8° and 8.8° of eccentricity (35.8% visible area, see **Figure 34B – second panels**); ten image fragments revealing different fractions of image area: 7.5% (size of all fragments 2.4°x2.4°), 2% (size of all fragments 1.2°x1.2°), 0.47% (size of all fragments 0.6°x0.6°) and 0.12% (size of all fragments 0.3°x0.3°; see **Figure 34B – third to sixth panels**, respectively). In these cases, the rest of the image was covered by uniform grey pixels. For each area, image fragments were randomly selected from all possible combinations satisfying the following conditions: i) they had to be comprised in the 4.8°- 8.8° eccentricity frame (stimuli presented within this eccentricity are well visible even if observers have to maintain fixation in the center, as shown with other tasks; see for example, Larson & Loschky, 2009; Staugaard et al., 2016); ii) they had to be evenly distributed within the frame (three fragments on the top and bottom sides of the frame, and two fragments on each lateral side; iii) they could not overlap with each other. The chosen frame width guarantees that criteria ii) and iii) are met. For each image, five different fragments' configurations were created to minimize memory effects (see **Figure 34C** for an example), for a total of 1635 different stimuli for each area. A total of 3000 trials per observer were run (300 trials for the control and frame conditions and 600 trials for each other condition). Each specific image configuration in each condition has been shown on average 1.2 times to each participant, preventing the association of a specific configuration of fragments to a target.

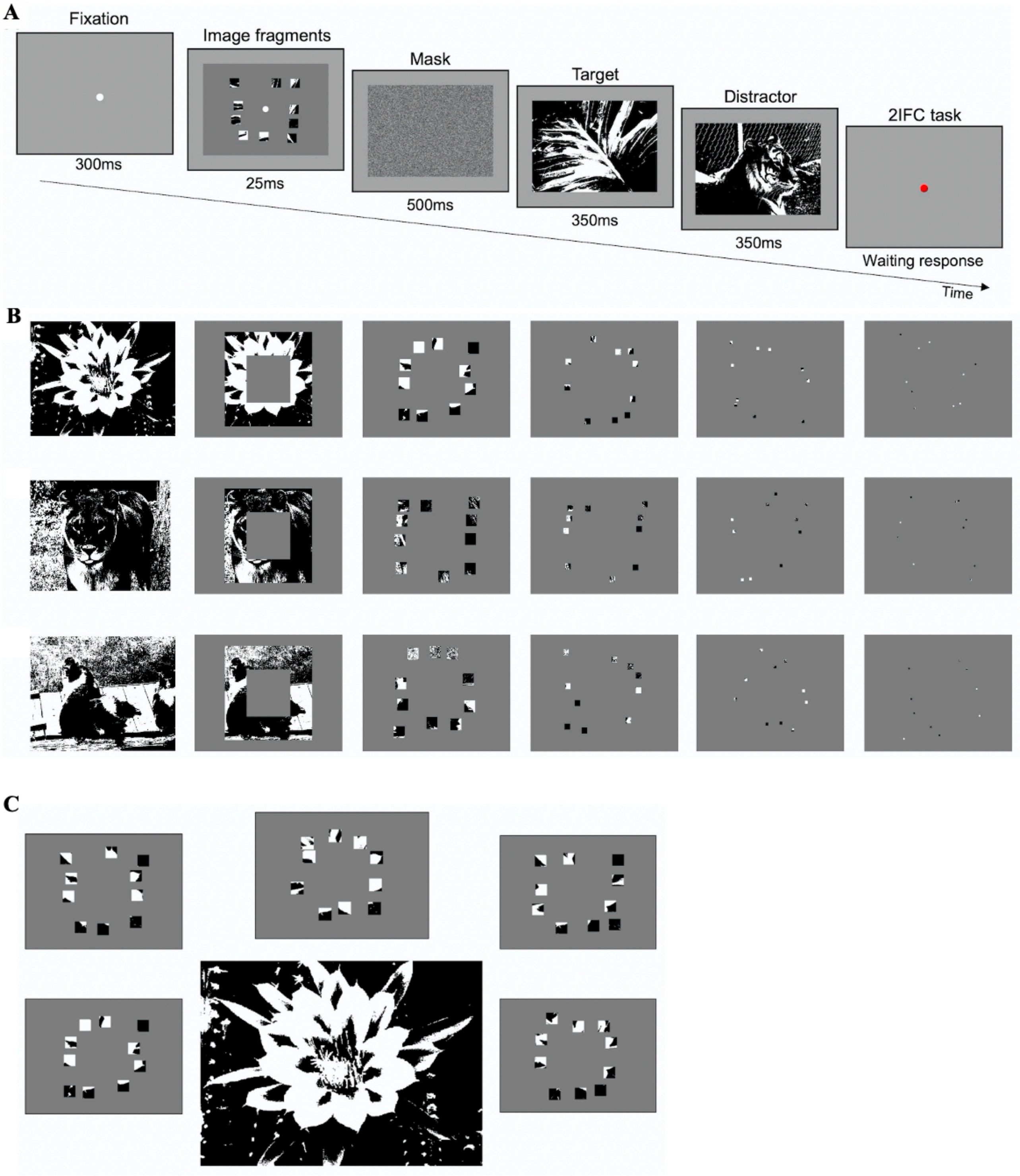

Figure 34. Study 4 – Preliminary experiment 1: procedure and stimuli. (A) Representation of experimental paradigm. (B) Examples of stimuli. Examples of different stimuli configurations for three images. From left to right: the first image is the control stimulus, the second is the "frame" stimulus, and the others show 10 fragments of decreasing size (in order: 7.5%, 2%, 0.47%, and 0.12%), positioned within the frame. **(C) Examples of different fragments' configurations for a specific image.** Five different stimuli with ten 2.40°x2.40° fragments, covering 7.5% of the image area. Figure retrieved from (Castellotti et al., 2023b).

5.2.3.2 Preliminary experiment 2

Preliminary experiment 2 followed the same procedure as Preliminary experiment 1 (see Figure 34A). We measured discrimination as a function of the number of fragments of different sizes covering two different visible fractions of image areas (2% and 7.5%). The fragments were still positioned in the 4.8°-8.8° eccentricity frame. Some examples of stimuli are reported in **Figure 35**. For 2% of the area we used: three 2.4°x2.4° fragments (randomly distributed across the frame), ten 1.2°x1.2° fragments (three fragments located on the top and bottom sides of the frame, and two fragments on the left and right sides), and forty 0.6°x0.6° fragments (twelve fragments located in the upper and lower side, and eight fragments in the left and right sides; see **Figure 35 – left side panels,** from top to bottom respectively). For 7.5% of the area we used: ten 2.40°x2.40° fragments (three fragments located on the top and bottom sides of the frame, and two fragments on the left and right sides), forty 1.2°x1.2° fragments (twelve fragments located on the top and bottom sides of the frame, and eight fragments on the left and right sides), and one hundred and sixty 0.6°x0.6° fragments (forty fragments located in the top, bottom, left, and right part of the image frame) (see **Figure 35 – right side panels,** from top to bottom respectively). For each image, five different fragments' configurations were created, for a total of 1635 different stimuli for each area (see Figure 34C). A total of 3600 trials per observer were run (600 trials for each condition). Each specific image configuration in each condition has been shown on average 1.1 times to each participant.

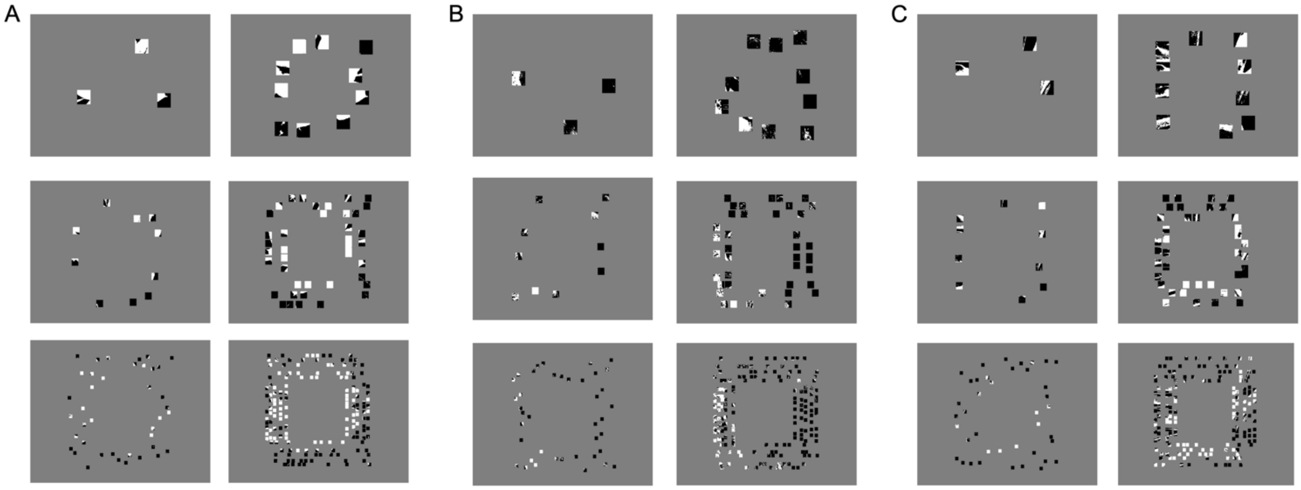

Figure 35. Study 4 – Preliminary experiment 2: stimuli. (A-C) Examples of stimuli. Examples of different stimuli configurations for three images. In the left columns, fragments revealed 2% of the image area, and in the right columns, fragments revealed 7.5% of the image area. Fragments' size in the images of each column decreases by fifty percent going from top to bottom; whereas fragments in the same row have the same size but vary in number. Figure retrieved from (Castellotti et al., 2023b).

5.3.3.3 Main experiment

The Main experiment follows the same procedure (2IFC) and used the same set of images (Olmos & Kingdom, 2004) as those of the Preliminary experiments 1 and 2, but participants were engaged in two different tasks: a task with original-contrast images and a task with randomly inverted-contrast images. In the first task, both the target and the distractor were digitized versions of the original images (as in Figure 34A). In the second task, in some randomly selected trials, the target and/or the distractor had their contrast inverted with respect to their original version (**Figure 36A**). Therefore, in some trials, both the target and the distractor could be presented with their original or inverted contrast, while, in other trials, only one of them could have inverted contrast. With this manipulation, we aim at reducing the probability of solving the task by matching the position of black and white spots in the fragments to those in the images (see **Figure 37**). Each image has been presented to each participant on average 37.7 times, either as a target or distractor.

In both tasks, the same stimuli conditions were tested. Some examples of stimuli are reported in **Figure 36B, C, D**). Stimuli consisted of one or ten fragments (see **Figure 36B, C, D – first and second columns**, respectively) with different sizes: 2.4°x2.4° and 1.2°x1.2° (see **Figure 36B, C, D – first and second rows**, respectively). The total area revealed by these fragments was 0.2% and 0.75% with one fragment, 2% and 7.5% with ten fragments. The characteristics of the stimuli (luminance, fragments distribution, and eccentricity) were the same as those used for Preliminary experiments 1 and 2. In the condition with 10 fragments, for each image, five different fragments' configurations were created, for a total of 1635 different stimuli for each area (Figure 34C). In the condition with 1 fragment four/five different configurations were created, for a total of 1144 and 1253 different stimuli for 0.2% and 0.75% area, respectively (see **Figure 36E**). In the Main experiment, each observer performed 2400 trials in total: 1200 trials in the task with original-contrast images (300 trials for each stimulus condition), and 1200 trials in the task with randomly inverted-contrast images (300 trials for each stimulus condition). Each specific image configuration in each condition has been shown on average 1.1 times to each participant.

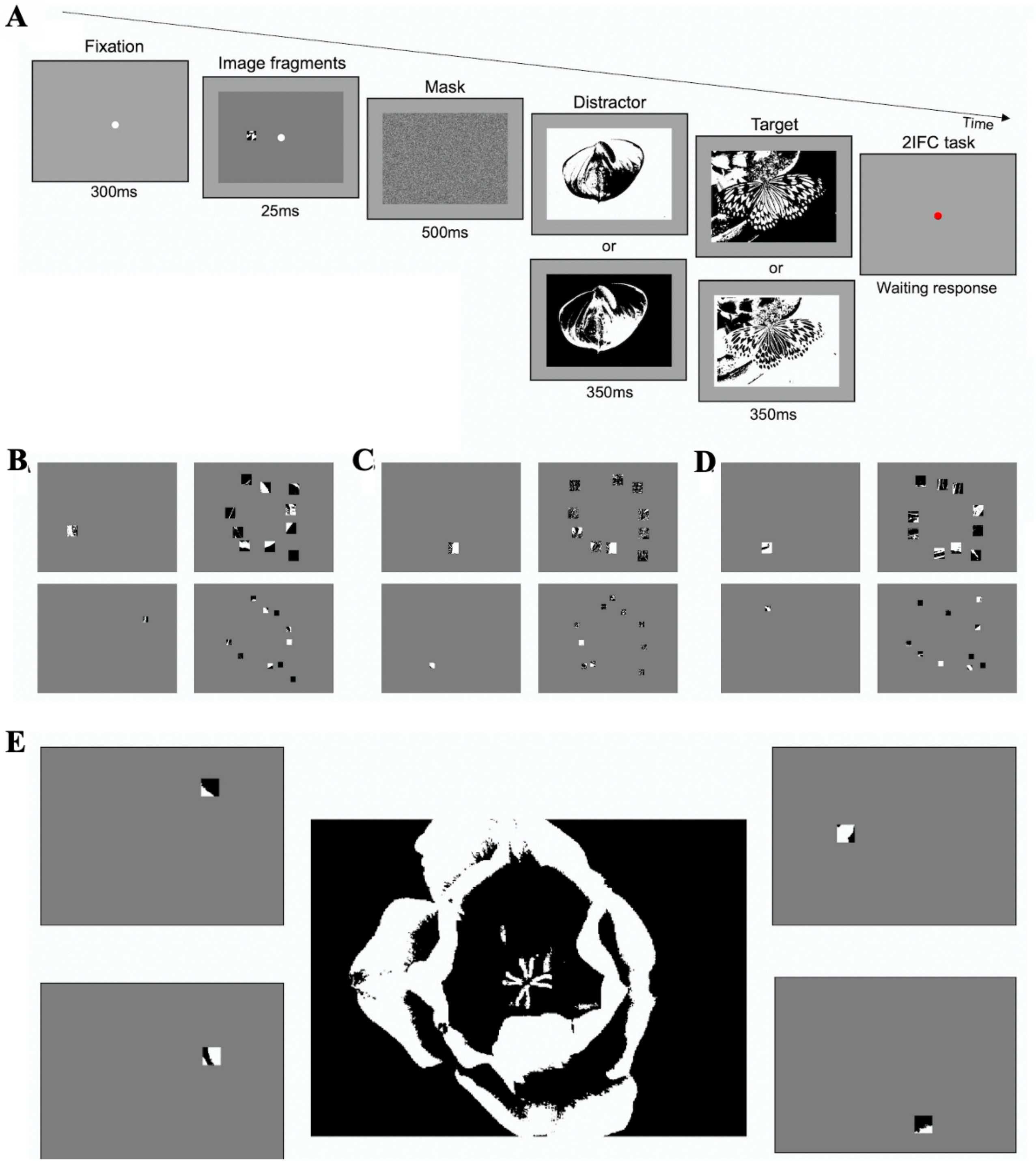

Figure 36. Study 4 – Main experiment: procedure and stimuli. (A) Experimental paradigm. In the upper row is shown a trial where the distractor has inverted contrast and the target has its original. The two panels below indicate that target or distractor can have a contrast inverted with respect to those shown above. **(B-D)** Examples of different stimuli configurations for three images. Fragments in the images of each column are the same number but their size decreases by fifty percent from top to bottom; fragments in the images of each row have the same size but vary in number (one or ten). **(E)** Examples of different fragments' configurations for a specific image. Four different stimuli with one 2.40°x2.40° fragment, covering 0.75% of the image area. Figure retrieved from (Castellotti et al., 2023b).

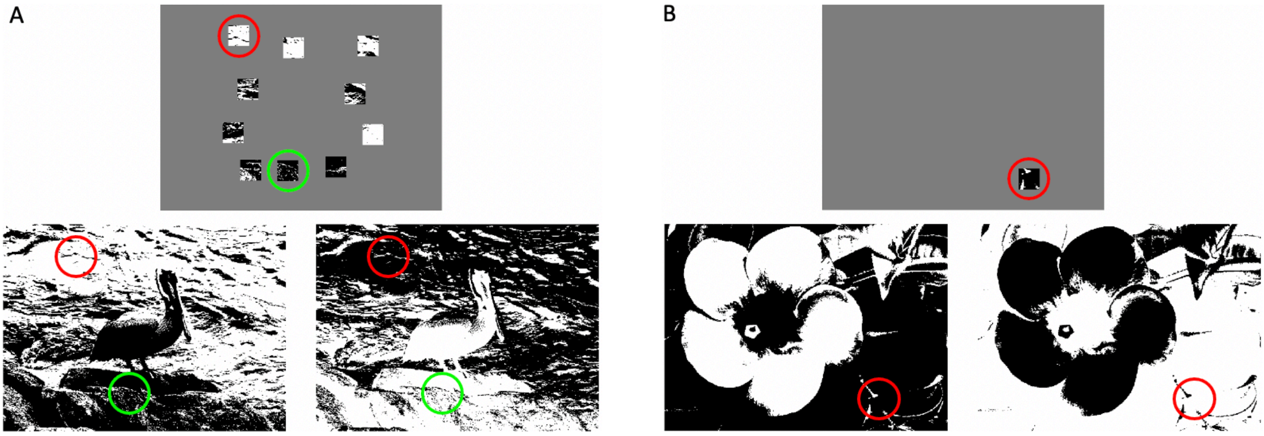

Figure 37. Study 4 – Main experiment: examples of tasks with original and inverted contrast images. The upper panels show two examples of fragmented images. The lower left-side panels show the original-contrast images; the lower right-side panels show inverted-contrast images. In these examples, the position of fragments with large black or white parts (red/green circles) can be easily matched in the original contrast images. **(A) Ten fragments (7.5% area).** In the original-contrast image (left-side panel), observers can match the position of the almost all-white fragment presented in the upper-left part of the image and that of the almost all-black fragment presented in the lower part to discriminate the target. Instead, with the inverted-contrast image (right-side panel), this positional match cannot be done. **(B) One fragment (0.75% area).** In the original-contrast image (left-side panel), observers can match the position of the almost all-black fragment presented in the lower-right part of the image to discriminate the target. In the inverted-contrast image (right-side panel), observers cannot find the black spot in the lower-right part of the target. Figure retrieved from (Castellotti et al., 2023b).

5.3.4 Data processing and statistical analysis

In all experiments, we measured the percentage of correct responses of each observer in each condition of visible area.

In Preliminary experiments 1 and 2, non-parametric one-way repeated-measures ANOVAs (Friedman's tests) with Conover post hoc comparisons (Bonferroni correction) were used to test differences between averaged performances across conditions. In Preliminary experiment 1, we also performed a one-sample Wilcoxon signed-rank test to assess whether the averaged performance in the condition with the smallest visible image area was still above the chance level (i.e., statistically different from 50%).

In the Main experiment, non-parametric two-way repeated-measures ANOVAs (Durbin tests) with Conover post hoc comparisons (Bonferroni correction) were used to test differences between average participants' performances in each condition of visible area in the original vs. inverted contrast tasks.

In addition, all observers' data were pooled together to calculate the performance as a function of fragments' contrast and signal-to-noise ratio (SNR) in each condition of visible area.

We calculated the Weber contrast of the fragment as follows: we first averaged the pixel values within the fragment (black = 0, white = 255), then this averaged value was subtracted from the background value (grey = 127), and finally the absolute value of the ratio between the result of the subtraction and the background was calculated. In the stimuli containing ten fragments, the average contrast of the fragments was considered. The performance was then analyzed as a function of Weber contrast (bins of 0.2 each).

To quantify the saliency of each fragment we calculated the signal-to-noise ratio (SNR), that is the number of optimal features, predicted salient by the reference model, over the total number of features. Specifically, we considered a set of 50 *optimal* features, 3x3 pixel large (see Figure 8C), each subtending ~0.1°x0.1° of visual angle (about 12 c/deg spatial frequency). In the stimuli containing ten fragments, the average SNR of the fragments was considered. The performance was then analyzed as a function of SNR (bins of 0.05 each).

For each SNR bin, we calculated the average contrast of fragments with the standard error. The Pearson linear-correlation coefficient between SNR and contrast was then calculated.

Given the strong correlation between fragments' contrast and SNR, to quantify their relative contribution to the performance, we created a new variable by subtracting, in each trial, the standardized values from each other (SNR – contrast).

Data from all conditions of visible area (7.5%, 2%, 0.75%, 0.2%) were pooled together and GLMMs with a binomial error structure were performed. In the task with original contrast images, the model included three fixed factors: i) SNR-contrast difference (standardized); ii) target order presentation, to test whether the performance depended on the fact that the target was in the first vs. second interval; iii) image repetition number (i.e., the frequency of occurrence of each image as target or distractor), to control for possible effects of visual memory. Participants and stimuli were included as random effects. In the task with randomly inverted-contrast images an additional fixed factor was included: iiii) target contrast inversion, to test whether the performance changed in the trials where the target was presented with original or inverted contrast.

We then compared (z-tests) the probability of correct responses (with binomial standard deviations) between the task with original-contrast images and the one with random contrast inversion. This was done separately for the trials where the target had original contrast and for those where the target had inverted contrast.

Finally, a GLMM was run in the task with randomly contrast-inverted images including only the trials where the target had original contrast.

5.4 Results

5.4.1 Preliminary experiment 1

Performance in Preliminary experiment 1 averaged across participants is reported in **Figure 38**. As expected, the percentage of correct responses increases with the size of the image fragments (i.e., the amount of visible area of the image). On average, observers' performance ranges from 55% for the smallest visible area to 83% when the full image is shown (100% area). Friedman's test showed a main effect of the visible area ($\chi2(5) = 45.3$, $p < 0.001$, W = 0.46). All Conover post hoc comparisons (Bonferroni correction) are reported in **Table 2**.

The average performance obtained by showing the smallest image area also resulted statistically different from 50% ($Z(9) = 55$, $p = 0.002$), showing that observers are able to discriminate an image based on very little information.

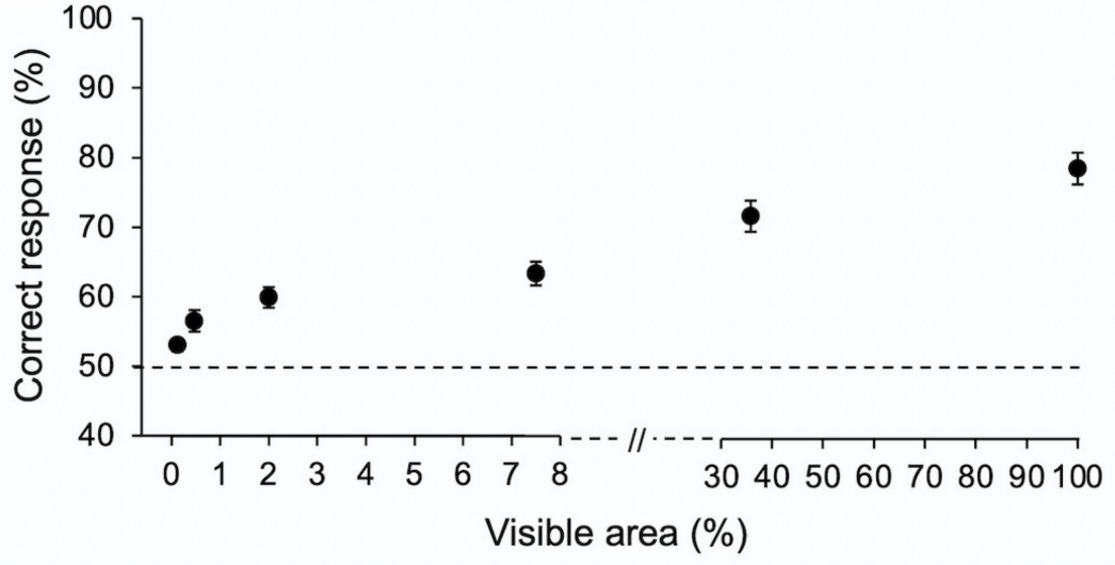

Figure 38. Study 4 – Preliminary experiment 1: performance as a function of images' visible area. Performance averaged across participants with SE. Figure retrieved from (Castellotti et al., 2023b).

Table 2. Study 4 – Preliminary experiment 1: Conover's post-hoc comparisons (Bonferroni correction) across average performances for each area condition (3000 trials in total per observer: 300 trials for 100% and 35.8% area conditions and 600 trials for each of other conditions).

	Area 100%	Area 35.8%	Area 7.5%	Area 2%	Area 0.47%	Area 0.12%
Area 100%	-	*t*=0.9, *p*=1	*t*=2.4, *p*=0.2	*t*=3.5, *p*=0.01	*t*=4.6, *p*<0.001	*t*=5.2, *p*<0.001
Area 35.8%	-	-	*t*=1.5, *p*=1	*t*=2.6, *p*=0.21	*t*=3.6, *p*=0.01	*t*=4.4, *p*<0.001
Area 7.5%	-	-	-	*t*=1.0, *p*=1	*t*=2.1, *p*=0.63	*t*=2.9, *p*=0.08
Area 2%	-	-	-	-	*t*=1.0, *p*=1	*t*=1.9, *p*=1
Area 0.47%	-	-	-	-	-	*t*=0.8, *p*=1
Area 0.12%	-	-	-	-	-	-

5.4.2 Preliminary experiment 2

In preliminary experiment 2, we compared the observers' performance when the same amount of image area is revealed by showing a different number of fragments of different sizes. Performances are reported in **Figure 39**. For both areas tested (2% and 7.5%), the percentage of correct responses tends to be greater with few big fragments than with more small fragments even if none of the results are statistically significant. When the size of the patches remains constant but their number increases, thus revealing a bigger amount of image area to the observers, the performance slightly increases in all conditions, although not significantly.

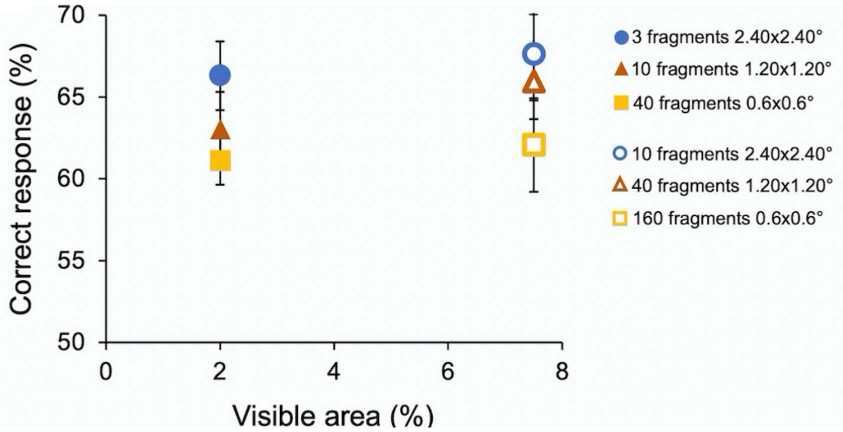

Figure 39. Study 4 – Preliminary experiment 2: performance as a function of the number and size of image fragments. Performance averaged across participants (n=5) with SE. Filled symbols indicate fragments revealing 2% of the area; empty symbols indicate fragments revealing 7.5% of the area. Symbols with the same shape indicate a different number of fragments of the same size. Figure retrieved from (Castellotti et al., 2023b).

5.4.3 Main experiment

In the Main experiment, we first analyzed the percentage of correct discrimination in the two tasks. In the task with original-contrast images (**Figure 40A**), when ten fragments are presented, observers' discrimination is 63.3% ± 1.8% (SE) for 2% area and 68.8% ± 2.5% for 7.5% area (**Figure 40A – left panel**). With one single fragment, the average observers' performance is 60.7% ± 2% at 0.2% area and 64.3% ± 1.6% at 0.75% area (**Figure 40A – right panel**). In the task with randomly inverted-contrast images (**Figure 40B**), with ten fragments discrimination performance is 61.1% ± 1.8% at 2% area and 66.7% ± 2.2% at 7.5% area (see **Figure 40B – left panel)**. With one single fragment, the average observers' performance is 58.3% ± 1.3% at 0.2% area and 63.6% ± 2.1% at 0.75% area (**Figure 40B – right panel**).

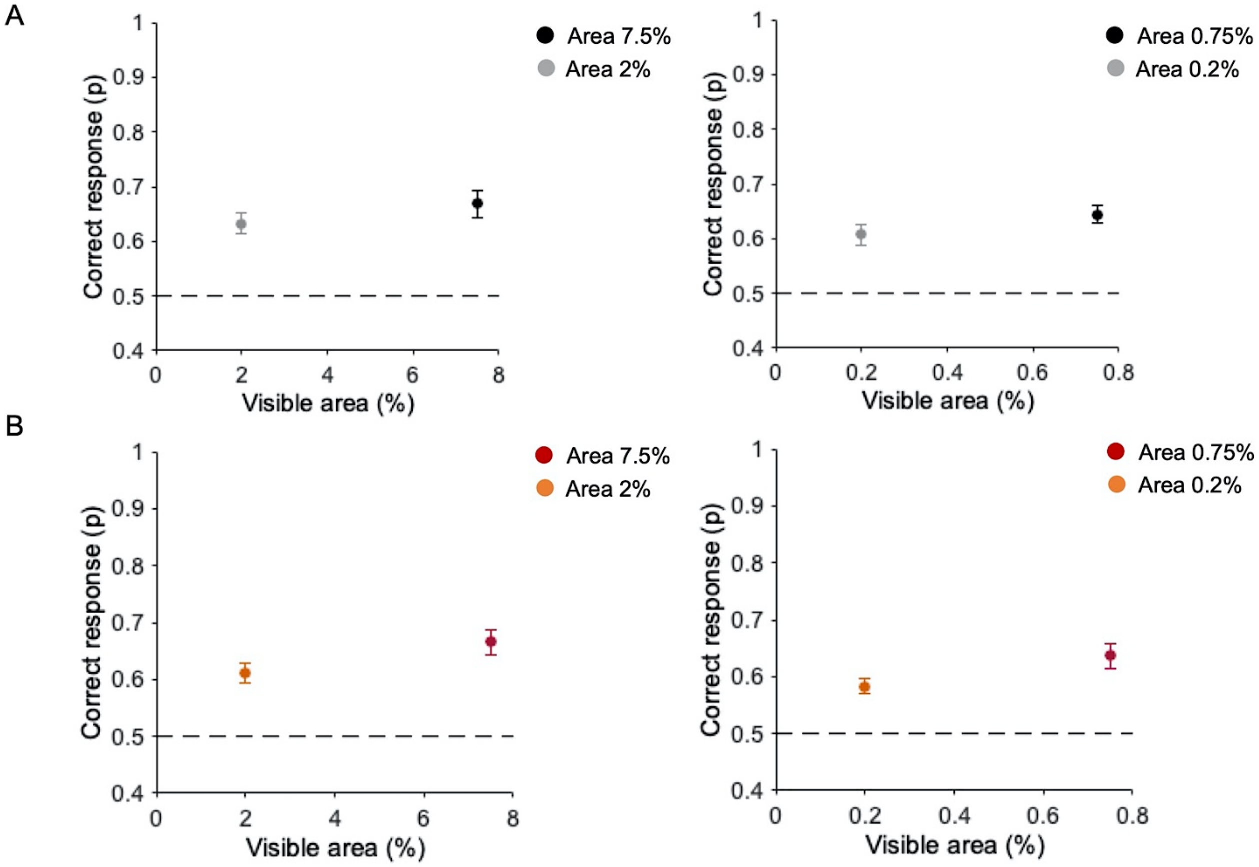

Figure 40. Study 4 – Main experiment: performance for different areas and number of fragments. (A) Task with original-contrast images. (B) Task with randomly inverted-contrast images. Left panels: average performance (n = 10) for ten fragments (2% and 7.5% of area); Right panels: average performance (n = 10) for one fragment (0.2% and 0.75% of area). Errors are SE across participants. Observers performed 2400 trials in total (300 trials for each condition). Figure retrieved from (Castellotti et al., 2023b).

Durbin test between performances with original- vs. randomly inverted-contrast images confirmed the effect of visible area ($\chi^2(1) = 9.2$, $p = 0.002$, W = -20) but no statistical differences emerged across the two tasks ($\chi^2(1) = 0.2$, $p = 0.61$). This suggests that, even if in some trials of this task there is no correspondence between the contrast of the fragments and that of the target image, the overall performance is comparable to that obtained in the task with original-contrast images.

We then investigated to what extent the performance depended on the *saliency* of the local high-frequency features contained in the fragments presented (as predicted by the constrained maximum-entropy model), or on the global luminance information (Weber *contrast*). Firstly, we calculated performance as a function of SNR and contrast separately. In the task with original contrast images, performance does not depend on SNR, and it does not seem to be related to fragments' contrast as well, although there is a tendency to increase with contrast with multiple fragments (**Figure 41A-B)**. Instead, in the task with randomly inverted-contrast images, the performance is higher for lower contrasts and decreases for higher contrasts, whereas it increases from lower to higher SNR (**Figure 41C-D)**.

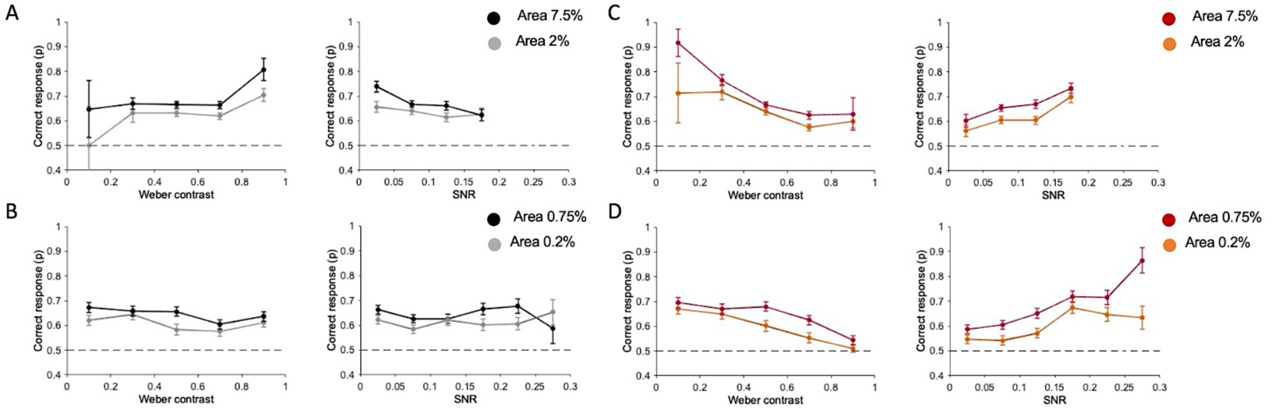

Figure 41. Study 4 – Performance as a function of fragments' Weber contrast and SNR. (A) Task with original-contrast images with ten fragments. (B) Task with original-contrast images with one fragment. (C) Task with randomly inverted-contrast images with ten fragments. (D) Task with randomly inverted-contrast images with one fragment. Error bars are binomial standard deviations. Observers performed 1200 trials in total (300 trials for each stimulus condition). Figure retrieved from (Castellotti et al., 2023b).

Note however that fragments' contrast and SNR are negatively correlated (**Figure 42**; 7.5% area: r = -.63, $p < 0.001$; 2% area: r = -.72, $p < 0.001$; 0.75% area: r = -.60, $p < 0.001$; 0.2% area: r = -.69, $p < 0.001$). This correlation depends on the nature of the fragments and the way the two variables have been calculated: fragments with lower contrast are those containing a higher number of *optimal* features (high SNR), because high SNR reflects into a textured stimulus, and averaging alternations of many black and white pixels, leads to low Weber contrast. On the other end, fragments with higher contrast are those with large black/white parts and therefore contain a few *optimal* features (see **Figure 43**). Note that the maximum SNR in the case of ten fragments (0.2) is lower than for one fragment (0.3) because, being the contrast mediated across ten different parts, the probability of having fragments with large black and white parts (and consequently low SNR) is higher.

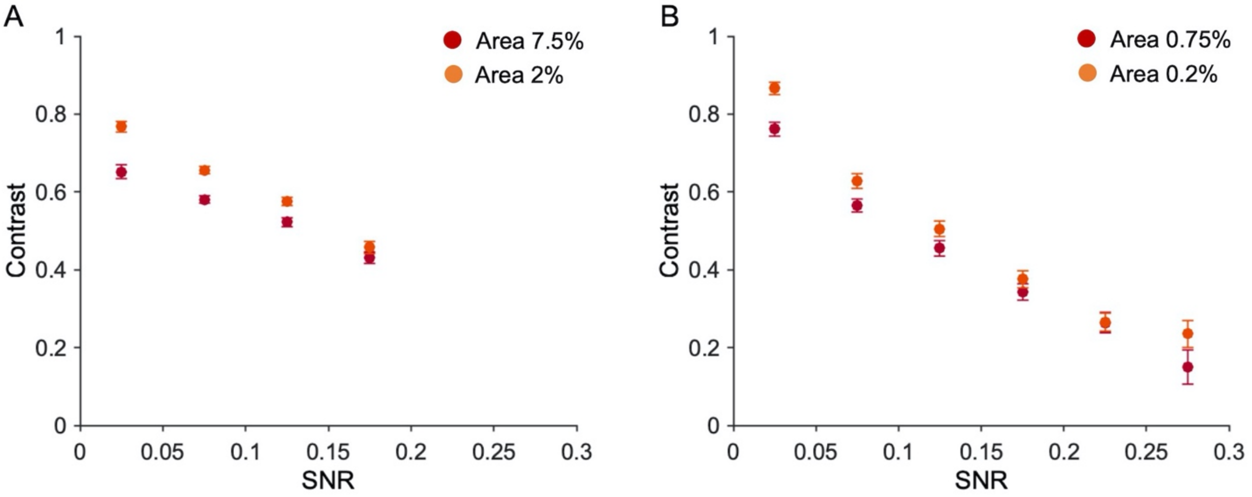

Figure 42. Study 4 – Main experiment: fragments' contrast vs. their SNR. (A) Ten fragments. Number of occurrences in each bin (from the first to the last bin) = 7.5% area: 345,1380,783,491; 2% area: 497,1202,857,444. **(B) One fragment (bins of 0.05 each).** Error bars are standard errors. Number of occurrences in each bin (from the first to the last bin) = 0.75% area: 834,873,586,425,235,47; 2% area: 748,764,620,440, 308,120. Figure retrieved from (Castellotti et al., 2023b).

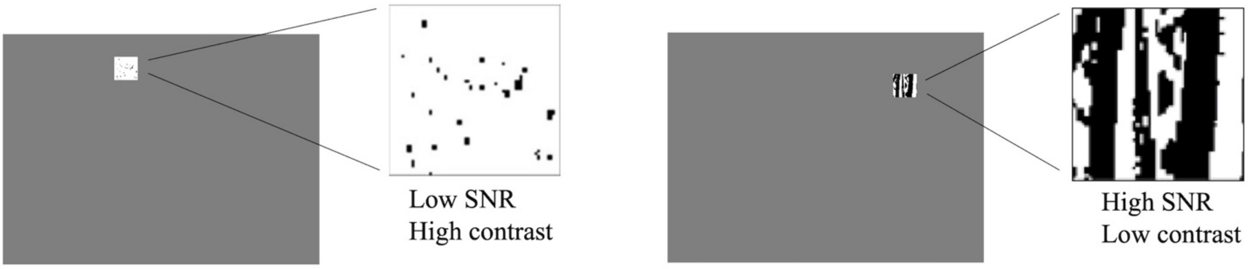

Figure 43. Study 4 – Relationship between contrast and SNR in a fragment. Left panel: example of an almost totally white fragment, which has high contrast with respect to the background, but contains few *optimal* features (low SNR). Right panel: example of a fragment containing a textured internal structure, which has a contrast similar to the background (low contrast) but contains a high number of *optimal* features (high SNR). Figure retrieved from (Castellotti et al., 2023b).

Since the correlations between SNR and contrast are quite high, in the following analysis we used the difference between standardized SNR and contrast, instead of considering them as two separate variables. In this way, the contributions of SNR and contrast to the performance can be separated. Moreover, in a 2IFC task, the order of target presentation might affect the performance, as well as the frequency of occurrence of each image: repeated presentations of the same image as target or distractor might induce visual learning of the images.

For the task with original contrast images, we then performed a GLMM with three fixed factors: SNR-contrast difference (standardized), target order presentation, and image repetition number. Participants and stimuli were included as random effects. The GLMM reveals no effect of the difference between standardized SNR and contrasts ($\chi^2(1) = 0.24$, $p = 0.62$), but a main effect of order ($\chi^2(1) = 9.1$, $p = 0.002$) and image repetition number ($\chi^2(1) = 19.2$, $p < 0.001$) emerged.

Overall, these results indicate that, in the task with original-contrast images, the performance does not depend on SNR (as shown in **Figure 44A**), and it does not seem to be related to fragments' contrast either (although there is a tendency to increase with contrast with multiple fragments; see Figure 41A-B). Given our hypotheses, we argue that in this condition observers do not rely on local cues and possibly use the position of black and white spots to solve the task. This hypothesis seems to be further supported by the fact that the performance is higher when the target is presented in the first interval of the 2IFC. Indeed, the match between the fragments and the corresponding image is easier if the target is temporally closer and its presentation is not interspersed with the appearance of the distractor.

We then performed the same analysis in the task with randomly inverted-contrast images (**Figure 44B**), used to reduce the contribution of positional global cues and to bring out the contribution of high-frequency *optimal* features (see Figure 37).

In this task, an additional factor was included in the GLMM. Considering all visible area conditions (12000 trials in total), due to the random nature of inversion, the target contrast alone was inverted in 24.5% of trials, the distractor contrast alone was inverted in 25.2% of trials, the contrast of both target and distractor was inverted in 22.8% of trials, and the contrast of both target and distractor was kept original in 27.4% of trials. In principle, these different target conditions could affect performance.

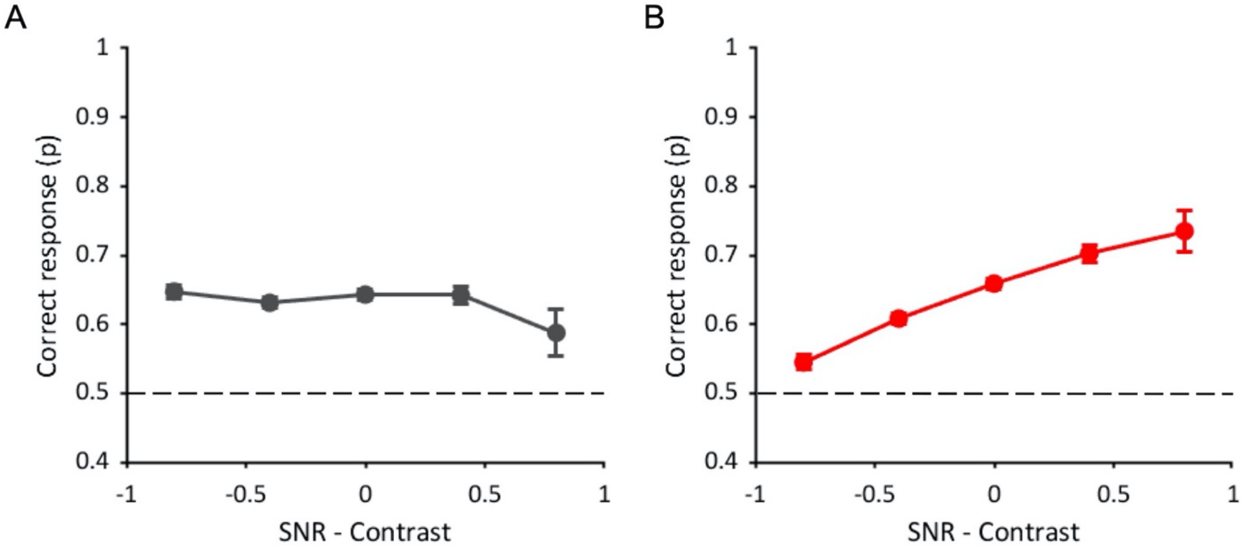

Figure 44. Study 4 – Main experiment: performance as a function of the difference between standardized SNR and contrast. (A) Task with original-contrast images. (B) Task with randomly inverted-contrast images. Data from all tested areas are pooled together. Errors are binomial standard deviation. Dashed lines represent chance level. Figure retrieved from (Castellotti et al., 2023b).

The GLMM analysis was thus performed with four fixed factors (standardized SNR-contrast difference, target order presentation, image repetition number, and target contrast inversion) and two random effects: participants and stimuli. The analysis shows a significant effect of SNR-contrast difference ($\chi^2(1) = 128.4$, $p < 0.001$) on performance. Indeed, performance increases with this difference (Figure 44B), suggesting that SNR prevails over contrast in driving the performance. The target order factor is instead not statistically significant ($\chi^2(1) = 0.07$, $p = 0.78$), meaning that the performance does not change whether the target image is shown in the first or the second interval of the 2IFC task. These results confirm further our hypothesis that, in this condition, participants change their strategy: they do not rely on positional cues anymore, but rather they use local information, therefore target order does not affect the performance. Again, the analysis reveals an effect of the image repetition number ($\chi^2(1) = 36.2$, $p < 0.001$). The target contrast inversion factor is also statistically significant ($\chi^2(3) = 45.5$, $p < 0.001$). Indeed, the performance with original-contrast target (65% ± 0.006%) is higher than with inverted-contrast target (60% ± 0.006%).

Interestingly, the performance in the task with randomly inverted-contrast images in the trials with original-contrast target is also higher than that obtained in the task with original-contrast images (63% ± 0.004%; $z = 2$, $p = 0.04$), although these two conditions are exactly the same.

The GLMM analysis, including only the trials with original-contrast target of the task with randomly inverted-contrast images, reveals a main effect of the difference between SNR - contrast ($\chi^2(1) = 33.9$, $p < 0.001$; see **Figure 45**), and of image repetition number ($\chi^2(1) = 18.4$, $p < 0.001$), but there is no effect of target presentation order ($\chi^2(1) = 0.31$, $p = 0.58$). These results are compatible with those found when considering all trials, independently of target contrast inversion (see Figure 44B). On the other end, these results are different from those found in the task with original-contrast images (see Figure 44A), although these two conditions are exactly the same. See the Discussion section for the interpretation of these results.

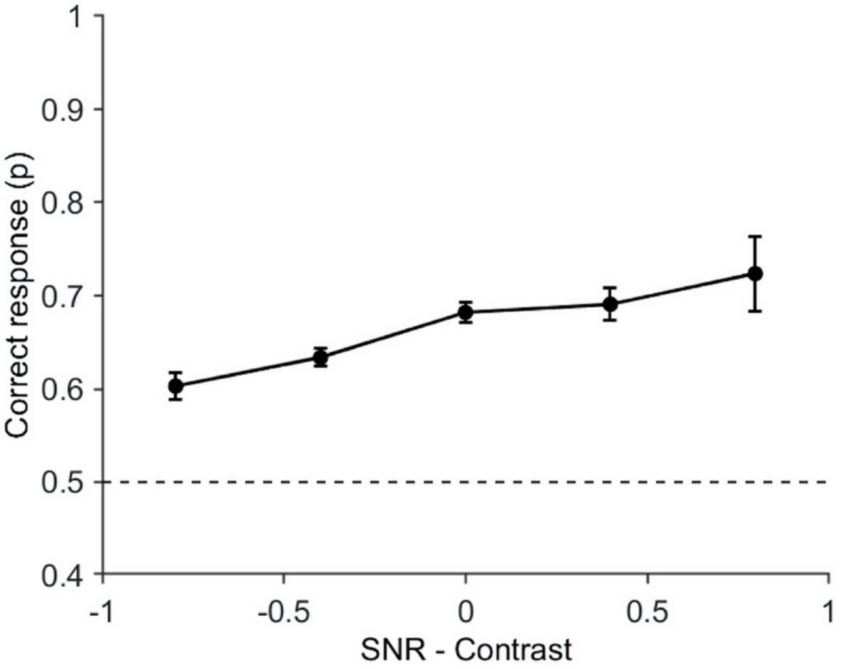

Figure 45. Study 4 – Main experiment: performance as a function of the difference between standardized SNR and contrast in trials with original-contrast targets in task with randomly inverted-contrast images. In the graph, data from all participants (n = 10) and area conditions (7.5%, 2%, 0.75%, 0.2%) are pooled together (bins of 0.2 each - binomial standard deviations), considering only trials where the target was not contrast-inverted (6311 trials in total). Figure retrieved from (Castellotti et al., 2023b).

5.5 Discussion

In the present work, we investigated the visual system's ability to quickly discriminate a scene, based on the salience of high-frequency local visual features.

Over the years, different studies have argued that the selection of relevant local elements is based on the simultaneous processing of different visual properties at multiple spatial scales, then combined into a single saliency map (Itti et al., 1998; Itti & Koch, 2001; Torralba, 2003). However, these models do not consider the amount of computing power required by each parallel process. Our reference model, instead, considers the system's computational costs. Considering the finest spatial scale as the most computationally demanding part of the processing and the need for fast analysis, the model applies a lossy data compression algorithm to images at a fine spatial scale (Del Viva et al., 2013). The result of this process is the extraction of a limited number of informative high-frequency visual features, that are used for fast image discrimination and to drive bottom-up attention (Castellotti et al., 2021, 2022).

Before investigating their role in fast discrimination of fragmented images, often presented to the visual system due to occlusions, we showed that observers can discriminate an image presented only for 25ms even when it's almost totally occluded. As expected, correct discrimination increases with the visible area, but is still possible with very little information (0.12%). These findings confirm that humans are very skilled in fast visual discrimination, as already broadly demonstrated (for a review, see Serre et al., 2007). Note however that we pushed the visual system's capacity to its limit, by showing images for the minimum duration necessary for a visual stimulus to reach primary cortical visual structures (Grill-Spector et al., 2000; Kirchner & Thorpe, 2006) and by using a paradigm that is known to be challenging for the observers (i.e., 2IFC tasks lead to higher error than 2AFC, Jäkel & Wichmann, 2006); This might explain why observers did not reach top performance even when the full image is displayed (100% area). Despite this, the minimal percentage of visible area needed to perform the task is much lower (0.12%) than that found in previous studies. For example, Tang et al. (2018) conducted an experiment similar to ours, with occluded or partially visible images presented for different durations, finding that in 25 ms observers robustly recognized objects when they were rendered <15% visible (Tang et al., 2018). The higher performance with a smaller visible area found here could be explained by the different tasks involved: their participants had to choose the right association between the occluded content and five different label options, while ours discriminate between two images.

We also investigated which factors mostly influence the correct discrimination of occluded pictures. That is, we studied whether, with the same amount of visible area, discrimination depends more on the number of visible fragments or on their size. Results show a slight (not significant) preference for a few large fragments, rather than for many small parts. This is somewhat unexpected. However, some have hypothesized that perceptual systems suffer from overload, so the higher the perceptual load of current information, the lower the ability to perceive additional information (Greene et al., 2017). Here a low number of fragments could produce a lower cognitive load (Nejati, 2021; Xing, 2007), hence better performance.

In the main experiment, we investigated the role of the high-frequency model-predicted *optimal* features in fragmented image discrimination by quantifying the saliency of the fragments as the ratio of *optimal* features over the total number of features they contain. That is, the question is whether observers focus on the local internal content of the fragments and use embedded *optimal* features to discriminate the target, or whether they covertly attend to the global contrast information (low frequency). Indeed, since we use black and white stimuli and a 2IFC discrimination task, observers could simply solve the task by matching the position of black and white parts of the fragmented image and the target, without the need to analyze the internal content of the patches.

When low frequencies can be used to perform the task (original contrast), the performance does not depend on the number of *optimal* features contained in the fragments, rather there is a slight tendency to increase with fragments contrast (particularly when ten fragments are shown). These results suggest that in this condition observers do not use local information but possibly use the fragments' global luminance distribution. This hypothesis is further supported by the evidence that, only in the task with original-contrast images, the performance increases if the target is shown in the first interval of the 2IFC task. Indeed, we can assume that the match between the position of the black and white parts of the fragmented image and the target is easier if the latter is temporally closer to the stimulus and there is no other image before it.

A higher performance in the task with original-contrast images than in the task with random contrast inversion would be expected, since, in the former, positional cues can always be used. The fact that the performances in the two tasks are similar suggests that, when the contribution of global information is decreased (random inversion of contrast), observers rely on a different kind of information to discriminate the scene. In fact, we found that the probability of correct discrimination increases with the number of *optimal* features in the fragments, both with one and ten fragments, indicating that observers' responses in the task with random inversion of contrast are based on the local content of the fragments. This change of strategy is further supported by the evidence that, in this condition, the performance does not depend on the target order of presentation. We argue that, since observers do not base their choice on positional cues, it doesn't matter anymore if the target is presented in the first or in the second interval.

In the task with randomly inverted-contrast images in some trials the target still has the original contrast, therefore the global luminance structure of the fragments could still drive discrimination.

Interestingly, considering only these specific trials, the performance is even higher than that obtained in the task where only original-contrast images are used, even though the two conditions are exactly the same. More importantly, correct responses depend on the number of *optimal* features in the fragments, and they are independent of target order, unlike in the task with original-contrast images. These results confirm that the contrast manipulation we applied in this task can change the observers' strategy. In this condition, participants seem to use both global and local information reaching a higher performance than when they rely only on global information. We, therefore, conclude that when less global information is available, local information plays a crucial role.

Note that the set of *optimal* features comprises spatial structures with both contrast polarities; this could explain why the inversion of contrast does not affect discrimination based on local information. The insensitivity to contrast inversion (Baylis & Driver, 2001; Niell & Stryker, 2008) found in V1 complex cells, together with the similarity of spatial structure between model-predicted *optimal* features and the bar and edge-like V1 receptive fields (Hubel & Wiesel, 1965; Figure 3B), strongly suggests that these cells represent the optimal way to transmit information in fast vision. This also highlights the strong predictive power of the constrained maximum-entropy model.

Overall, our findings suggest that local and global analyses interact in fast image processing and that the contribution of the high-frequency *optimal* features significantly emerges when the visual system is tested in very challenging conditions. This means that local information, when derived from maximum-entropy optimization criteria coupled with strict computational limitations, allows fast image discrimination even when the information about the scene is drastically reduced.

This fast local extraction of salient features must be operated very early in the visual pathway (Del Viva et al., 2013; Zhaoping, 2002), and integrated into a global percept at later visual stages. Indeed, in real scenes the visual system "goes beyond the information given" in a local region (Meng & Potter, 2008) and fills in the missing information of occluded images by binding the visible image fragments (Bruno et al., 1997; Johnson & Olshausen, 2005; Meng & Potter, 2008). Also, in daily life, the *a priori* knowledge of the objects helps the visual system in image recognition (Pinto et al., 2015; Stein & Peelen, 2015). Long-term memory, which is capable of storing a massive number of details from the images (Brady et al., 2008), contributes as well. Visual learning effects also occurred in our experiment, since the performance is affected by repeated presentations of the same image. This indicates that participants might have become acquainted with image details, revealing that there are some memory effects at play. Studies of the mechanisms of recognition of incomplete images have also developed information-statistical approaches, the concepts of the extraction of the signal from noise, and models of matched filtration (for a review, see Shelepin, Chikhman & Foreman, 2009).

What cannot be ignored is the fact that while viewing a scene, humans make eye movements several times per second. Therefore, we have recently extended this study including eye movement recording in similar conditions, testing whether saccades are directed toward the most informative fragments to reconstruct the images (Del Viva et al., 2022). To study this, we briefly present a few small image fragments and invite participants to look at one fragment to later perform a discrimination task. Our preliminary results show that eye movements preferentially head toward fragments containing a larger number of optimal features, especially in the session with inverted-contrast images. In this condition, image discrimination is also more accurate when saccades are directed towards fragments containing a large number of optimal features. These preliminary results then confirm that *optimal* local features are used for image discrimination when global information is not fully available and suggest that they are preferential targets of eye movements because of their saliency.

One could still wonder why in the task with randomly inverted-contrast images when using just one fragment the performance depends on fragment size (**Figure 40D**), even though the ratio of *optimal* features to the total is the same. Clearly, some other factors are at play here. A plausible explanation could be that here we have considered for saliency only the finest spatial scale, while these fragments contain edges and bars a lower scale (Callaway, 2005; DeYoe & Van Essen, 1988; Nassi & Callaway, 2009), to the point that a single fragment could be considered as a single feature itself rather than a set of smaller features. Hypothetically, it would be interesting then to run the constrained maximum-entropy model with larger feature sizes to study what kind of optimality emerges on larger scales, and if these new *optimal* features are present in large quantity in the fragments used for discrimination. The problem is that the need for such small feature size is a corollary of the central idea of compression proposed by Del Viva et al. (2013): the number of possible features, that we assume to be a limited resource, increases exponentially with the size of the feature itself and so the amount of computing needed to calculate them. For example, considering that a feature size of 1° (see paragraph 1.1.1.2) corresponds in our experiment to a 28x28 pixel feature, it would require sorting out 10^{236} possible 1-bit features. Clearly, a number that is not manageable by any natural or man-made device. To solve this problem then we have to hypothesize a general system, composed of many layers, in which an image is represented at a progressively large spatial scale, and in which each level feeds the next one, and the level of image reconstruction becomes progressively more complex. Such a framework matches the hierarchical architectures of ventral visual streams and has been exploited by many models based on deep convolutional neural networks, attempting to mimic the multilayer processing of the visual system for image classification (e.g., Cichy et al., 2016, 2014; Eickenberg et al., 2017; Güçlü & van Gerven, 2015; Yamins et al., 2014). A future goal of this group of research is then to develop such a parallel model that extrapolates the idea of constrained maximum-entropy optimal feature selection to multiple layers. In this framework, the *optimal* features extracted at a given layer will be the only information feed to the next and therefore all the possible features in this higher layer are only combinations of the optimal features extracted in the layer below. In this way, it will be possible to reduce the information to a level that becomes computable and compliant with the idea of compression of the model and to extract *optimal* larger features.

Chapter 6

General discussion and conclusions

6. GENERAL DISCUSSION AND CONCLUSIONS

The studies presented in this thesis exploit a variety of experimental designs and techniques and each one has its own specific research question to answer. Besides, all the studies share the same final objective. Namely, they aim to assess whether the saliency of local visual features in fast vision can be determined by information maximization criteria coupled with the computational limitations of our visual system, as predicted by a recent model of early vision (Del Viva et al., 2013).

This model simulates the first stages of visual processing where the demand for fast analysis of large amounts of data requires that the implicit constraints of a biological system with finite capacity cannot be ignored. Therefore, the model assumes that this early-stage filter is subjected to some computational limitations, formalized as limited storage and limited rate of transmission. Finally, it assumes that the system is optimized to preserve the maximum amount of information in the output. These assumptions lead to a unique procedure of image data reduction, extracting simplified *sketches*, that are input to further processing, created by retaining only a limited number of features that are *optimal* information carriers, dropping all the remaining information.

Here we are interested in understanding the role of these local optimally informative visual features in the creation of a saliency map that the oculomotor system could use to drive eye movements toward potentially relevant locations, therefore, ultimately, in their contribution to image reconstruction.

In our first study, *optimal* features turned out to be more salient than others, despite the lack of any clues coming from a global meaningful structure, suggesting that they get preferential treatment during fast image analysis. Also, peripheral fast visual processing of these informative local features seems able to guide gaze orientation (Castellotti et al., 2021). In the second study, we demonstrated that *optimal* features can rapidly and automatically attract the subject's attention in covert and overt attentional tasks, in which "saliency" is implicitly manipulated rather than explicitly cued. That is, they are able to locally boost contrast sensitivity, reduce latencies of target-oriented saccades, decrease the number of saccadic direction errors, and increase anticipatory saccades at their locations, although observers are not required to pay attention to features saliency (Castellotti et al., 2022). The third study adds a new piece to our investigation, showing that *optimal* features interfere with the path of saccades toward a target acting as salient visual distractors (Castellotti et al., 2023a). This suggests that the visuo-oculomotor system rapidly and automatically processes optimally informative features while programming visually guided eye movements. The last study tried to test a more ecological condition by using occluded natural images and found that optimal local information also plays a role in image reconstruction based on little information, as well as global visual elements (Castellotti et al., 2023b).

On the other end, the results of our studies showed that other features, discarded by the model as *non-optimal* features, do not produce the same effects as the *optimal* ones. Namely, they are not judged visually salient, they do not attract attention, and they do not induce saccadic curvature.

Very interestingly, all the effects found with *optimal* vs. *non-optimal* features are comparable to those obtained with features of different luminance. The attention-grabbing effect and the saccadic curvature induced by *optimal* features are indeed the same as those found with high-luminance features, suggesting that the saliency provided by *optimal* features is comparable to that of high-luminance stimuli, if compared on equal grounds. Therefore, visual saliency instantiated by specific spatial structures, determined by information maximization, is comparable to that of luminance-based saliency.

Let me also mention that our studies, besides testing the predicting power of our reference model, employ some novel paradigms to investigate visual saliency that may be useful for other research. Indeed, in our second study, we used a rare double spatial-cueing paradigm, where two simultaneous peripheral cues, with either the same or different assumed saliency, are presented. In the third study, we exploit the phenomenon of saccadic curvature to compare the effects on saccades trajectories produced by different types of visual features used as distractors. Given the results found, we think that in the future such paradigms may be useful tools to test the relative saliency of two stimuli, even if ontologically different from each other, by directly comparing their ability to capture attention or the magnitude of deviation they induce in the saccade trajectory.

Overall, the findings presented in this thesis point to a rapid orientation of saccades toward the salient information provided by features *optimality*. Humans can only fixate and extract detailed information from one small region of space at a time. This makes an efficient selection of relevant local features critical for visual processing and optimal behavior. Decades of work in vision science have argued for such dynamic selection to be based on multiple saliency maps (Itti et al., 1998; Itti and Koch, 2001; Parkhurst et al., 2002; Torralba, 2003). The saliency of *optimal* features is independent of the global image context, leading to the speculation that they may play an important role within the multi-scale analysis of saliency performed by the human visual system.

The saliency map is not derived, in our case, from an algorithm trying to make sense of visual properties determined a-priori (e.g., color, motion, texture) competing at individual image locations. Our salient features are instead a consequence of both the early input data reduction and of the frequency with which they occur in the input. *Optimal* and *non-optimal* features do not differ in low-level properties, such as average luminance or spatial frequency. Therefore, it is worthwhile to reflect upon the properties that makes *optimal* features so much more significant, to the point of eliciting the same effects as if they had different luminance. According to the reference model adopted here (Del Viva et al., 2013), the visual system, to produce an early saliency map of a visual scene, extracts just a very limited salient features*,* based on criteria of maximal entropy within strict bounds on data output rate. O*ptimal* features then represent a compromise between the amount of information they carry about the visual scene and the cost for the system to process them. On the other hand, the *non-optimal* features used in our studies are individually the most informative, but do not meet computational limitations criteria (Del Viva et al., 2013). Therefore, the computational limitations do much more than simply limit the performance of the system; they seem to take a significant role, not only in compression, but also in shaping what the system selects as salient in the input.

Several past studies have explored the mechanisms of fast vision at different scales and stimulus durations, finding that both coarse and fine spatial information are simultaneously used in fast categorization of images (Oliva & Schyns, 1997a; Schyns & Oliva, 1999). Some models build bottom-up saliency maps, based on simultaneous processing of different visual properties at multiple spatial scales that are then somehow combined into a single saliency-map (Itti & Koch, 2001). These models do not address the issue of the amount of computing power required by each of these parallel processes which varies greatly across scales and modalities. The reference model instead revolves entirely around the concept of computational costs. From this viewpoint, the finest usable spatial scale takes naturally a central role. In fact, as a consequence of the properties of the Fourier transform, the information content is proportional to the square of spatial frequency, making the finest scale by large the most computationally demanding part of the processing. Saliency extraction at this scale then, with strong reduction of information, becomes a pressing necessity, and one expects it to play an important role amongst all possible maps involved.

There is still a debate on whether this fast bottom-up extraction of visual saliency map is based mainly on local (Zhaoping, 2002) or rather global clues (Itti et al., 1998; Oliva, 2005; Oliva & Schyns, 1997b; Torralba, 2003). The contribution of local analysis to the global percept of an image has been studied in a past work, within the framework of the present model, by replacing in a sketch the *optimal* features (typically located along within objects contours), with other features that are *non-optimal* carriers of information, keeping their localization in the image unchanged. The disruption of these *optimal* local cues causes a decrease in image recognizability, in spite of its global structure being preserved (Del Viva et al., 2016; paragraph 1.4.6). Our current findings also suggest that image reconstruction processes use local information. While the existence of other mechanisms in addition to what is analyzed here has been proved beyond doubt, this new result allows establishing the existence of a bottom-up reference frame for the extraction of saliency that can efficiently drive the process.

Many studies have proposed that bottom-up saliency maps are represented in early sensory cortices (Zhaoping, 2006; Zhaoping, 2019; Zhaoping & Zhe, 2015) and rely on specific sensory properties. Priority maps are instead less dependent on the detailed physical properties of the sensory input, and account for both the global properties of the scene, the behavioral goals and high cognitive information (as reviewed in Itti & Koch, 2001; Zelinsky & Bisley, 2015). They would rather be represented in higher cortical sensory areas (including parietal and prefrontal areas, Bisley & Goldberg, 2010; Thompson & Bichot, 2005), as well as in subcortical regions closer to the motor output, as the Superior Colliculus for saccade planning (Veale et al., 2017; White et al., 2017). Both earlier studies, supporting the view of saliency maps as represented in early visual cortex (Zhaoping, 2006), and more recent works, suggesting the existence of a priority map in the superior colliculus (Veale et al., 2017; White et al., 2017), agree on the fast nature of such representations. The gaze could then be rapidly oriented toward the maximum-saliency locations highlighted by these maps (Garcia-Diaz et al., 2012; Itti & Baldi, 2009; Itti & Koch, 2000, 2001; Najemnik & Geisler, 2005, 2008).

The saliency map extracted by the constrained maximum-entropy algorithm, for efficient compression, must be created very early in the visual system, and several converging evidence indicates the primary visual cortex as the most likely candidate. First of all, *optimal* features are good approximations, within the limitations of a 3 x 3 grid, of the structure of some receptive fields of neurons found in primary visual areas (Hubel & Wiesel, 1965; Figure 3B). Such elongated edge- and bar-shaped structures haven't been found in the thalamus and superior colliculus (DeAngelis et al., 1995; Derrington & Webb, 2004; Drager & Hubel, 1975; Harutiunian Kozak et al., 1973; Kara et al., 2002), although some studies found orientation selectivity in the superior colliculus (Ahmadlou & Heimel, 2015; De Franceschi & Solomon, 2018; Feinberg & Meister, 2015; Gale & Murphy, 2014; Wang et al., 2010). Then, *optimal* features extraction supports a fine-scale local analysis consistent with V1 (Hubel & Wiesel, 1962; Hubel & Wiesel, 1974; Lennie, 1998). V1 is also the most extended visual area (Lennie, 1998), with larger energy consumption (Lennie, 2003) and higher input/output neural ratio with respect to the retina and other extrastriate areas (Lennie, 1998), making it a good candidate for the information bottleneck required by our model. Finally, V1 is involved in very fast visual analysis (Grill-Spector et al., 2000; Kirchner & Thorpe, 2006). All these observations are consistent with the idea, previously advanced, that the function of V1 is to create a "bottom-up saliency map" enabling a "lossy pre-attentive selection of information", so that data rate can be further reduced for detailed processing (Zhaoping, 2006; Zhaoping, 2019; Zhaoping & May, 2007). Theoretically, V1 neurons can work together to filter information in a manner compatible with the reference model following a hierarchical multilayer processing. For example, cells selective to vertical bars in different points of the image are connected by lateral connections. Therefore, when one or more of these cells detect their preferred stimulus, a higher-order cell to which they are connected integrates this information detecting an extended vertical bar located in a wider region of the image. This means that raw input data are analyzed locally and then combined into more complex features at a lower spatial scale. Then, the characteristics and sizes of the receptive field found in V1 indicate that the processing of higher-order visual areas (V2, V3, V4, V5) is necessary to create a rich representation of the visual scene in terms of structure, spatial position, and semantic meaning.

Another important question is how the visual system could have developed to use the optimal pattern set. We can hypothesize that the model allows for easy algorithms for unsupervised learning. The optimal patterns have probabilities falling within a limited range; a system that is initially sensitive to a wide variety of patterns could converge towards optimality simply by discarding patterns that occur too rarely, or too frequently, in the input. This process might even happen during normal activity, allowing for continuous updating and adaptation to changing external conditions. We can hypothesize that the visual system follows a learning process updating optimal information following the changes in the image statistics. This means that modifying the visual environment would lead to modifications of the system's mechanisms. Indeed, a key property of our brain is its plasticity and, unlike others, the reference model, given its heuristic nature, has the advantage of being adaptable to changes in the input allowing for this dynamic adaptation. In fact, to determine the set of patterns to be recognized in the input data, the model uses a heuristic approximate algorithm, that is easier to implement and adapt to changing situations compared to its exact version (Del Viva et al., 2017). It has been shown that a more efficient and exact version of the same algorithm would be less biologically-plausible and would require more complex computations that would change completely the structure of the underlying network in case of changing external and/or internal conditions. Besides, preliminary experiments show that the heuristic solution is a better predictor of human image discrimination than the exact one (Del Viva et al., 2017). At this point, it would be interesting to further investigate learning processes testing how they could be framed within this model approach.

Coming back to the main results, our findings indicate that the saliency map provided by the model-predicted *optimal* features is used to automatically guide attention and eye movements toward informative locations and is comparable to that of luminance-based saliency. They also suggest that these salient features participate in the visual reconstruction process.

To conclude, the findings presented in this thesis suggest that visual saliency may be derived naturally in a system that, under the pressure of fast visual analysis, operates maximum information transmission under computational limitation constraints, as predicted by the reference model (Del Viva et al., 2013). On this basis, we speculate that active vision is efficiently adapted to maximize information in natural visual scenes under specific processing constraints.

BIBLIOGRAPHY

Ahmadlou, M., & Heimel, J. A. (2015). Preference for concentric orientations in the mouse superior colliculus. *Nature Communications*. https://doi.org/10.1038/ncomms7773

Aizawa, H., & Wurtz, R. H. (1998). Reversible inactivation of monkey superior colliculus. I. Curvature of saccadic trajectory. *Journal of Neurophysiology*, *79*(4). https://doi.org/10.1152/jn.1998.79.4.2082

Atick, J. J. (1992). Could information theory provide an ecological theory of sensory processing? In *Network: Computation in Neural Systems*. https://doi.org/10.1088/0954-898X_3_2_009

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*. https://doi.org/10.1037/h0054663

Attwell, D., & Laughlin, S. B. (2001). An energy budget for signaling in the grey matter of the brain. In *Journal of Cerebral Blood Flow and Metabolism*. https://doi.org/10.1097/00004647-200110000-00001

Barlow, H. (1961). Possible Principles Underlying the Transformations of Sensory Messages. *Sensory Communication*, *1*. https://doi.org/10.7551/mitpress/9780262518420.003.0013

Baylis, G. C., & Driver, J. (2001). Shape-coding in IT cells generalizes over contrast and mirror reversal, but not figure-ground reversal. *Nature Neuroscience*, *4*(9). https://doi.org/10.1038/nn0901-937

Baylor, D. A. (1987). Photoreceptor signals and vision: Proctor lecture. In *Investigative Ophthalmology and Visual Science* (Vol. 28, Issue 1).

Benson, V. (2008). A comparison of bilateral versus unilateral target and distractor presentation in the remote distractor paradigm. *Experimental Psychology*. https://doi.org/10.1027/1618-3169.55.5.334

Bergen, J. R., & Julesz, B. (1983). Parallel versus serial processing in rapid pattern discrimination. *Nature*, *303*(5919). https://doi.org/10.1038/303696a0

Bisley, J. W., & Goldberg, M. E. (2010). Attention, intention, and priority in the parietal lobe. In *Annual Review of Neuroscience*. https://doi.org/10.1146/annurev-neuro-060909-152823

Blakemore, B. C., & Campbell, F. W. (1969). *By C. BLAKEMORE AND F. W. CAMPBELL From the*. 237–260.

Boeschoten, M. A., Kemner, C., Kenemans, J. L., & Van Engeland, H. (2005). The relationship between local and global processing and the processing of high and low spatial frequencies studied by event-related potentials and source modeling. *Cognitive Brain Research*, *24*(2). https://doi.org/10.1016/j.cogbrainres.2005.01.021

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(38). https://doi.org/10.1073/pnas.0803390105

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*. https://doi.org/10.1163/156856897X00357

Briand, K. A., Larrison, A. L., & Sereno, A. B. (2000). Inhibition of return in manual and saccadic response systems. *Perception and Psychophysics*, *62*(8). https://doi.org/10.3758/BF03212152

Brown, J. M., & Koch, C. (2000). Influences of occlusion, color, and luminance on the perception of fragmented pictures. *Perceptual and Motor Skills*, *90*(3). https://doi.org/10.2466/pms.2000.90.3.1033

Bruce, N. D. B., & Tsotsos, J. K. (2005). Saliency based on information maximization. In *Advances in Neural Information Processing Systems* (pp. 155–162).

Bruno, N., Bertamini, M., & Domini, F. (1997). Amodal Completion of Partly Occluded Surfaces: Is There a Mosaic Stage? *Journal of Experimental Psychology: Human Perception and Performance*, *23*(5), 1412–1426. https://doi.org/10.1037/0096-1523.23.5.1412

Buonocore, A., & Hafed, Z. M. (2021). A sensory race between oculomotor control areas for coordinating motor timing. *Journal of Vision*, *21*(9). https://doi.org/10.1167/jov.21.9.2420

Callaway, E. M. (2005). Structure and function of parallel pathways in the primate early visual system. In *Journal of Physiology* (Vol. 566, Issue 1). https://doi.org/10.1113/jphysiol.2005.088047

Calvert, G. A. (2001). Crossmodal processing in the human brain: Insights from functional neuroimaging studies. In *Cerebral Cortex* (Vol. 11, Issue 12). https://doi.org/10.1093/cercor/11.12.1110

Carpenter, R. H. S. (1988). *Movements of the Eyes, 2nd Rev*. Pion Limited.

Carrasco, M. (2006). Chapter 3 Covert attention increases contrast sensitivity: psychophysical, neurophysiological and neuroimaging studies. *Progress in Brain Research*, *154*(SUPPL. A), 33–70. https://doi.org/10.1016/S0079-6123(06)54003-8

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, *51*(13), 1484–1525. https://doi.org/10.1016/j.visres.2011.04.012

Casteau, S., & Vitu-Thibault, F. (2012). On the effect of remote and proximal distractors on saccadic behavior: A challenge to neural-field models. *Journal of Vision*, *12*(12). https://doi.org/10.1167/12.12.14

Castellotti, S., D'Agostino, O., & Del Viva, M. M. (2023b). Fast discrimination of fragmentary images: the role of local optimal information. *Frontiers in Human Neuroscience*, *17*. https://doi.org/10.3389/fnhum.2023.1049615

Castellotti, S., Montagnini, A., & Del Viva, M. M. (2021). Early Visual Saliency Based on Isolated Optimal Features. *Frontiers in Neuroscience*, *15*. https://doi.org/10.3389/fnins.2021.645743

Castellotti, S., Montagnini, A., & Del Viva, M. M. (2022). Information-optimal local features automatically attract covert and overt attention. *Scientific Reports*, *12*(1), 9994. https://doi.org/10.1038/s41598-022-14262-2

Castellotti, S., Szinte, M., Del Viva, M. M., & Montagnini, A. (2023a). Saccadic trajectories deviate toward or away from optimally informative visual features. *IScience*, 107282. https://doi.org/https://doi.org/10.1016/j.isci.2023.107282

Cavaletti, G. A., & Marmiroli, P. L. (2009). *Elementi di anatomia ed istologia oculare*. Aracne. https://books.google.it/books?id=_0IHQgAACAAJ

Cerkevich, C. M., Lyon, D. C., Balaram, P., & Kaas, J. H. (2014). Distribution of cortical neurons projecting to the superior colliculus in macaque monkeys. *Eye and Brain*, *6*. https://doi.org/10.2147/EB.S53613

Chalupa, L. M., & Rhoades, R. W. (1977). Responses of visual, somatosensory, and auditory neurones in the golden hamster's superior colliculus. *The Journal of Physiology*, *270*(3). https://doi.org/10.1113/jphysiol.1977.sp011971

Chica, A. B., Martín-Arévalo, E., Botta, F., & Lupiáñez, J. (2014). The Spatial Orienting paradigm: How to design and interpret spatial attention experiments. In *Neuroscience and Biobehavioral Reviews* (Vol. 40). https://doi.org/10.1016/j.neubiorev.2014.01.002

Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A., & Oliva, A. (2016). Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Scientific Reports*, *6*. https://doi.org/10.1038/srep27755

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience*, *17*(3). https://doi.org/10.1038/nn.3635

Coëffé, C., & O'regan, J. K. (1987). Reducing the influence of non-target stimuli on saccade accuracy: Predictability and latency effects. *Vision Research*, *27*(2). https://doi.org/10.1016/0042-6989(87)90185-4

Coren, S., & Hoenig, P. (1972). Effect of Non-Target Stimuli upon Length of Voluntary Saccades. *Perceptual and Motor Skills*, *34*(2), 499–508. https://doi.org/10.2466/pms.1972.34.2.499

Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, and Computers*. https://doi.org/10.3758/BF03195489

Cowey, A., & Rolls, E. T. (1974). Human cortical magnification factor and its relation to visual acuity. *Experimental Brain Research*, *21*(5). https://doi.org/10.1007/BF00237163

Damasse, J. B., Perrinet, L. U., Madelain, L., & Montagnini, A. (2018). Reinforcement effects in anticipatory smooth eye movements. *Journal of Vision*. https://doi.org/10.1167/18.11.14

Daniel, P. M., & Whitteridge, D. (1961). The representation of the visual field on the cerebral cortex in monkeys. *The Journal of Physiology*, *159*(2). https://doi.org/10.1113/jphysiol.1961.sp006803

Danziger, S., & Kingstone, A. (1999). Unmasking the inhibition of return phenomenon. *Perception and Psychophysics*, *61*(6). https://doi.org/10.3758/BF03207610

Davis, A. R., Sloper, J. J., Neveu, M. M., Hogg, C. R., Morgan, M. J., & Holder, G. E. (2006). Differential changes of magnocellular and parvocellular visual function in early- and late-onset strabismic amblyopia. *Investigative Ophthalmology and Visual Science*, *47*(11). https://doi.org/10.1167/iovs.06-0382

De Franceschi, G., & Solomon, S. G. (2018). Visual response properties of neurons in the superficial layers of the superior colliculus of awake mouse. *Journal of Physiology*. https://doi.org/10.1113/JP276964

DeAngelis, G. C., Ohzawa, I., & Freeman, R. D. (1995). Receptive-field dynamics in the central visual pathways. In *Trends in Neurosciences*. https://doi.org/10.1016/0166-2236(95)94496-R

Del Viva, M. M., Punzi, G., & Benedetti, D. (2013). Information and Perception of Meaningful Patterns. *PLoS ONE*, *8*(7). https://doi.org/10.1371/journal.pone.0069154

Del Viva, M. M., Punzi, G., & Shevell, S. K. (2016). Chromatic information and feature detection in fast visual analysis. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0159898

Del Viva, M. M., Budinich, R., Palmieri, L., Georgiev, V. S., & Punzi, G. (2017). Role of the cost of plasticity in determining the features of fast vision in humans. *MODVIS 2017 Computational and Mathematical Models in Vision (Abstract).*, 4.

Del Viva, M.M., Castellotti, S., D'Agostino, O., & Montagnini, A. (2022). Which salient features attract our gaze in fast vision of natural scenes? *Journal of Vision*, *22*(14), 3934. https://doi.org/10.1167/jov.22.14.3934

Delorme, A., Richard, G., & Fabre-Thorpe, M. (2000). Ultra-rapid categorisation of natural scenes does not rely on colour cues: A study in monkeys and humans. *Vision Research*, *40*(16). https://doi.org/10.1016/S0042-6989(00)00083-3

Derrington, A. M., & Webb, B. S. (2004). Visual System: How Is the Retina Wired up to the Cortex? In *Current Biology*. https://doi.org/10.1016/j.cub.2003.12.014

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, *36*(12). https://doi.org/10.1016/0042-6989(95)00294-4

Deuble, H., Wolf, W., & Hauske, G. (1984). The Evaluation of the Oculomotor Error Signal. *Advances in Psychology*, *22*(C). https://doi.org/10.1016/S0166-4115(08)61818-X

DeYoe, E. A., & Van Essen, D. C. (1988). Concurrent processing streams in monkey visual cortex. In *Trends in Neurosciences* (Vol. 11, Issue 5). https://doi.org/10.1016/0166-2236(88)90130-0

Donk, M., & van Zoest, W. (2008). Effects of Salience Are Short-Lived. *Psychological Science*. https://doi.org/10.1111/j.1467-9280.2008.02149.x

Douglas, R. J., & Martin, K. A. C. (2007). Recurrent neuronal circuits in the neocortex. In *Current Biology* (Vol. 17, Issue 13). https://doi.org/10.1016/j.cub.2007.04.024

Doyle, M., & Walker, R. (2001). Curved saccade trajectories: Voluntary and reflexive saccades curve away from irrelevant distractors. *Experimental Brain Research*, *139*(3). https://doi.org/10.1007/s002210100742

Drager, U. C., & Hubel, D. H. (1975). Responses to visual stimulation and relationship between visual, auditory, and somatosensory inputs in mouse superior colliculus. *Journal of Neurophysiology*. https://doi.org/10.1152/jn.1975.38.3.690

Echeverri, E. (2006). Limits of capacity for the exchange of information in the human nervous system. *IEEE Transactions on Information Technology in Biomedicine*. https://doi.org/10.1109/TITB.2006.879585

Eickenberg, M., Gramfort, A., Varoquaux, G., & Thirion, B. (2017). Seeing it all: Convolutional network layers map the function of the human visual system. *NeuroImage*, *152*. https://doi.org/10.1016/j.neuroimage.2016.10.001

Enns, J. T., & Lollo, V. Di. (2000). What's new in visual masking? In *Trends in Cognitive Sciences* (Vol. 4, Issue 9). https://doi.org/10.1016/S1364-6613(00)01520-5

Fecteau, J. H., & Munoz, D. P. (2006). Salience, relevance, and firing: a priority map for target selection. *Trends in Cognitive Sciences*, *10*(8), 382–390. https://doi.org/10.1016/j.tics.2006.06.011

Feinberg, E. H., & Meister, M. (2015). Orientation columns in the mouse superior colliculus. *Nature*. https://doi.org/10.1038/nature14103

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*(1). https://doi.org/10.1093/cercor/1.1.1

Ferster, D., & Lindström, S. (1983). An intracellular analysis of geniculo-cortical connectivity in area 17 of the cat. *The Journal of Physiology*, *342*(1).

https://doi.org/10.1113/jphysiol.1983.sp014846

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, *4*(12). https://doi.org/10.1364/josaa.4.002379

Findlay, J. M. (1982). Global visual processing for saccadic eye movements. *Vision Research*, *22*(8), 1033–1045. https://doi.org/https://doi.org/10.1016/0042-6989(82)90040-2

Finlay, B. L., Schneps, S. E., Wilson, K. G., & Schneider, G. E. (1978). Topography of visual and somatosensory projections to the superior colliculus of the golden hamster. *Brain Research*, *142*(2). https://doi.org/10.1016/0006-8993(78)90632-7

Fischer, B., & Ramsperger, E. (1984). Human express saccades: extremely short reaction times of goal directed eye movements. *Experimental Brain Research*, *57*(1). https://doi.org/10.1007/BF00231145

Gale, S. D., & Murphy, G. J. (2014). Distinct representation and distribution of visual information by specific cell types in mouse superficial superior colliculus. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.2768-14.2014

Garcia-Diaz, A., Fdez-Vidal, X. R., Pardo, X. M., & Dosil, R. (2012). Saliency from hierarchical adaptation through decorrelation and variance normalization. *Image and Vision Computing*. https://doi.org/10.1016/j.imavis.2011.11.007

Garey, L. J., & Powell, T. P. (1971). An experimental study of the termination of the lateral geniculo-cortical pathway in the cat and monkey. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *179*(54). https://doi.org/10.1098/rspb.1971.0080

Gegenfurtner, K. R., & Rieger, J. (2000). Sensory and cognitive contributions of color to the recognition of natural scenes. *Current Biology*. https://doi.org/10.1016/S0960-9822(00)00563-7

Geisler, W. S., Perry, J. S., Super, B. J., & Gallogly, D. P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research*, *41*(6). https://doi.org/10.1016/S0042-6989(00)00277-7

Giordano, A. M., McElree, B., & Carrasco, M. (2009). On the automaticity and flexibility of covert attention: A speed-accuracy trade-off analysis. *Journal of Vision*, *9*(3). https://doi.org/10.1167/9.3.30

Glimcher, P. W. (2001). Making choices: The neurophysiology of visual-saccadic decision making. In *Trends in Neurosciences* (Vol. 24, Issue 11). https://doi.org/10.1016/S0166-2236(00)01932-9

Godijn, R., & Theeuwes, J. (2002). Programming of Endogenous and Exogenous Saccades: Evidence for a Competitive Integration Model. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(5). https://doi.org/10.1037/0096-1523.28.5.1039

Godijn, R., & Theeuwes, J. (2004). The relationship between inhibition of return and saccade trajectory deviations. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(3). https://doi.org/10.1037/0096-1523.30.3.538

Goffart, L., Hafed, Z. M., & Krauzlis, R. J. (2012). Visual fixation as equilibrium: Evidence from superior colliculus inactivation. *Journal of Neuroscience*, *32*(31). https://doi.org/10.1523/JNEUROSCI.0696-12.2012

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. In

Trends in Neurosciences (Vol. 15, Issue 1). https://doi.org/10.1016/0166-2236(92)90344-8

Greene, C. M., Murphy, G., & Januszewski, J. (2017). Under High Perceptual Load, Observers Look but Do Not See. *Applied Cognitive Psychology*, *31*(4), 431–437. https://doi.org/10.1002/acp.3335

Grill-Spector, K., Kushnir, T., Hendler, T., & Malach, R. (2000). The dynamics of object-selective activation correlate with recognition performance in humans. *Nature Neuroscience*. https://doi.org/10.1038/77754

Güçlü, U., & van Gerven, M. A. J. (2015). Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *Journal of Neuroscience*, *35*(27). https://doi.org/10.1523/JNEUROSCI.5023-14.2015

Hall, W. C., & Moschovakis, A. (2003). The superior colliculus: New approaches for studying sensorimotor integration. In *The Superior Colliculus: New Approaches for Studying Sensorimotor Integration*.

Hare, R. D. (1973). Orienting and Defensive Responses to Visual Stimuli. *Psychophysiology*. https://doi.org/10.1111/j.1469-8986.1973.tb00532.x

Harutiunian Kozak, B., Dec, K., & Wrobel, A. (1973). The organization of visual receptive fields of neurons in the cat colliculus superior. *Acta Neurobiologiae Experimentalis*.

Heeman, J., Nijboer, T. C. W., Van der Stoep, N., Theeuwes, J., & Van der Stigchel, S. (2016). Oculomotor interference of bimodal distractors. *Vision Research*, *123*. https://doi.org/10.1016/j.visres.2016.04.002

Herzog, M. (2016). Bachmann, T., & Francis, G. Visual masking: Studying perception, attention, and consciousness. *Perception*, *45*(6). https://doi.org/10.1177/0301006615623413

Hickey, C., & Van Zoest, W. (2012). Reward creates oculomotor salience. In *Current Biology* (Vol. 22, Issue 7). https://doi.org/10.1016/j.cub.2012.02.007

Hikosaka, O., Miyauchi, S., & Shimojo, S. (1996). Orienting of spatial attention - Its reflexive, compensatory, and voluntary mechanisms. *Cognitive Brain Research*, *5*(1–2). https://doi.org/10.1016/S0926-6410(96)00036-5

Holmes, G. (1918). DISTURBANCES OF VISION BY CEREBRAL LESIONS. *British Journal of Ophthalmology*, *2*(7). https://doi.org/10.1136/bjo.2.7.353

Honda, H., & Findlay, J. M. (1992). Saccades to targets in three-dimensional space: Dependence of saccadic latency on target location. *Perception & Psychophysics*. https://doi.org/10.3758/BF03206770

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, *79*(8). https://doi.org/10.1073/pnas.79.8.2554

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*, *148*(3). https://doi.org/10.1113/jphysiol.1959.sp006308

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*. https://doi.org/10.1113/jphysiol.1962.sp006837

Hubel, D. H., & Wiesel, T. N. (1963). Shape and arrangement of columns in cat's striate cortex. *The Journal of Physiology*, *165*(3). https://doi.org/10.1113/jphysiol.1963.sp007079

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, *195*(1). https://doi.org/10.1113/jphysiol.1968.sp008455

Hubel, D. H., & Wiesel, T. N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *Journal of Neurophysiology*, *28*. https://doi.org/10.1152/jn.1965.28.2.229

Hubel, D. H. (1982). Evolution of ideas on the primary visual cortex, 1955-1978: A biased historical account - Nobel lecture, 8 December 1981. *Bioscience Reports*, *2*(7). https://doi.org/10.1007/BF01115245

Hubel, D. H., & Wiesel, T. N. (1965). Receptive archi- tecture in two nonstriate visual areas (18 and 19) of the cati. *Journal of Neurophysiology*, *28*(2), 229–289.

Hubel, D. H., & Wiesel, T. N. (1974). Uniformity of monkey striate cortex: A parallel relationship between field size, scatter, and magnification factor. *Journal of Comparative Neurology*. https://doi.org/10.1002/cne.901580305

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(11), 1254–1259. https://doi.org/10.1109/34.730558

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research*. https://doi.org/10.1016/j.visres.2008.09.007

Itti, L., & Borji, A. (2013). Computational models: Bottom-up and top-down aspects. In *The Oxford Handbook of Attention* (A. C. Nobr, pp. 1–20). http://arxiv.org/abs/1510.07748

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*. https://doi.org/10.1016/S0042-6989(99)00163-7

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*. https://doi.org/10.1038/35058500

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(11), 1254–1259. https://doi.org/10.1109/34.730558

Jäkel, F., & Wichmann, F. A. (2006). Spatial four-alternative forced-choice method is the preferred psychophysical method for naïve observers. *Journal of Vision*, *6*(11), 1307–1322. https://doi.org/10.1167/6.11.13

Jay, M. F., & Sparks, D. L. (1987). Sensorimotor integration in the primate superior colliculus. II. Coordinates of auditory signals. *Journal of Neurophysiology*, *57*(1). https://doi.org/10.1152/jn.1987.57.1.35

Jay, M. F., & Sparks, D. L. (1984). Auditory receptive fields in primate superior colliculus shift with changes in eye position. *Nature*, *309*(5966). https://doi.org/10.1038/309345a0

Johnson, J. S., & Olshausen, B. A. (2005). The recognition of partially visible natural objects in the presence and absence of their occluders. *Vision Research*, *45*(25–26), 3262–3276. https://doi.org/10.1016/j.visres.2005.06.007

Jonides, J. (1998). Voluntary versus automatic control over the mind's eye's movement. *Psychonomic Bulletin & Review*. https://doi.org/10.1037/0096-1523.29.5.835

Jonikaitis, D., & Belopolsky, A. V. (2014). Target-distractor competition in the oculomotor system is spatiotopic. *Journal of Neuroscience*, *34*(19). https://doi.org/10.1523/JNEUROSCI.4453-

13.2014

Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S. A., & Hudspeth, A. J. (2013). Principles of Neural Science, Fifth Editon | Neurology Collection | McGraw-Hill Medical. In *2013*.

Kaplan, E., Lee, B. B., & Shapley, R. M. (1990). Chapter 7 New views of primate retinal function. In *Progress in Retinal Research* (Vol. 9, Issue C). https://doi.org/10.1016/0278-4327(90)90009-7

Kara, P., Pezaris, J. S., Yurgenson, S., & Reid, R. C. (2002). The spatial receptive field of thalamic inputs to single cortical simple cells revealed by the interaction of visual and electrical stimulation. *Proceedings of the National Academy of Sciences*, *99*(25), 16261 LP – 16266. https://doi.org/10.1073/pnas.242625499

Kauffmann, L., Ramanoël, S., & Peyrin, C. (2014). The neural bases of spatial frequency processing during scene perception. In *Frontiers in Integrative Neuroscience* (Vol. 8, Issue MAY). https://doi.org/10.3389/fnint.2014.00037

Kean, M., & Lambert, A. (2003). The influence of a salience distinction between bilateral cues on the latency of target-detection saccades. *British Journal of Psychology*, *94*(3), 373–388. https://doi.org/10.1348/000712603767876280

Kehoe, D. H., Aybulut, S., & Fallah, M. (2018). Higher order, multifeatural object encoding by the oculomotor system. *Journal of Neurophysiology*, *120*(6). https://doi.org/10.1152/jn.00834.2017

Kehoe, D. H., & Fallah, M. (2017). Rapid accumulation of inhibition accounts for saccades curved away from distractors. *Journal of Neurophysiology*, *118*(2). https://doi.org/10.1152/jn.00742.2016

Kehoe, D. H., Lewis, J., & Fallah, M. (2021). Oculomotor target selection is mediated by complex objects. *Journal of Neurophysiology*, *126*(3). https://doi.org/10.1152/jn.00580.2020

Kehoe, D. H., Rahimi, M., & Fallah, M. (2018). Perceptual color space representations in the oculomotor system are modulated by surround suppression and biased selection. *Frontiers in Systems Neuroscience*, *12*. https://doi.org/10.3389/fnsys.2018.00001

Kim, U. S., Mahroo, O. A., Mollon, J. D., & Yu-Wai-Man, P. (2021). Retinal Ganglion Cells—Diversity of Cell Types and Clinical Relevance. In *Frontiers in Neurology* (Vol. 12). https://doi.org/10.3389/fneur.2021.661938

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research*. https://doi.org/10.1016/j.visres.2005.10.002

Kleiner, M., Brainard, D. H., Pelli, D. G., Broussard, C., Wolf, T., & Niehorster, D. (2007). What's new in Psychtoolbox-3? *Perception*.

Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research*, *35*(13). https://doi.org/10.1016/0042-6989(94)00279-U

Krauzlis, R. J., & Miles, F. A. (1996). Initiation of saccades during fixation or pursuit: Evidence in humans for a single mechanism. *Journal of Neurophysiology*, *76*(6). https://doi.org/10.1152/jn.1996.76.6.4175

Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology*, *16*(1). https://doi.org/10.1152/jn.1953.16.1.37

Lambert, A., Spencer, E., & Mohindra, N. (1987). Automaticity and the capture of attention by a peripheral display change. *Current Psychology*, *6*(2). https://doi.org/10.1007/BF02686618

Larson, A. M., & Loschky, L. C. (2009). The contributions of central versus peripheral vision to scene gist recognition. *Journal of Vision*, *9*(10). https://doi.org/10.1167/9.10.6

Lavie, N., Beck, D. M., & Konstantinou, N. (2014). Blinded by the load: Attention, awareness and the role of perceptual load. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1641). https://doi.org/10.1098/rstb.2013.0205

LeDoux, J. (1996). *The Emotional Brain* (Simon and Schuster (ed.)).

Lee, C., Rohrer, W. H., & Sparks, D. L. (1988). Population coding of saccadic eye movements by neurons in the superior colliculus. *Nature*, *332*(6162). https://doi.org/10.1038/332357a0

Lennie, P. (1998). Single units and visual cortical organization. *Perception*. https://doi.org/10.1068/p270889

Lennie, P. (2003). The cost of cortical computation. *Current Biology*. https://doi.org/10.1016/S0960-9822(03)00135-0

Levy, W. B., & Baxter, R. A. (1996). Energy Efficient Neural Codes. *Neural Computation*, *8*(3). https://doi.org/10.1162/neco.1996.8.3.531

Li, Z. (2002). A saliency map in primary visual cortex. In *Trends in Cognitive Sciences*. https://doi.org/10.1016/S1364-6613(00)01817-9

Lock, T. M., Baizer, J. S., & Bender, D. B. (2003). Distribution of corticotectal cells in macaque. *Experimental Brain Research*, *151*(4). https://doi.org/10.1007/s00221-003-1500-y

Ludwig, C. J. H., & Gilchrist, I. D. (2003). Target similarity affects saccade curvature away from irrelevant onsets. *Experimental Brain Research*, *152*(1). https://doi.org/10.1007/s00221-003-1520-7

Ludwig, C. J. H., Gilchrist, I. D., & McSorley, E. (2004). The influence of spatial frequency and contrast on saccade latencies. *Vision Research*. https://doi.org/10.1016/j.visres.2004.05.022

Luo, D. G., Xue, T., & Yau, K. W. (2008). How vision begins: An odyssey. In *Proceedings of the National Academy of Sciences of the United States of America* (Vol. 105, Issue 29). https://doi.org/10.1073/pnas.0708405105

Macknik, S. L., & Martinez-Conde, S. (2004). Dichoptic visual masking reveals that early binocular neurons exhibit weak interocular suppression: Implications for binocular vision and visual awareness. *Journal of Cognitive Neuroscience*, *16*(6). https://doi.org/10.1162/0898929041502788

Marr, D. (1982). Vision: a computational investigation into the human representation and processing of visual information. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. https://doi.org/10.1016/0022-2496(83)90030-5

Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London - Biological Sciences*. https://doi.org/10.1098/rspb.1980.0020

Maunsell, J. H. R., & Newsome, W. T. (1987). Visual processing in monkey extrastriate cortex. *Annual Review of Neuroscience*, *Vol. 10*. https://doi.org/10.1146/annurev.ne.10.030187.002051

Maunsell, J. H. R., Ghose, G. M., Assad, J. A., Mcadams, C. J., Boudreau, C. E., & Noerager, B. D. (1999). Visual response latencies of magnocellular and parvocellular LGN neurons in macaque monkeys. *Visual Neuroscience*, *16*(1). https://doi.org/10.1017/S0952523899156177

Mayer, M. J., & Dowling, J. E. (1988). The Retina: An Approachable Part of the Brain. *The American Journal of Psychology*, *101*(4). https://doi.org/10.2307/1423238

Mcilwain, J. T. (1991). Distributed spatial coding in the superior colliculus: A review. *Visual Neuroscience*, *6*(1). https://doi.org/10.1017/S0952523800000857

McPeek, R. M. (2006). Incomplete suppression of distractor-related activity in the frontal eye field results in curved saccades. *Journal of Neurophysiology*, *96*(5). https://doi.org/10.1152/jn.00564.2006

McPeek, R. M., Han, J. H., & Keller, E. L. (2003). Competition between saccade goals in the superior colliculus produces saccade curvature. *Journal of Neurophysiology*, *89*(5). https://doi.org/10.1152/jn.00657.2002

McSorley, E., Haggard, P., & Walker, R. (2006). Time course of oculomotor inhibition revealed by saccade trajectory modulation. *Journal of Neurophysiology*, *96*(3). https://doi.org/10.1152/jn.00315.2006

Meeter, M., Van der Stigchel, S., & Theeuwes, J. (2010). A competitive integration model of exogenous and endogenous eye movements. *Biological Cybernetics*, *102*(4), 271–291. https://doi.org/10.1007/s00422-010-0365-y

Meng, M., & Potter, M. C. (2008). Detecting and remembering pictures with and without visual noise. *Journal of Vision*, *8*(9), 1–10. https://doi.org/10.1167/8.9.7

Montagnini, A., & Castet, E. (2007). Spatiotemporal dynamics of visual attention during saccade preparation: Independence and coupling between attention and movement planning. *Journal of Vision*, *7*(14). https://doi.org/10.1167/7.14.8

Morgan, M. J. (2011). Features and the “primal sketch.” In *Vision Research*. https://doi.org/10.1016/j.visres.2010.08.002

Morris, J. S., Öhman, A., & Dolan, R. J. (1999). A subcortical pathway to the right amygdala mediating “unseen” fear. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.96.4.1680

Morrone, M. C., & Burr, D. C. (1988). Feature detection in human vision: a phase-dependent energy model. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character. Royal Society (Great Britain)*. https://doi.org/10.1098/rspb.1988.0073

Movshon, J. A., Thompson, I. D., & Tolhurst, D. J. (1978a). Receptive field organization of complex cells in the cat’s striate cortex. *The Journal of Physiology*, *283*(1). https://doi.org/10.1113/jphysiol.1978.sp012489

Movshon, J. A., Thompson, I. D., & Tolhurst, D. J. (1978b). Spatial summation in the receptive fields of simple cells in the cat’s striate cortex. *The Journal of Physiology*, *283*(1). https://doi.org/10.1113/jphysiol.1978.sp012488

Müller, H. J., & Rabbitt, P. M. A. (1989). Reflexive and Voluntary Orienting of Visual Attention: Time Course of Activation and Resistance to Interruption. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(2). https://doi.org/10.1037/0096-1523.15.2.315

Munoz, D. P., & Everling, S. (2004). Look away: The anti-saccade task and the voluntary control of eye movement. In *Nature Reviews Neuroscience* (Vol. 5, Issue 3). https://doi.org/10.1038/nrn1345

Murray, R. F., Sekuler, A. B., & Bennett, P. J. (2001). Time course of amodal completion revealed by a shape discrimination task. *Psychonomic Bulletin and Review*, *8*(4). https://doi.org/10.3758/BF03196208

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*.

https://doi.org/10.1038/nature03390

Najemnik, J., & Geisler, W. S. (2008). Eye movement statistics in humans are consistent with an optimal search strategy. *Journal of Vision*. https://doi.org/10.1167/8.3.4

Nakayama, K., & Mackeben, M. (1989). Sustained and transient components of focal visual attention. *Vision Research*, *29*(11). https://doi.org/10.1016/0042-6989(89)90144-2

Nassi, J. J., & Callaway, E. M. (2009). Parallel processing strategies of the primate visual system. In *Nature Reviews Neuroscience* (Vol. 10, Issue 5). https://doi.org/10.1038/nrn2619

Nejati, V. (2021). Effect of stimulus dimension on perception and cognition. *Acta Psychologica*, *212*, 103208. https://doi.org/10.1016/j.actpsy.2020.103208

Niell, C. M., & Stryker, M. P. (2008). Highly selective receptive fields in mouse visual cortex. *Journal of Neuroscience*, *28*(30). https://doi.org/10.1523/JNEUROSCI.0623-08.2008

Nothdurft, H.C. (2000). Salience from feature contrast: additivity across dimensions. *Vision Research*, *40*(10), 1183–1201. https://doi.org/https://doi.org/10.1016/S0042-6989(00)00031-6

Nothdurft, H. C. (1993a). The Conspicuousness of Orientation and Motion Contrast. *Spatial Vision*. https://doi.org/10.1163/156856893X00487

Nothdurft, H. C. (1993b). The role of features in preattentive vision: Comparison of orientation, motion and color cues. *Vision Research*. https://doi.org/10.1016/0042-6989(93)90020-W

Nothdurft, H. C. (2002). Attention shifts to salient targets. *Vision Research*. https://doi.org/10.1016/S0042-6989(02)00016-0

Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/0096-3445.130.3.466

Oliva, A. (2005). Gist of the scene. In *Neurobiology of Attention*. https://doi.org/10.1016/B978-012375731-9/50045-8

Oliva, A., & Schyns, P. G. (1997a). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology*. https://doi.org/10.1006/cogp.1997.0667

Oliva, A., & Schyns, P. G. (1997b). Coarse Blobs or Fine Edges? Evidence That Information DiagnosticityC hanges the Perception of Complex Visual Stimuliings of their distinctive properties. In the object recognition literature, scene. *Cognitive Psychology*, *34*(1), 72–107.

Oliva, A., & Schyns, P. G. (2000). Diagnostic Colors Mediate Scene Recognition. *Cognitive Psychology*. https://doi.org/10.1006/cogp.1999.0728

Olmos, A., & Kingdom, F. A. A. (2004). A biologically inspired algorithm for the recovery of shading and reflectance images. *Perception*. https://doi.org/10.1068/p5321

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*. https://doi.org/10.1038/381607a0

Olshausen, B. A., & Field, D. J. (2004). Sparse coding of sensory inputs. In *Current Opinion in Neurobiology* (Vol. 14, Issue 4). https://doi.org/10.1016/j.conb.2004.07.007

Østerberg, G. A. (1937). Topography of the Layer of Rods and Cones in the Human Retina. *Journal of the American Medical Association*, *108*(3), 232. https://doi.org/10.1001/jama.1937.02780030070033

Otero-Millan, J., Troncoso, X. G., Macknik, S. L., Serrano-Pedraza, I., & Martinez-Conde, S. (2008). Saccades and microsaccades during visual fixation, exploration, and search: Foundations for a common saccadic generator. *Journal of Vision*, *8*(14). https://doi.org/10.1167/8.14.21

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*. https://doi.org/10.1016/S0042-6989(01)00250-4

Párraga, C. A., Brelstaff, G., Troscianko, T., & Moorehead, I. R. (1998). Color and luminance information in natural scenes. *Journal of the Optical Society of America A*, *15*(3). https://doi.org/10.1364/josaa.15.000563

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*. https://doi.org/10.1163/156856897X00366

Perrinet, L. U., & Bednar, J. A. (2015). Edge co-occurrences can account for rapid categorization of natural versus animal images. *Scientific Reports*. https://doi.org/10.1038/srep11400

Pinto, Y., van Gaal, S., de Lange, F. P., Lamme, V. A. F., & Seth, A. K. (2015). Expectations accelerate entry of visual stimuli into awareness. *Journal of Vision*, *15*(8), 1–15. https://doi.org/10.1167/15.8.13

Pokorny, J. (2011). Review: Steady and pulsed pedestals, the how and why of post-receptoral pathway separation. In *Journal of Vision* (Vol. 11, Issue 5). https://doi.org/10.1167/11.5.1

Posner, M. I., Cohen, Y., & Rafal, R. D. (1982). Neural systems control of spatial orienting. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *298*(1089). https://doi.org/10.1098/rstb.1982.0081

Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. https://doi.org/10.1080/00335558008248231

Qian, N. (1997). Binocular disparity and the perception of depth. In *Neuron* (Vol. 18, Issue 3). https://doi.org/10.1016/S0896-6273(00)81238-6

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. In *Neural Computation*. https://doi.org/10.1162/neco.2008.12-06-420

Ristori, L., & Punzi, G. (2010). Triggering on heavy flavors at hadron colliders. *Annual Review of Nuclear and Particle Science*, *60*. https://doi.org/10.1146/annurev.nucl.012809.104501

Rizzolatti, G., Riggio, L., Dascola, I., & Umiltá, C. (1987). Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention. *Neuropsychologia*, *25*(1, Part 1), 31–40. https://doi.org/https://doi.org/10.1016/0028-3932(87)90041-8

Rockel, A. J., Hiorns, R. W., & Powell, T. P. S. (1980). The basic uniformity in structure of the neocortex. *Brain*, *103*(2). https://doi.org/10.1093/brain/103.2.221

Rodieck, R. W. (1998). The first steps in seeing. In *The first steps in seeing.* Sinauer Associates.

Ross, L. E., & Ross, S. M. (1980). Saccade latency and warning signals: Stimulus onset, offset, and change as warning events. *Perception & Psychophysics*, *27*(3). https://doi.org/10.3758/BF03204262

Sadun, A. A., Johnson, B. M., & Smith, L. E. H. (1986). Neuroanatomy of the human visual system: Part II retinal projections to the superior colliculus and pulvinar. *Neuro-Ophthalmology*, *6*(6). https://doi.org/10.3109/01658108609016476

Schiller, P. H. (1984). The superior colliculus and visual function. In *Comprehensive Physiology*. https://doi.org/10.1002/cphy.cp010311

Scholl, B., Burge, J., & Priebe, N. J. (2013). Binocular integration and disparity selectivity in mouse primary visual cortex. *Journal of Neurophysiology*, *109*(12). https://doi.org/10.1152/jn.01021.2012

Schütz, A. C., Trommershäuser, J., & Gegenfurtner, K. R. (2012). Dynamic integration of information about salience and value for saccadic eye movements. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1115638109

Schyns, P. G., & Oliva, A. (1999). Dr. Angry and Mr. Smile: When categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition*, *69*(3). https://doi.org/10.1016/S0010-0277(98)00069-9

Serre, T., Kreiman, G., Kouh, M., Cadieu, C., Knoblich, U., & Poggio, T. (2007). A quantitative theory of immediate visual recognition. *Progress in Brain Research*, *165*, 33–56. https://doi.org/10.1016/S0079-6123(06)65004-8

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, *27*(4), 623–656. https://doi.org/10.1002/j.1538-7305.1948.tb00917.x

Shelepin, Y. E., Chikhman, V. N., & Foreman, N. (2009). Analysis of the studies of the perception of fragmented images: Global description and perception using local features. *Neuroscience and Behavioral Physiology*, *39*(6). https://doi.org/10.1007/s11055-009-9171-1

Sheliga, B. M., Riggio, L., & Rizzolatti, G. (1994). Orienting of attention and eye movements. *Experimental Brain Research*, *98*(3). https://doi.org/10.1007/BF00233988

Sheliga, B. M., Riggio, L., Craighero, L., & Rizzolatti, G. (1995). Spatial attention-determined modifications in saccade trajectories. *NeuroReport*, *6*(3). https://doi.org/10.1097/00001756-199502000-00044

Shepherd, G. M. (2004). The Synaptic Organization of the Brain. In *The Synaptic Organization of the Brain*. https://doi.org/10.1093/acprof:oso/9780195159561.001.1

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. In *Annual Review of Neuroscience* (Vol. 24). https://doi.org/10.1146/annurev.neuro.24.1.1193

Simoncelli, E. P. (2003). Vision and the statistics of the visual environment. In *Current Opinion in Neurobiology* (Vol. 13, Issue 2). https://doi.org/10.1016/S0959-4388(03)00047-3

Slotnick, S. D., Klein, S. A., Carney, T., & Sutter, E. E. (2001). Electrophysiological estimate of human cortical magnification. *Clinical Neurophysiology*, *112*(7). https://doi.org/10.1016/S1388-2457(01)00561-2

Smith, A. T., Singh, K. D., Williams, A. L., & Greenlee, M. W. (2001). Estimating receptive field size from fMRI data in human striate and extrastriate visual cortex. *Cerebral Cortex*, *11*(12). https://doi.org/10.1093/cercor/11.12.1182

Smith, E. C., & Lewicki, M. S. (2006). Efficient auditory coding. *Nature*, *439*(7079). https://doi.org/10.1038/nature04485

Spatz, W. B. (1979). The retino-geniculo-cortical pathway in Callithrix. II. The geniculo-cortical projection. *Experimental Brain Research*, *36*(3). https://doi.org/10.1007/BF00238512

Staugaard, C. F., Petersen, A., & Vangkilde, S. (2016). Eccentricity effects in vision and attention. *Neuropsychologia*, *92*, 69–78. https://doi.org/https://doi.org/10.1016/j.neuropsychologia.2016.06.020

Stein, T., & Peelen, M. V. (2015). Content-specific expectations enhance stimulus detectability by increasing perceptual sensitivity. *Journal of Experimental Psychology. General*, *144*(6), 1089–1104. https://doi.org/10.1037/xge0000109

Szinte, M., Aagten-Murphy, D., Jonikaitis, D., Wollenberg, L., & Deubel, H. (2020). Sounds are remapped across saccades. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-78163-y

Talgar, C. P., & Carrasco, M. (2002). Vertical meridian asymmetry in spatial resolution: Visual and attentional factors. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/BF03196326

Tang, H., Schrimpf, M., Lotter, W., Moerman, C., Paredes, A., Caro, J. O., Hardesty, W., Cox, D., & Kreiman, G. (2018). Recurrent computations for visual pattern completion. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(35). https://doi.org/10.1073/pnas.1719397115

Tatler, B. W., & Melcher, D. (2007). Pictures in mind: Initial encoding of object properties varies with the realism of the scene stimulus. *Perception*. https://doi.org/10.1068/p5592

Tepin, M. B., & Dark, V. J. (1992). Do Abrupt-onset Peripheral Cues Attract Attention Automatically? *The Quarterly Journal of Experimental Psychology Section A*, *45*(1). https://doi.org/10.1080/14640749208401318

Theeuwes, J. (1991). Exogenous and endogenous control of attention: The effect of visual onsets and offsets. *Perception & Psychophysics*, *49*(1). https://doi.org/10.3758/BF03211619

Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychologica*, *135*(2). https://doi.org/10.1016/j.actpsy.2010.02.006

Theeuwes, J., & Godijn, R. (2001). Attentional and oculomotor capture. In *Attraction, distraction and action: Multiple perspectives on attentional capture.* (pp. 121–149). Elsevier Science. https://doi.org/10.1016/S0166-4115(01)80008-X

Theeuwes, J., & Godijn, R. (2004). Inhibition-of-return and oculomotor interference. *Vision Research*, *44*(12). https://doi.org/10.1016/j.visres.2003.09.035

Theeuwes, J., & van der Burg, E. (2008). The role of cueing in attentional capture. *Visual Cognition*, *16*(2–3). https://doi.org/10.1080/13506280701462525

Thompson, K. G., & Bichot, N. P. (2005). A visual salience map in the primate frontal eye field. In *Progress in Brain Research*. https://doi.org/10.1016/S0079-6123(04)47019-8

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*(6582). https://doi.org/10.1038/381520a0

Tipper, S. P., Howard, L. A., & Paul, M. A. (2001). Reaching affects saccade trajectories. *Experimental Brain Research*, *136*(2). https://doi.org/10.1007/s002210000577

Tipper, S. P, Howard, L. A., & Jackson, S. R. (1997). Selective Reaching to Grasp: Evidence for Distractor Interference Effects. *Visual Cognition*, *4*(1), 1–38. https://doi.org/10.1080/713756749

Torralba, A. (2003). Contextual priming for object detection, {International Journal} of {Computer Vision}. *International Journal of Computer Vision*, *53*(2), 169–191.

Torralba, A. (2003). Modeling global scene factors in attention. *Journal of the Optical Society of America A*. https://doi.org/10.1364/josaa.20.001407

Trappenberg, T. P., Dorris, M. C., Munoz, D. P., & Klein, R. M. (2001). A model of saccade initiation based on the competitive integration of exogenous and endogenous signals in the superior colliculus. *Journal of Cognitive Neuroscience*, *13*(2). https://doi.org/10.1162/089892901564306

Treisman, A. (1985). Preattentive processing in vision. *Computer Vision, Graphics and Image Processing*. https://doi.org/10.1016/S0734-189X(85)80004-9

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1). https://doi.org/10.1016/0010-0285(80)90005-5

Tudge, L., Brandt, S. A., & Schubert, T. (2018). Salience from multiple feature contrast: Evidence from saccade trajectories. *Attention, Perception, and Psychophysics*, *80*(3). https://doi.org/10.3758/s13414-017-1480-9

Ullman, S., Assif, L., Fetaya, E., & Harari, D. (2016). Atoms of recognition in human and computer vision. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(10). https://doi.org/10.1073/pnas.1513198113

Valberg, A., & Lee, B. B. (1992). Main cell systems in primate visual pathways. In *Current Opinion in Ophthalmology* (Vol. 3, Issue 6). https://doi.org/10.1097/00055735-199212000-00015

Van der Stigchel, S., De Vries, J. P., Bethlehem, R., & Theeuwes, J. (2011). A global effect of capture saccades. *Experimental Brain Research*, *210*(1). https://doi.org/10.1007/s00221-011-2602-6

Van der Stigchel, S. (2010). Recent advances in the study of saccade trajectory deviations. In *Vision Research* (Vol. 50, Issue 17). https://doi.org/10.1016/j.visres.2010.05.028

Van der Stigchel, S., Meeter, M., & Theeuwes, J. (2006). Eye movement trajectories and what they tell us. In *Neuroscience and Biobehavioral Reviews* (Vol. 30, Issue 5). https://doi.org/10.1016/j.neubiorev.2005.12.001

Van der Stigchel, S., Meeter, M., & Theeuwes, J. (2007a). The spatial coding of the inhibition evoked by distractors. *Vision Research*, *47*(2). https://doi.org/10.1016/j.visres.2006.11.001

Van der Stigchel, S., Meeter, M., & Theeuwes, J. (2007b). Top-down influences make saccades deviate away: The case of endogenous cues. *Acta Psychologica*, *125*(3), 279–290. https://doi.org/https://doi.org/10.1016/j.actpsy.2006.08.002

Van der Stigchel, S., Mulckhuyse, M., & Theeuwes, J. (2009). Eye cannot see it: The interference of subliminal distractors on saccade metrics. *Vision Research*, *49*(16). https://doi.org/10.1016/j.visres.2009.05.018

van Leeuwen, J., Smeets, J. B. J., & Belopolsky, A. V. (2019). Forget binning and get SMART: Getting more out of the time-course of response data. *Attention, Perception, and Psychophysics*, *81*(8). https://doi.org/10.3758/s13414-019-01788-3

Van Zoest, W., Donk, M., & Theeuwes, J. (2004). The role of stimulus-driven and goal-driven control in saccadic visual selection. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(4). https://doi.org/10.1037/0096-1523.30.4.749

Van Zoest, W., Donk, M., & Van der Stigchel, S. (2012). Stimulus-salience and the time-course of saccade trajectory deviations. *Journal of Vision*, *12*(8). https://doi.org/10.1167/12.8.16

Veale, R., Hafed, Z. M., & Yoshida, M. (2017). How is visual salience computed in the brain? Insights from behaviour, neurobiology and modeling. In *Philosophical Transactions of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rstb.2016.0113

Walker, R., Kentridge, R. W., & Findlay, J. M. (1995). Independent contributions of the orienting of attention, fixation offset and bilateral stimulation on human saccadic latencies. *Experimental Brain Research*, *103*(2). https://doi.org/10.1007/BF00231716

Walker, R., Deubel, H., Schneider, W. X., & Findlay, J. M. (1997). Effect of remote distractors on saccade programming: Evidence for an extended fixation zone. *Journal of Neurophysiology*. https://doi.org/10.1152/jn.1997.78.2.1108

Walker, R., & McSorley, E. (2008). The Influence of Distractors on Saccade-Target Selection: Saccade Trajectory Effects. *Journal of Eye Movement Research*, *2*(3). https://doi.org/10.16910/jemr.2.3.7

Walker, R., McSorley, E., & Haggard, P. (2006). The control of saccade trajectories: Direction of curvature depends on prior knowledge of target location and saccade latency. *Perception and Psychophysics*, *68*(1). https://doi.org/10.3758/BF03193663

Wang, L., Sarnaik, R., Rangarajan, K., Liu, X., & Cang, J. (2010). Visual receptive field properties of neurons in the superficial superior colliculus of the mouse. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.3305-10.2010

Wang, Z., Kruijne, W., & Theeuwes, J. (2012). Lateral interactions in the superior colliculus produce saccade deviation in a neural field model. *Vision Research*, *62*. https://doi.org/10.1016/j.visres.2012.03.024

Wang, Z., & Theeuwes, J. (2014). Distractor evoked deviations of saccade trajectory are modulated by fixation activity in the superior colliculus: computational and behavioral evidence. *PloS One*, *9*(12). https://doi.org/10.1371/journal.pone.0116382

Watt, R. J., & Morgan, M. J. (1983). The recognition and representation of edge blur: Evidence for spatial primitives in human vision. *Vision Research*. https://doi.org/10.1016/0042-6989(83)90158-X

Webster, M. A., & de Valois, R. L. (1985). Relationship between spatial-frequency and orientation tuning of striate-cortex cells. In *Journal of the Optical Society of America, A, Optics, Image & Science* (Vol. 2, Issue 7, pp. 1124–1132). Optical Society of America. https://doi.org/10.1364/JOSAA.2.001124

Whalen, P. J., Rauch, S. L., Etcoff, N. L., McInerney, S. C., Lee Michael, B., & Jenike, M. A. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *Journal of Neuroscience*, *18*(1). https://doi.org/10.1523/jneurosci.18-01-00411.1998

White, B. J., Kan, J. Y., Levy, R., Itti, L., & Munoz, D. P. (2017). Superior colliculus encodes visual saliency before the primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1701003114

White, B. J., Theeuwes, J., & Munoz, D. P. (2012). Interaction between visual- and goal-related neuronal signals on the trajectories of saccadic eye movements. *Journal of Cognitive Neuroscience*, *24*(3). https://doi.org/10.1162/jocn_a_00162

Wollenberg, L., Deubel, H., & Szinte, M. (2018). Visual attention is not deployed at the endpoint of averaging saccades. *PLoS Biology*, *16*(6). https://doi.org/10.1371/journal.pbio.2006548

Wright, R. D., & Ward, L. M. (1994). Shifts of visual attention: An historical and methodological overview. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *48*(2). https://doi.org/10.1037/1196-1961.48.2.151

Xing, J. (2007). *LNCS 4553 - Information Complexity in Air Traffic Control Displays*. 797–806.

Yamins, D. L. K., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex.

Proceedings of the National Academy of Sciences of the United States of America, *111*(23). https://doi.org/10.1073/pnas.1403112111

Yarbus, A. L. (1967). *Eye movements and vision*. Plenum Press. https://doi.org/https://doi.org/10.1007/978-1-4899-5379-7

Zeki, S. M. (1978). Uniformity and diversity of structure and function in rhesus monkey prestriate visual cortex. *The Journal of Physiology*, *277*(1). https://doi.org/10.1113/jphysiol.1978.sp012272

Zelinsky, G. J., & Bisley, J. W. (2015). The what, where, and why of priority maps and their interactions with visual working memory. *Annals of the New York Academy of Sciences*. https://doi.org/10.1111/nyas.12606

Zhang, Xilin, Zhaoping, L., Zhou, T., & Fang, F. (2012). Neural Activities in V1 Create a Bottom-Up Saliency Map. *Neuron*. https://doi.org/10.1016/j.neuron.2011.10.035

Zhang, Xukun, Sun, Y., Liu, W., Zhang, Z., & Wu, B. (2020). Twin mechanisms: Rapid scene recognition involves both feedforward and feedback processing. *Acta Psychologica*, *208*, 103101. https://doi.org/10.1016/j.actpsy.2020.103101

Zhao, Y., Humphreys, G. W., & Heinke, D. (2012). A biased-competition approach to spatial cueing: Combining empirical studies and computational modelling. *Visual Cognition*, *20*(2). https://doi.org/10.1080/13506285.2012.655806

Zhaoping, L. (2006). Theoretical understanding of the early visual processes by data compression and data selection. In *Network (Bristol, England)*. https://doi.org/10.1080/09548980600931995

Zhaoping, L. (2016). From the optic tectum to the primary visual cortex: migration through evolution of the saliency map for exogenous attentional guidance. In *Current Opinion in Neurobiology* (Vol. 40). https://doi.org/10.1016/j.conb.2016.06.017

Zhaoping, L. (2019). A new framework for understanding vision from the perspective of the primary visual cortex. In *Current Opinion in Neurobiology*. https://doi.org/10.1016/j.conb.2019.06.001

Zhaoping, L., & May, K. A. (2007). Psychophysical tests of the hypothesis of a bottom-up saliency map in primary visual cortex. *PLoS Computational Biology*. https://doi.org/10.1371/journal.pcbi.0030062

Zhaoping, L., & Zhe, L. (2015). Primary Visual Cortex as a Saliency Map: A Parameter-Free Prediction and Its Test by Behavioral Data. *PLoS Computational Biology*. https://doi.org/10.1371/journal.pcbi.1004375

SUPPLEMENTARY MATERIAL

To compare the spatial frequency content of our features, we computed the distance of the spatial frequency (SF) spectra (defined as the Euclidian distance of the vectors of Fourier components' amplitude) within and between our two sets of stimuli (*optimal* and *non-optimal* features; Figure 8C and 8E). The spectra have been calculated after subtracting the mean value from each feature, to remove the irrelevant constant component. The spread of frequency spectra within a given set of features can be visualized by plotting a histogram of the above-defined distances, taken between all possible pairs within the set. First, the histogram of distances within the set of all possible 512 3x3 binary features (excluding degenerate cases with distance < 10^{-4}, due to symmetrical/negative features), shows that they are comprised in the range [0.3, 1.7] (**Figure S1**).

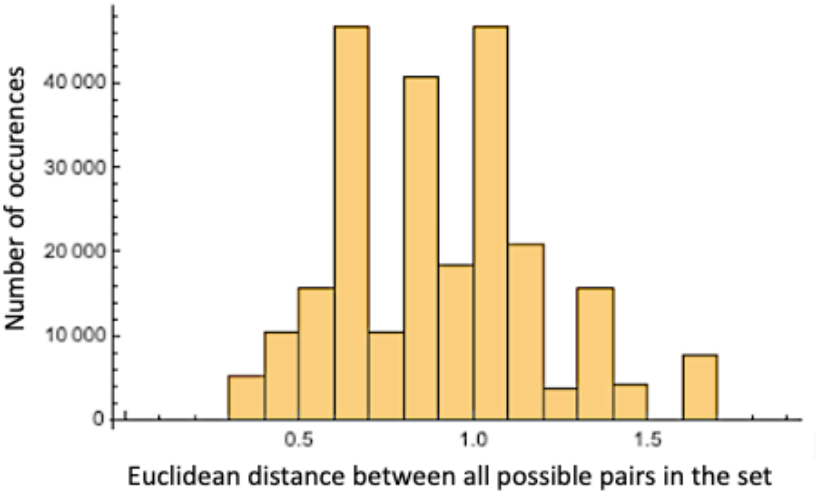

Figure S1. Distances within the set of all possible features. Figure retrieved from (Castellotti et al., 2023a).

The histograms of inter-feature distances taken within the two (much smaller) sets of *non-optimal* and *optimal* features (**Figure S2**) indicate the diameter of the respective feature sets in the 9-dimensional space of frequency spectra. The spread of frequency spectra within the set of *non-optimal* features [0, 1.2] is not much lower than the diameter of the entire feature space. For the optimal set of features, it is even the same as the whole space; all one can see is a slight tendency for lower values.

These results show that the two sets of features do not occupy specific corners of the frequency spectrum but are rather spread over the entirety of the theoretically available space.

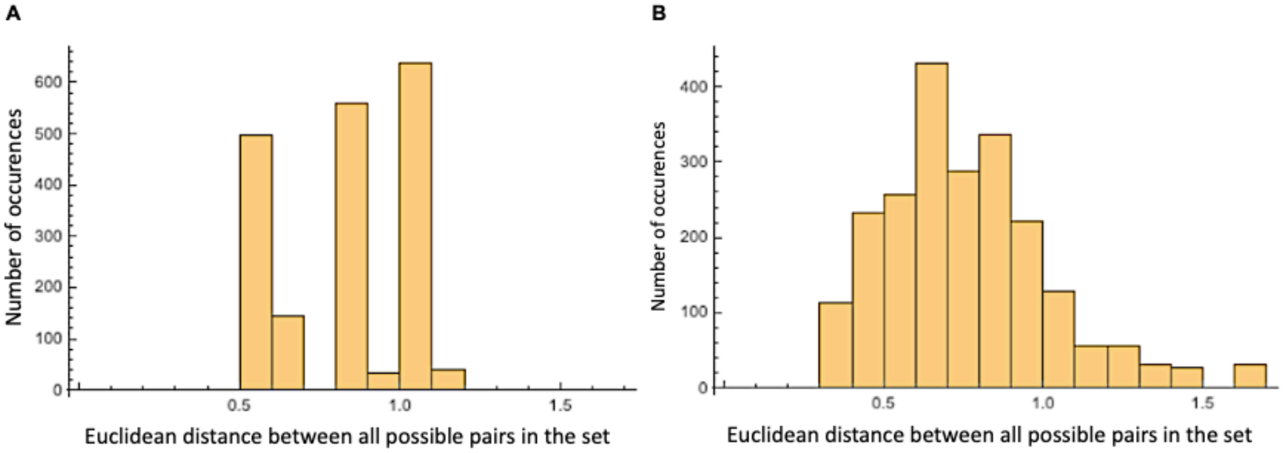

Figure S2. Distances within the set of *non-optimal* (A) and *optimal features* (B). Figure retrieved from (Castellotti et al., 2023a).

Furthermore, it can be seen that the distribution of distances between all possible pairs of features, formed by picking one in each of the two sets (*optimal* vs. *non-optimal*), covers again the same [0.3, 1.6] range (**Figure S3**).

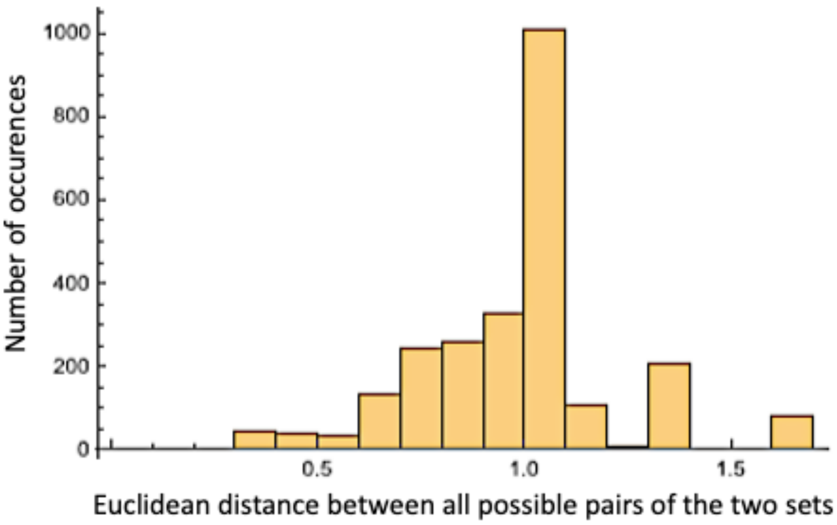

Figure S3. Distances between *optimal* and *non-optimal* feature sets. Figure retrieved from (Castellotti et al., 2023a).

Also, comparing the means of individual components of the nine-dimensional spectra of the two sets, the resulting z-score was equal to or less than 1.

In sum, the two sets have spectra that do not differ by much more than the typical distances within each individual set, and they both extend over essentially the whole frequency space theoretically allowed for their size.

Also, by looking at the closest *non-optimal* feature to each of our *optimal* features, it turns out that, for 48 out of our 50 *optimal* features there is at least one *non-optimal* feature at a distance of less than 0.5 - that is less than the minimum distance between any pairs that can be formed between *non-optimal* features themselves.

In light of these results, the different effects of *optimal* vs. *non-optimal* features found in our studies cannot be explained by their spatial frequency content.

Acknowledgments

First and foremost, I would like to thank my supervisor Maria Michela Del Viva who provided me with so much valuable guidance and support throughout the duration of my PhD. I am extremely appreciative of the opportunities she gave me and of the huge amount of time and effort she put into my development as a researcher. I couldn’t have asked for a better supervisor.

Thanks to Anna Montagnini, who host me in Marseille providing me with a work and life experience that I will never forget.

My PhD experience wouldn’t have been so much special without my amazing lab mates. A special thanks to my friends Irene, Viola, and Paula for the unceasing emotional support: they were tough years, but we made it!

Thanks to Valentina and the guys of LaComune for reminding me that sometimes you have to stop being on the computer.

Last, but not least, a special thanks to my family and Chiara who supported me during these years with so much love.