UNIVERSITÀ DI PARMA

DIPARTIMENTO DI MEDICINA E CHIRURGIA

**CORSO DI LAUREA MAGISTRALE IN
PSICOBIOLOGIA E NEUROSCIENZE COGNITIVE**

Visual Exploration of Natural Movies in Freely-viewing Macaques

Esplorazione visiva di scene naturali dinamiche in macachi in libera visione

Relatore:
Chiar.mo *Prof*. LUCA BONINI

Correlatore:
Chiar.mo *Prof*. MARCO LANZILOTTO

Laureando:
FRANCESCO ROMANO

ANNO ACCADEMICO 2022-2023

INDEX

ABSTRACT

In the study of the visual system, basic artificial stimuli such as spots or bars of light have been instrumental in uncovering the receptive field (RF) structure of retinal and cortical neurons, serving as pivotal milestones in vision research. Yet, the sensory input encountered in real-world scenarios is rarely composed of such isolated simple patterns. A growing body of evidence emphasizes the critical need to investigate visual processing within the context of active natural vision, showing that naturalistic stimuli, mirroring the complexity and dynamism of real-world environments, are not only more functionally relevant but also occasionally more effective in eliciting neuronal responses. Here, we introduce a paradigm that facilitates an ecologically valid exploration of the anatomical and physiological mechanisms of visual processing and attention in the macaque's early vision pathway, where studies utilizing naturalistic paradigms are still lacking. Moreover, in a preliminary phase prior to neural acquisition, we conducted analyses of ocular behavior to verify whether, and to what extent of ocular behavior follows repeatable patterns across repeated presentations of the same vs different clips, and to determine if high- and low-level properties of the stimuli or the overall level of attentional engagement drive the observed variability. The results show that attentional deployment in primates is influenced by both high-level factors, such as the presence of conspecifics, and low-level features, like salience of the stimuli. However, these factors rarely result in highly reproducible ocular patterns, particularly in simple, static scenes or when salience is elevated. Additionally, our examination of attentional engagement levels reveals that while attention and gaze duration tend to be consistent across various content categories when the average looking time towards the movie throughout the session is high, when the time in which the gaze is directed at the video window is slightly higher than when it is directed elsewhere, stimuli featuring primates capture attention more strongly and sustainably compared to non-biological stimuli. This suggests a predominant role for conspecific and socially relevant stimuli in directing attention and eye movements, surpassing the effects of saliency alone.

ABSTRACT (ITALIAN VERSION)

Nell'ambito di studio del sistema visivo, stimoli semplici come punti o barre luminose hanno storicamente ricoperto un ruolo fondamentale per delineare la struttura dei campi recettivi dei neuroni retinici e corticali. Tuttavia, è raro che l'input sensoriale proveniente dal mondo reale sia composto da pattern semplici e isolati. Sono sempre più numerose le evidenze che sottolineano la necessità di indagare l'elaborazione visiva in contesti di libera visione e con l'ausilio di stimoli naturalistici che, riflettendo la complessità e il dinamismo degli ambienti reali, sono sia più rilevanti da un punto di vista funzionale, che, talvolta, efficaci nell'elicitare l'attività neuronale. Nel presente lavoro, introduciamo un paradigma che permette uno studio ecologicamente rilevante dei meccanismi anatomo-fisiologici dell'elaborazione e dell'attenzione visiva nella via visiva sottocorticale del macaco, dove studi che utilizzano paradigmi di visione naturale sono ancora pochi. Inoltre, in fase preliminare all'acquisizione neurale, abbiamo analizzato il comportamento oculare per verificare se, ed in che misura, questo sia riproducibile tra diverse presentazioni di stesse o differenti clip e se fattori attentivi, sia di alto che basso livello, o il livello generale di engagement attentivo possano avere un'influenza sul comportamento oculare. Dai risultati emerge come l'allocazione dell'attenzione nei primati sia influenzato sia da fattori di alto livello, come la presenza di conspecifici, sia da proprietà di basso livello, come la salienza degli stimoli. Tuttavia, questi fattori determinano solo raramente pattern oculari altamente riproducibili, che sembrano verificarsi principalmente in scene poco complesse e statiche oppure quando il livello di salienza è particolarmente elevato. Inoltre, dalle nostre analisi emerge che, quando il tempo medio di osservazione del video è alto, attenzione e durata dello sguardo tendono ad essere diretti in egual misura verso tutte le categorie di stimolo presentate, mentre quando il tempo in cui lo sguardo rimane focalizzato sul video è solo leggermente superiore rispetto a quando è diretto altrove, gli stimoli che raffigurano conspecifici catturano l'attenzione in maniera più forte e sostenuta rispetto a quanto fatto dagli

stimoli non biologici, suggerendo un ruolo prioritario per gli stimoli conspecifici e socialmente

rilevanti nel dirigere l'attenzione e i movimenti oculari, superando gli effetti della salienza.

Abstract

1. INTRODUCTION

The environment is populated with objects of varying significance: some are irrelevant, while others can prove crucial for our survival. Life, therefore, demands all living organisms a certain ability to ignore irrelevant information and to pay attention towards potentially relevant ones. To this end, a crucial role is played by the eyes, sensory organs that move continuously to gather visual information about the surrounding environment (Noton & Stark, 1971).

1.1. We Need to Move our Eyes

The first reason why humans need to move their eyes lies in their anatomical and physiological features. The fovea, the region of maximum vision acuity, is remarkably small. The human foveal depression has a diameter of about 0.8-1.5 mm and covers only about one-four thousandth of the retinal surface (Bringmann et al., 2018; Land & Tatler, 2009). Moving away from the fovea, acuity falls rapidly, reaching one-tenth of its maximum value at 20° of eccentricity (*Figure 1*). Therefore, eye movement is essential for centering objects on the fovea and, consequently, obtaining detailed visual information of the surrounding environment (Land & Tatler, 2009).

Figure 1. Visual acuity decreases with eccentricity (Snowden et al., 2012).

Humans use a "saccade and fixate" strategy when exploring the visual scene, with information gathered during stabilized fixations, and saccades used to shift gaze direction as rapidly as possible. While foveal focusing is a key aspect of vision in humans and primates, a repertoire of eye movements can also be observed in animals that lack a fovea. The same "saccade and fixate" strategy is indeed shared by nearly all vertebrates and found in many invertebrates: primates, fishes, crabs, flies and cuttlefish all exhibit the same consistent pattern of fast eye movements and stable fixations (Land & Tatler, 2009; Land, 1999; Land, 2011). The saccade and fixate pattern is so ancient that it can also be observed in the lampreys, relatives of the jawless fish of the Ordovician period that lived 450 million years ago. Building upon data on our early fish ancestors, in 1962 Gordon Walls published a landmark article where he stated that the primary function of eye movement was not to scan the surroundings but rather to maintain a stable image on the retina. In this view, the real reason for having eye movements was not so much to shift gaze but to avoid motion blur, which would inevitably occur in a mobile animal with fixed eyes (Land, 2019). This blurring results from the long response time of photoreceptors: cones take 20 milliseconds to register a change in light intensity and consequently fail to accurately capture an object moving across them in less than 20 milliseconds (Land, 2019; Land & Tatler, 2009).

In all vertebrates the image is stabilized by two reflexes, the vestibulo-ocular reflex (VOR) and the optokinetic reflex (OKR), both of which were present in ancient fish. In VOR the semi-circular canals of the inner ear provide a head velocity signal which the oculomotor system translates into instructions to the eye muscles to move the eye at a speed equal and opposite to the head movement, thus keeping gaze stationary. OKR is a reflex initiated by ganglion cells in the retina that detect residual image motion. This signal is sent back to the eye muscles, causing the eyes to move in the same direction of the image motion. This action helps

stabilize the retina relative to the image (Land, 2019). However, as an animal turns while navigating its environment, stabilization alone isn't enough. The eyes need to move periodically to recenter the gaze, and saccades allow this minimizing the duration of image blurring thanks to their remarkable speed (Land, 1999).

1.1.1. A Closer Look on Saccades and Fixations

Humans spend about 10% of their waking hours making saccades, during which they are effectively blind, either through blur or "saccadic suppression" (Ibbotson & Krekelberg, 2011; Land, 1999). These rapid eye movements, lasting approximately 20-50 milliseconds (Dragoi & Sur, 2006), promptly move the fovea from one fixation point to another at a frequency of up to 3 times per second (Land, 1999). Saccades are characterized by a highly standardized waveform, featuring a smooth increase and decrease in eye velocity, reaching speeds of up to 900° per second. The velocity of a saccade is exclusively determined by its amplitude. Although we can consciously adjust the amplitude and direction of saccades, their speed remains beyond our voluntary control (Kandel et al., 2021).

As highlighted in the previous paragraph, eye movements are primarily important for analyzing the details of the visual world. Consequently, between successive saccades, the eyes maintain relative stillness during fixations, typically lasting about 100-400 milliseconds (Dragoi & Sur, 2006). It's important to note that even during fixation, the eyes are not entirely motionless but engage in continuous miniature movements, such as ocular drift and microsaccades (Krauzlis et al., 2017). These subtle eye movements during fixation exert significant effects on visual perception. Smooth ocular drift, by subtly shifting the retinal image, enhances spatial acuity. Microsaccades contribute instead to improved vision by repositioning the fovea and foveola but can also introduce momentary interruptions and visual distortions akin to those caused by larger saccades. In contrast to past beliefs, fixation is not a passive

interval between eye movements; rather, it is an active and dynamic process (Krauzlis et al., 2017).

1.1.2. Neural Substrates of Saccades and Fixations

The neural circuit responsible for controlling saccadic eye movements is one of the most comprehensively understood systems in the brain (Carpenter, 2000). A significant body of research, comprising lesion studies, functional neuroimaging, animal neurophysiology and anatomical investigations, has highlighted a network of cortical and subcortical areas crucial for controlling visual fixation and saccadic eye movements. These areas, organized in a hierarchical manner, include regions of the cerebral cortex, basal ganglia, thalamus, superior colliculus (SC), brainstem, and cerebellum (see *Figure 2*) (Munoz & Everling, 2004).

At the highest level of the hierarchy, two cortical areas come into play: the lateral intraparietal area of the posterior parietal cortex (LIP) and the frontal eye fields (FEF). Area LIP, located in the lateral bank of the intraparietal sulcus in the parietal lobe (Wardak et al., 2011), is situated at the interface between visual information and motor command and is mainly responsible for the control of both saccades and visual attention (Bisley & Goldberg, 2006; Colby et al., 1996; Powell & Goldberg, 2000). It receives projections from extrastriate visual cortices and is reciprocally connected both to the intermediate layers of the superior colliculus (SCi) and to the frontal cortical oculomotor areas, including the frontal eye fields (FEF) (Blatt et al., 1990; Wardak et al., 2011). In contrast to neurons in the parietal cortex, those in the frontal eye field are more closely linked with the execution of saccades, firing before and during saccades to their movement fields (Bruce & Goldberg, 1985; Curtis & Connolly, 2008; Segraves & Goldberg, 1987). Moreover, saccade-related neurons in FEF promote the activity of the SCi both directly, through excitatory signals from direct projections (Munoz & Everling, 2004; Segraves & Goldberg, 1987), and indirectly, by releasing it from the inhibition exerted by the

substantia nigra through the activation of the caudate nucleus (Hikosaka et al., 1989; Hikosaka et al., 2000; Hikosaka & Wurtz, 1983; Munoz & Everling, 2004; Stanton et al., 1988). After passing through the superior colliculus, which will be discussed in detail below, the signal reaches the pontine and mesencephalic reticular formations in the brainstem, which supply the motor signals required to activate the extraocular muscles during saccades (Cohen & Henn, 1972; Keller, 1974; Scudder et al., 2002; Sparks, 2002).
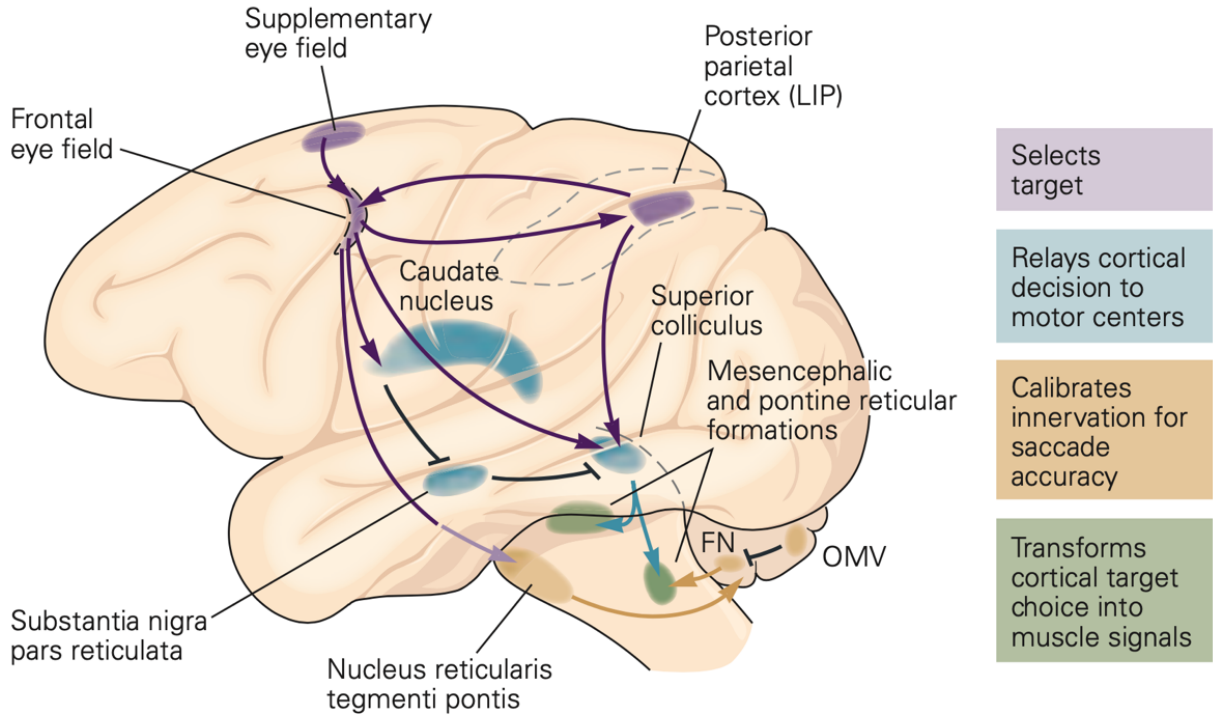

Figure 2. Neural circuitry of oculomotor control (Kandel et al., 2021).

1.1.2.1. The Superior Colliculus

Commonly referred to as the final common path for saccades, the superior colliculus serves as a crucial hub in the circuit of eye movements, positioned midway between higher-order cortices, which govern when and where to execute saccades, and the saccadic burst generators in the brainstem, which give rise to the extraocular muscle signals (Girard & Berthoz, 2005).

The superior colliculus (SC) is a laminar structure located anteriorly in the roof of the midbrain and is the mammalian homolog of the optic tectum in nonmammalian vertebrates (May, 2006). Thanks to its remarkably conserved organization across vertebrates, the superior colliculus/optic tectum is a key structure for comparative research (Allen et al., 2021). The superior colliculus can be categorized into two functional regions (King, 2004; May, 2006) (see *Figure 3*): the superficial layers, which primarily encode visual information, and the intermediate and deep layers, which orient eye and head movements.

The superficial visual layers receive direct input from the retina and a projection from the striate cortex, representing the entire contralateral visual hemifield. Neurons in the visual layers of the superior colliculus, like those in visual areas, create retinotopic visual maps, with neighboring neurons responding to adjacent locations in the visual field (see *Figure 3*(**A**)) (Gandhi & Katnani, 2011; Krauzlis et al., 2013). The SC representation mirrors the contralateral visual fields from both eyes, with the upper visual field (lower retina) mapped medially and the lower visual field (upper retina) mapped laterally (Daroff & Aminoff, 2014). The fovea is overrepresented at the rostral pole of the SC, and the size of receptive fields increases with eccentricity (Cynader & Berman, 1972), resembling cortical visual maps. Notably, a disproportionately large part of the SC map is devoted to the central area of the visual field. This topographic arrangement extends from the periphery represented caudally in the layer to

the vertical meridian and fovea represented rostrally (*Figure 3*(**B**)) (Daroff & Aminoff, 2014; Gandhi & Katnani, 2011; Krauzlis et al., 2013).
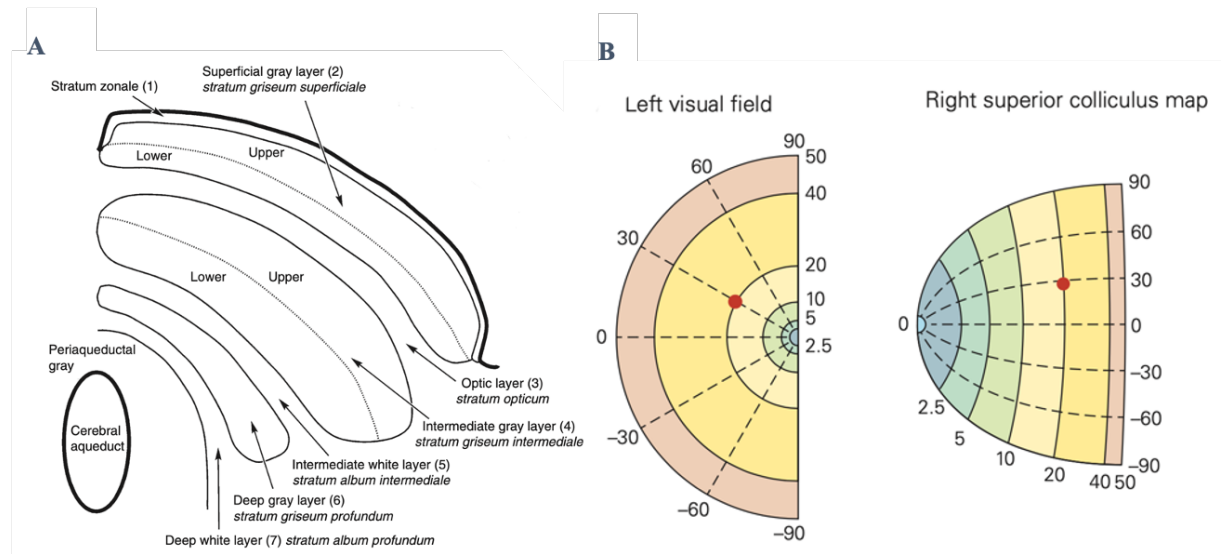

Figure 3. (**A**). Diagram illustrating the layered organization of the mammalian superior colliculus and the corresponding terminology for each layer (Daroff & Aminoff, 2014). (**B**). Retinotopic organizations of neurons in the superior colliculus (Kandel et al., 2021).

As introduced earlier, neurons in the superficial layers exhibit responses to visual stimuli. In monkeys, approximately half of these vision-related neurons show quantitative enhancement in their responses when an animal is preparing to make a saccade to a stimulus within the cell's receptive field (Wurtz & Goldberg, 1972; Wurtz et al., 1982).

The intermediate and deep layers are primarily related to oculomotor activity. Within these layers, movement-related neurons selectively discharge before saccades with specific amplitudes and directions (Gandhi & Katnani, 2011; Wurtz & Goldberg, 1972). Each movement-related neuron in the superior colliculus possesses a "movement field", representing a region in the visual field targeted by saccades controlled by that neuron (Goldberg & Wurtz, 1972; Sparks et al., 1976; Wurtz & Goldberg, 1972). The map of movement fields in the intermediate layers is in register with the map of visual receptive fields in the overlying superficial layers (Schiller & Stryker, 1972; Wurtz & Goldberg, 1972), ensuring the integration

of visual and motor signals within these layers. Additionally, movement fields are large (Goldberg & Wurtz, 1972): that's why neurons in the superior colliculus fire for saccades of various amplitudes and directions, even though they preferentially discharge for a specific direction and amplitude. This results in the activation of a substantial population of cells before each saccade, collectively encoding eye movements through the broadly tuned responses of each neuron (Kandel et al., 2021).

Regarding fixations, as observed earlier, they are not static intervals; instead, they involve subtle eye movements contributing to functional aspects of visual perception. These movements appear to be controlled by neuronal mechanisms involving many of the same brain regions responsible for generating voluntary eye movements (Krauzlis et al., 2017). The neurons closest to the motor output that are believed to be involved in fixation are the omnipause neurons, situated in the nucleus raphe interpositus of the paramedian pontine reticular formation (Scudder et al., 2002). These neurons sustain firing during fixation and cease firing during saccades in all directions, inhibiting saccade-related burst neurons and preventing the generation of saccades (Krauzlis et al., 2017). Among the structures providing inputs to the omnipause neurons, the superior colliculus is particularly noteworthy (Krauzlis et al., 2017). The most rostral portion of the superior colliculus receives inputs from the fovea and the foveal representation in primary visual cortex (V1) (Kandel et al., 2021; Krauzlis et al., 2017). Neurons in the intermediate layers in this region exhibit robust discharge during active visual fixation and before small saccades to the contralateral visual field (Everling et al., 1998; Munoz & Wurtz, 1993). This region is commonly referred to as the fixation zone due to the observed neuronal activity during visual fixation (Krauzlis et al., 2017). Neurons in this zone inhibit the movement-related neurons in the more caudal parts of the colliculus and also project directly to the nucleus of the dorsal raphe, where they inhibit saccade generation by exciting the

omnipause neurons (Munoz & Wurtz, 1993). However, some authors argue against the "fixation zone" hypothesis, citing evidence that these neurons are not a functionally distinct class but rather an extension of saccade-related neurons found elsewhere in the SC. According to the more recently proposed "equilibrium hypothesis", fixation corresponds to an equilibrium state in which target-related activity is balanced across the two SC (Krauzlis et al., 2017), and saccades are triggered when this activity becomes sufficiently imbalanced (Goffart et al., 2012).

1.2. Where Do We Look?

The visual world is an extremely rich and complex source of information. Primates, which have a very small region of acute vision, manage this overwhelming influx of information by scanning the environment through frequent and precise eye movements. Similarly, cognitive and perceptual systems also exhibit limited processing capacity. In this regard, a pivotal role is played by "visual attention", which comprises a set of cognitive mechanisms capable of modulating the signal across multiple levels of the visual system (Evans et al., 2011; Moore et al., 2003; Wolfe & Horowitz, 2004).

Visual attention contributes to many different functions within the visual system. Among these, the filtering of visual information appears undeniably essential. Attention can fulfill this role by selecting a small subset of potentially relevant stimuli for more in-depth processing while relegating the rest to only limited analysis (Borji et al., 2013; Evans et al., 2011; Itti & Koch, 2001; Wolfe & Horowitz, 2004). While the fundamental role of visual attention in directing an organism's limited resources is widely accepted, the criteria for selecting certain stimuli over others remain a subject of debate.

1.2.1. What Factors Drive the Allocation of Attention?

The deployment of gaze is not random but is rather guided by several factors. Broadly speaking, two primary factors can be identified: 1) bottom-up, low-level factors that direct attention toward the most conspicuous locations and 2) top-down, high-level factors in which attention is guided by the observer's goals, potentially relevant objects or by the "gist" of the scene (Connor et al., 2004; Einhäuser et al., 2008; Evans et al., 2011; Van der Linden, 2018; Wolfe & Horowitz, 2017). Based on these two factors, researchers have constructed models of guidance that seek to explain and predict eye movements. Although models have been created to explain eye movements in both free viewing and visual search, the focus of this thesis lies primarily on free-viewing models.

1.2.1.1. Low-level, Saliency-based Guidance

When observing a scene, certain locations stand out due to their visual conspicuity. These locations, termed "salient", attract the gaze solely based on low-level visual features such as color, contrast, orientation or luminance. Low-level, saliency-based guidance is labeled as "bottom-up" because it relies only on visual features and is independent of the semantic content or the observer's goals. The interest in bottom-up saliency guidance, supported by an extensive body of psychophysical literature demonstrating how low-level features can rapidly capture attention (Tatler et al., 2011; Treisman & Gelade, 1980; Wolfe & Horowitz, 2004, 2017), has led researchers to develop computational models of saliency, better known as "saliency models" (Koehler et al., 2014; Kümmerer & Bethge, 2023).
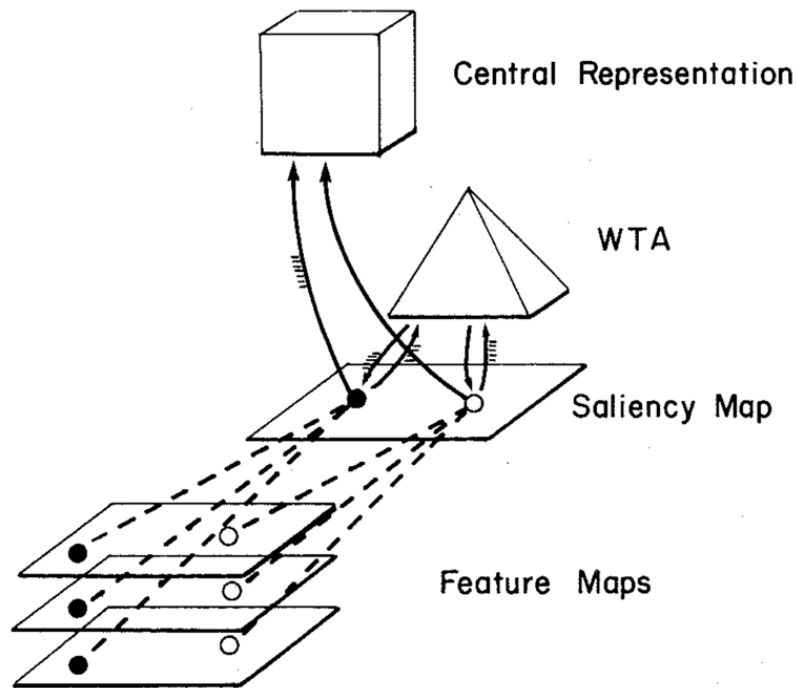
Figure 4. A schematic drawing of Koch & Ullman's saliency model (Koch & Ullman, 1985).

The saliency models, starting from a static image provided as input, generate a topographic map that quantifies the level of salience for each spatial position in the image (Koehler et al., 2014; Kümmerer & Bethge, 2023). The first saliency model (*Figure 4*) was proposed by Koch & Ullman in their 1985 landmark article (Koch & Ullman, 1985). The initial stage of the model includes a set of different topographical maps, each encoding a specific feature, as color, contrast, orientation or direction of movement. For each feature, conspicuity maps are computed, highlighting locations that significantly differ from their surroundings in terms of the corresponding feature. These conspicuity maps are then integrated into one global conspicuity map, known as saliency map, from which the most conspicuous location is selected using a biologically plausible winner-take-all (WTA) mechanism. The current WTA selection is then inhibited to shift attention to the next most salient location. The mechanism proposed by Koch & Ullman was later implemented by Itti et al. (1998), using color, intensity, and edges as elementary features and center-surround differences to detect feature pop-out.

A major strength of this model, which quickly made it widely adopted, is its "image computable" nature, allowing it to be applied to any arbitrary image for quick saliency map generation (*Figure 5*) (Kümmerer & Bethge, 2023).
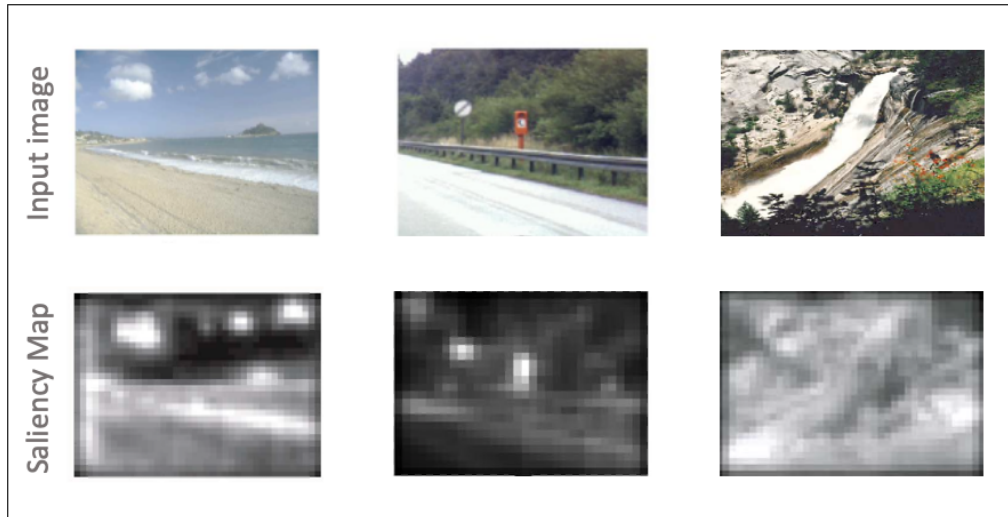

Figure 5. Examples of saliency maps with the respective input images (Itti et al., 1998; Parkhurst et al., 2002).

While the original visual saliency model was initially designed to predict covert attention allocation, numerous studies have assessed its efficacy in predicting eye movements during free-viewing conditions, where observers explore images without a specific task (Berg et al., 2009; Borji et al., 2012; Einhäuser et al., 2008; Parkhurst et al., 2002; Peters et al., 2005). The widespread use of free-viewing conditions stems from the assumption that the absence of a task would accentuate the impact of salience (Kümmerer & Bethge, 2023). Empirical assessments of the salience model using natural scenes reveal that salience at fixated locations is notably higher than at control locations (Foulsham & Underwood, 2008; Parkhurst et al., 2002). However, while statistically significant differences exist in the low-level content of fixated locations compared to control locations, the extent of these differences tends to be modest, indicating a weak correlation between low-level features and fixations (Tatler et al., 2011). Moreover, while low-level features may be good correlates of fixation probability, several

authors argue against their causal role in driving attention (Einhäuser et al., 2008; Tatler et al., 2011). In sum, saliency-based models have limitations and are not fully capable of accounting for human eye movements.

1.2.1.2. High-level Guidance

High-level or top-down guidance can refer to aspects related to the observer, such as their goals, as well as to the properties of the stimulus or the scene. Since the focus of this thesis is on free-viewing conditions rather than visual search, aspects related to the observer's mindset won't be assessed.

Among the high-level properties of a scene, objects constitute a crucial aspect (Einhäuser et al., 2008; Koehler et al., 2014; Kümmerer & Bethge, 2023; Tatler et al., 2011). We observed how early saliency models utilize information about individual features without distinguishing whether these features constitute part of an object or not. However, the world is populated by objects, and it is towards these objects that organisms direct their actions. Numerous studies demonstrated that the distinction between fixated and control locations is more accurate when considering object-level information rather than relying solely on low-level salience (Azadi et al., 2024; Einhäuser et al., 2008; Koehler et al., 2014). This observation is further supported by the finding that the best predicting models are those which consider both low-level and high-level, object-like properties of the visual scene (Azadi et al., 2024; Kummerer et al., 2017; Kümmerer & Bethge, 2023). Some object categories that attract human fixations include non-biological elements like text and cars, as well as biological entities such as animals and, notably, human body parts (Borji et al., 2013; End & Gamer, 2017; Kümmerer & Bethge, 2023; Rubo & Gamer, 2018; Skripkauskaite et al., 2023). Further evidence indeed indicates that biological objects with social features, including human faces, heads, and bodies, capture attention during free viewing, outperforming low-level saliency in predicting fixations (Azadi et al., 2024; End

& Gamer, 2017; Rubo & Gamer, 2018; Skripkauskaite et al., 2023). Moreover, social areas of interest such as faces and bodies attract more attention than non-biological objects (Skripkauskaite et al., 2023). Another higher-level aspect that influences gaze deployment is scene context. Observers are indeed capable of extracting the "gist" of the scene very rapidly, influencing both object recognition and the selection of certain objects over others (Van der Linden, 2018; Wolfe & Horowitz, 2017).

1.2.1.3. Where Do Monkeys Look?

Due to extensive homologies, monkeys are widely used as animal models for the study of human cognitive processes, such as visual attention. Although there are many similarities in the brain architecture, validating monkeys as a model for human visual attention is crucial, and this requires a comparison of their behavior under realistic free-viewing conditions (Berg et al., 2009).

Humans and monkeys do not employ identical strategies during free-viewing tasks. Despite these overall differences, the prediction accuracies across species are significantly above chance, indicating a shared degree of fixation selection strategies (McFarland et al., 2013; Wilming et al., 2017). Saliency models predict fixations in both monkeys and humans during the viewing of natural and artificial video clips, with human fixation prediction being stronger (Segraves et al., 2017; Wilming et al., 2017). However, as mentioned earlier, low-level models overlook crucial biological stimuli, such as faces, eyes, and bodies, which are selectively attended to by non-human primates (Kano et al., 2018; Shepherd et al., 2010). Furthermore, humans and monkeys differ in gaze behavior regarding features like the degree of central bias, the duration and regularity of fixation periods, and the amplitude of saccades (Berg et al., 2009; Shepherd et al., 2010). Monkeys exhibit short and stereotyped fixation intervals separated by large saccades, while humans tend to fixate for more prolonged and variable periods (Shepherd

et al., 2010). These differences may facilitate relatively fast threat and resource detection by monkeys and align with the observation that monkeys abbreviate fixations toward high-risk social targets, such as high-ranking male faces (Shepherd et al., 2010).

1.2.2. A Shared Neural Control of Attention and Eye-Movements

In the preceding paragraphs, we explored the close relationship between attention and eye movements, where the latter are employed as a metric to study the former. Gaze shifts correspond to an overt orientation of visual attention, as the target selection carried out by attention manifests in observable eye movements, aimed at a more detailed and in-depth processing (Kümmerer & Bethge, 2023; Steinmetz & Moore, 2012).

Exploring the connection between spatial attention and saccadic eye movements spans several decades of research (Jonikaitis & Moore, 2019). In everyday scenarios, gaze direction naturally corresponds to attention, as we overtly focus on and attend to objects of interest. Nevertheless, attention can disengage from the fixation point. This ability proves particularly crucial during primate social interactions, where direct gaze often signals aggression. As a result, subdominant macaques, for example, often avoid the gaze of the more dominant ones, even while covertly monitoring them (Moore et al., 2003).

While overt behaviors associated with attention, such as moving the eyes and head and orienting to targets of interest, are inherently driven by motor control mechanisms, there is also substantial evidence suggesting a significant overlap in neural mechanisms between covert attention and the control of saccadic eye movements (Corbetta, 1998; Corbetta et al., 1998; Kowler et al., 1995; Kustov & Lee Robinson, 1996; Moore et al., 2003; Rizzolatti et al., 1987).This idea was originally introduced by Rizzolatti and colleagues in their influential Premotor Theory of Attention, proposing that shifts in spatial attention result from saccade planning (Rizzolatti et al., 1987).

1.2.2.1. Cortical and Subcortical Pathways for Visual Attention

Neurophysiological investigations have highlighted key brain structures that seem to play a causal role in both oculomotor behavior and visual attention, including the superior colliculus (SC), the lateral intraparietal area of the parietal cortex (LIP), and the frontal eye field (FEF). Within each of these structures, neurons contribute in various ways to both functions (Awh et al., 2006; Jonikaitis & Moore, 2019; Steinmetz & Moore, 2012).

LIP and FEF emerge as primary candidates for orchestrating spatial attentional effects. They are densely interconnected and each receives visual input from numerous sources (Shipp, 2004; Wardak et al., 2011). LIP functions as a priority map, representing locations and objects based on their attentional priority, influenced by a combination of low-level salience and top-down control (Bisley & Mirpour, 2019; Bisley et al., 2011; Kandel et al., 2021). Beyond its anatomical suitability as a priority map, the responses of LIP neurons exhibit bottom-up features and robustly respond by top-down influences (Bisley et al., 2011). Studies indicate heightened LIP responses to sudden onsets or moving stimuli, resembling salience correlates. Additionally, LIP activity is evident in behaviorally relevant decision-making and reward-related contexts (Bisley & Mirpour, 2019; Bisley et al., 2011).

The current understanding of FEF involvement in attention relies primarily on inference techniques. Moore & Fallah employed a microstimulation approach, setting the intensity below the saccadic threshold (Moore & Fallah, 2001). Despite being insufficient to induce eye movements, the stimulation notably enhanced covert attentional deployment (Awh et al., 2006; Jonikaitis & Moore, 2019; Wardak et al., 2011). Further supporting the involvement of FEF in visual attention, reversible inactivation of this area demonstrated a significant impact on covert visual search (Jonikaitis & Moore, 2019; Wardak et al., 2011).

The subcortical circuit thought to facilitate attentional influences involves two structures, the superior colliculus (SC) and the thalamus's pulvinar nucleus (Krauzlis et al., 2013; Shipp, 2004). Similar to FEF, evidence supporting the superior colliculus' role in spatial attention primarily comes from studies using inference techniques, showing a role in both cover and overt target selection (Jonikaitis & Moore, 2019; Krauzlis et al., 2013; Moore et al., 2003; Veale et al., 2017). Moreover, further studies support a role of the superficial layers of the superior colliculus in salience encoding (Veale et al., 2017; White, Berg, et al., 2017; White, Kan, et al., 2017). White et al. correlated the firing rate of SCs neurons and the output of a well-established computational saliency model, finding discharge patterns that were well predicted by the level of saliency (White, Kan, et al., 2017). Furthermore, the observed lack of selectivity for any visual feature in the superficial layers of the superior colliculus aligns with the feature-agnostic nature of the saliency map (Veale et al., 2017).

As mentioned earlier, another subcortical center involved in visual deployment is the pulvinar, the largest and most posterior thalamic nucleus (Froesel et al., 2021; Tanaka & Kunimatsu, 2011; Wilke et al., 2010). The pulvinar has expanded significantly during primate evolution and now stands as the most extended visual nucleus within the thalamus in higher-order mammals, including humans (Cortes et al., 2023; Wilke et al., 2010). Its classical division comprises three main subnuclei: the lateral, the medial and the inferior pulvinar (Froesel et al., 2021; Wilke et al., 2010). Spatial attention enhances visual responses in the pulvinar. Interestingly, this modulation is stronger in the presence of distractors, suggesting a contribution of the pulvinar to the selection of behaviorally relevant stimuli (Froesel et al., 2021; Robinson & Petersen, 1992; Tanaka & Kunimatsu, 2011). Pulvinar lesions, observed in both humans and monkeys, lead to deficits in spatial and temporal attention. In severe cases, such lesions can give rise to hemispatial neglect syndrome, characterized by a lack of awareness regarding

contralesional visual stimuli (Froesel et al., 2021; Tanaka & Kunimatsu, 2011). Lastly, there is a hypothesis suggesting that the pulvinar contains a representation of visual salience. This is based on the fact that the inferior pulvinar receives input from the superficial layers of the superior colliculus and has a retinotopic map (Veale et al., 2017).

1.3. An Ecological Approach to the Study of Vision

In the first chapter, we looked closely to the intricate interplay between visual processing and eye movements. A comprehensive understanding of vision requires acknowledging the limitations of the retina and the frequent changes in gaze direction. Therefore, studying sensory and motor processing within a unified framework is fundamental to comprehend vision under natural condition (Ibbotson & Krekelberg, 2011; Yamane et al., 2023). An increasing body of evidence and theoretical considerations underline the significance of exploring visual processing under conditions of active vision (Dragoi & Sur, 2006; Felsen & Dan, 2005; Leopold & Park, 2020; Yamane et al., 2023). Adopting naturalistic paradigms entails some risks but allows to gain a profound understanding of the key organizational features of the visual brain that have evolved to thrive within inherently complex and dynamic environments (Leopold & Park, 2020; Russ & Leopold, 2015).

1.3.1 Requirements for an Ecologically Relevant Study of Vision

The goal of systems neuroscience is to unravel how neural circuits process stimuli encountered in the natural environment. Achieving this objective requires considering both the characteristics of natural stimuli and the response properties of neurons under natural stimulation (Felsen & Dan, 2005).

1.3.1.1. Natural Stimuli

Basic artificial stimuli have played a crucial role in uncovering the neural mechanisms of sensory processing. In the visual system, stimuli like spots or bars of light have unveiled the receptive field (RF) structure of retinal and cortical neurons, which constitute the landmarks of contemporary vision research. However, real-world sensory inputs rarely consist of isolated simple patterns (Felsen & Dan, 2005).

Natural stimuli present a complex and intricate structure across various spatial and temporal scales. They exhibit diverse Fourier spectra distribution, with light levels and spatial contrast undergoing drastic variations, while object boundaries contribute to sharp transitions in brightness and color. Additionally, surface textures and illumination gradients introduce smoothly varying elements (Gallant et al., 1998; Karamanlis et al., 2022). Finally, visual stimuli found in natural environments exhibit substantial spatial and temporal correlations, as observed in the similarity in luminance values among neighboring pixels (Felsen & Dan, 2005). Overall, natural stimuli prove functionally more relevant and, in some cases, more effective in driving neuronal responses (David et al., 2004; Felsen & Dan, 2005; Felsen et al., 2005). These results align with the non-linear nature of neural processing, where responses to natural stimuli are poorly predictable from the summation of responses to individual components (Felsen & Dan, 2005).

1.3.1.2. Natural Vision

The success of natural approaches to studying vision is grounded in the widely accepted assumption that the visual system has evolved to optimally respond to the statistical properties of encountered signals. These properties do not rely solely on the physical properties of the environment, justifying the use of natural scenes, but also on mechanisms through which organisms, particularly primates, acquire visual information — namely, eye movements

(Dragoi & Sur, 2006; Leopold & Park, 2020; Simoncelli & Olshausen, 2001). In fact, the visual input is dynamically shaped by oculomotor behavior with precise space-time statistics (Dragoi & Sur, 2006; Simoncelli & Olshausen, 2001). Dragoi & Sur discovered that V1 neurons process orientation in a manner adapted to the statistics of free-viewing (Dragoi & Sur, 2006). They recorded the eye positions of three macaques freely viewing natural scenes and observed that successive fixations tend to occur on stimuli with similar or largely dissimilar orientations. Simultaneously, the most significant improvement in stimulus discriminability in V1 occurs when neurons are adapted to a stimulus with similar or largely dissimilar orientation. Based on these findings, they concluded that V1 neurons have evolved to efficiently encode dynamic stimuli with similar and largely dissimilar orientation structures, as these stimuli have the highest probability of occurrence during natural viewing. Moreover, other studies have revealed different responses of visual areas under free-viewing conditions compared to responses obtained in traditional laboratory conditions (Berg et al., 2009; Gallant et al., 1998; Im & Fried, 2016; McMahon et al., 2015).

1.3.2. The Free-viewing Paradigm

An ecologically relevant study of the visual system should combine the use of natural and dynamic stimuli, such as videos, with natural viewing methods, like free-viewing. The free-viewing paradigm involves a condition in which a subject observes an image or a video without a specific task or instruction, allowing the individual to look freely without constraints (Kümmerer & Bethge, 2023; Tatler et al., 2011). In the neurophysiological field, the free-viewing paradigm typically involves head restriction and an eye-tracking system, with the latter being crucial when the goal is to correlate neural activity with the foveated stimulus content. The visual stimuli utilized can be either static, such as images, or dynamic, like videos. They may represent a variety of content, including natural landscapes, artificial environments,

animals, and even arrays of artificial stimuli (DiCarlo & Maunsell, 2000; Killian et al., 2012; Sakon & Suzuki, 2021; Wang et al., 2023; White, Berg, et al., 2017; White et al., 2021).

Although the free-viewing paradigm is particularly suitable for studying the visuo-oculomotor system, its application in the neurophysiological domain is widespread. Examining studies conducted on monkeys that utilized both free-viewing and videos depicting natural scenes reveals research conducted across various subcortical and cortical areas, including the hippocampus, amygdala, superior colliculus, STS, F7, and posterior parietal cortex (Aboharb et al., 2023; McMahon et al., 2015; Mosher et al., 2014; Wang et al., 2023; White, Berg, et al., 2017). Notably, the free-viewing paradigm is useful not only for investigating visual or oculomotor processing but also for revealing high-level properties such as memory recognition, object distance encoding, visual space mapping, representation of social information, and also mirror response properties (Aboharb et al., 2023; Caggiano et al., 2011; Killian et al., 2012; McMahon et al., 2015; Sakon & Suzuki, 2021). In conclusion, free-viewing, especially when coupled with natural scenes, is a highly versatile research method that allows the investigation of various functions and response properties. Moreover, it departs from reductionist methodologies while providing a good control over variables and a high degree of ecological validity (Russ & Leopold, 2015).

2. AIMS

This study aims to achieve two primary objectives. Firstly, it seeks to create a paradigm that allows for an ecologically relevant investigation of the anatomical and physiological mechanisms of visual attention in macaques early visual pathway. Building upon theoretical frameworks suggesting that free viewing conditions and the use of natural images optimally reveal the properties of the primate visual system, we have developed a two-step paradigm. This includes a phase of unconstrained visual exploration of natural dynamic scenes followed by receptive field mapping of targeted structures, employing natural images as stimuli. Secondly, this study aims at investigating, at the behavioral level, both the attentive engagement and the oculomotor patterns exhibited during movie exploration, to assess potential differences based on the category of the video content. Specifically, the present study aims to verify if, and to what extent, there is reproducibility of ocular behavior between repeated presentations of the same or different clips, and to determine if this varies depending on the nature of the stimulus. Finally, we also aim to investigate how ocular behavior changes based on different levels of attentive engagement exhibited during the session.

3. MATHERIALS AND METHODS

3.1. Subjects

Two male *Macaca mulatta* monkeys of 10 years, weighing 17 kg (Mk1) and 13 kg (Mk2), were trained for the behavioral tasks described below. They were housed together and had access to cage enrichments. Vegetables, seeds and pellets were provided in the home cage. The temperature, humidity and lighting conditions were customized to provide the most ideal environment for the animals. All experimental protocols complied with the European (Directive 2010/63/EU) and national (D.lgs 26/2014) laws on the protection of animals used for scientific purposes, they were approved by the Veterinarian Animal Care and Use Committee of the University of Parma and authorized by the Italian Ministry of Health.

3.2. Surgeries

During all surgeries, anesthesia was induced with ketamine hydrochloride (5 mg/kg intramuscular) and medetomidine hydrochloride (0.05 mg/kg i.m.) and maintained with 2% isoflurane vaporized in 100% oxygen. Surgery was performed in aseptic and stereotaxic conditions. Hydration of the monkey was maintained with continuous infusion of saline solution and eye hydration was ensured through vitamin A eye gel. A heating pad stabilized the monkey's body temperature throughout the surgical procedure. Heart rate, respiratory depth, and body temperature were continuously monitored. Analgesics were administered intra- and post- operatively. Upon recovery from anesthesia, each animal was returned to its home cage and closely monitored until complete recovery. Dexamethasone and prophylactic broad-spectrum antibiotics were administered pre- and post- operatively.

3.2.1. Headpost, Recording Chamber and Electrodes

Initially, the monkeys underwent a surgical procedure to implant a headpost, which served to stabilize their head positions during both training and testing sessions. For neuronal signal collection, a biocompatible plastic recording chamber was implanted in each monkey, dimensions being 44 × 23 × 35 mm (*Figure 7*). The chamber was designed with a grid of parallel grooves, each 1.25 mm wide with 3 mm between grooves, to accommodate up to eight Omnetics connector blocks for both monkeys. These connectors facilitated the interface between the multielectrode contacts and the headstages from Intan Technologies, which were connected to the Open Ephys Acquisition Board. The recording chambers were crafted, with precise cutting and shaping informed by a 3D reconstruction of the monkeys' cranial structures generated from 7T Magnetic Resonance images using 3D Slicer software (*Figure 6*).
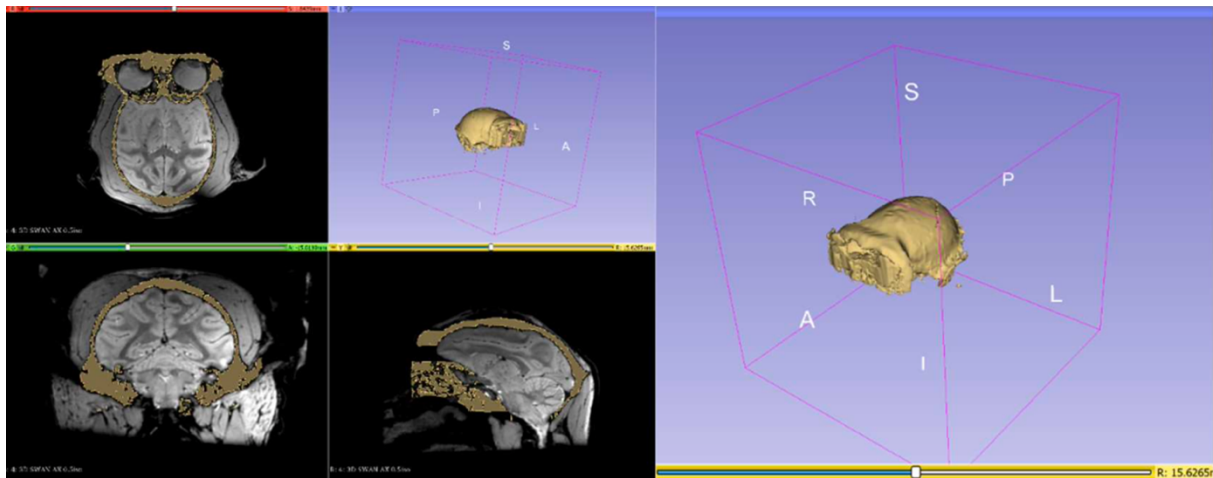

Figure 6. 3D slicer interface. On the left side, bone tissue is identified and annotated in two-dimensional MRI scans across three planes: axial (top left), sagittal (bottom left), and coronal (bottom right). The right side of the figure illustrates the upper right segment of the 3D Slicer environment, which is generated through the integration of these three sectional views.
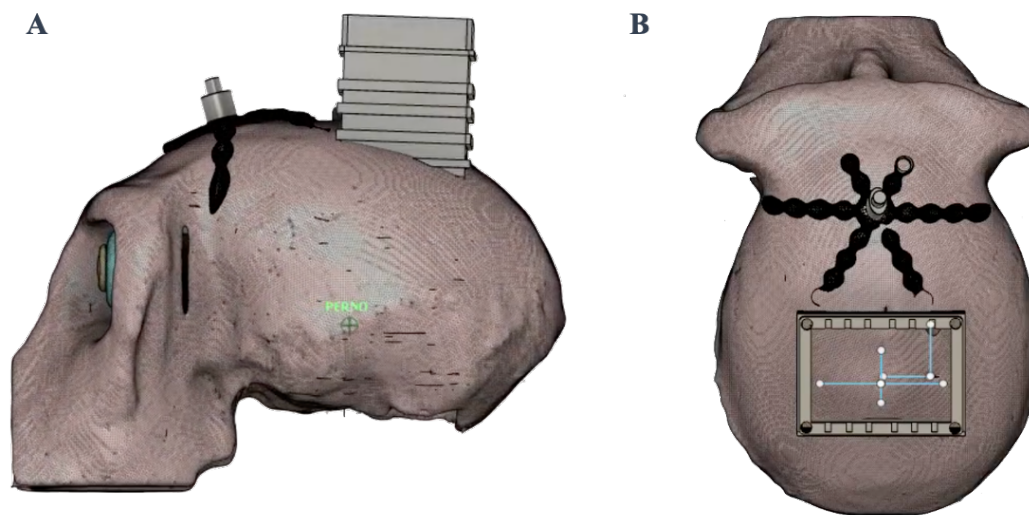
Figure 7. **(A)** Sagittal and **(B)** horizontal view of the headpost and the recording chamber implants on the skull.

Following complete recovery from the chamber implantation, both animals underwent an identical protocol for the insertion of recording probes. The insertion coordinates were determined through analysis of 7T magnetic resonance imaging data, ensuring precise targeting. The implantation included three sites per hemisphere: the superior colliculus, the medial pulvinar, and the lateral pulvinar. The implantation process began with the creation of a dental cement layer within the chamber, poured in liquid form and allowed to solidify. Next, six polyimide guide tubes—three per hemisphere, each 42 mm in length (outer diameter 820 μm, inner diameter 760 μm)—were permanently affixed. Vertical holes were drilled through the cement and bone layers using a stereotaxically guided drill to maintain a straight trajectory. These tubes acted as permanent guides for the insertion of linear probes to prescribed depths, ensuring accuracy and preventing deviation. The probes varied slightly depending on the implantation site. In the superior colliculus (SC) and medial pulvinar (PulM), Plexon V-Probes with 32 Platinum Iridium (Pt/Ir) channels and 125 μm inter-electrode spacing were utilized. These probes, measuring 49 mm in total length, were reinforced with a tube anchored to the ground (43 mm long, 640 μm diameter) (*Figure 8* (**A**)). In the lateral pulvinar (PulL), a 64-

channel probe with 100 μm spacing was employed, accompanied by the same type of reinforcement tube (*Figure 8* (**B**)). The probes were inserted into the guide tubes within the awake monkeys on a roughly weekly schedule and were semi-chronically secured using Kwik-Cast™ silicone sealant.
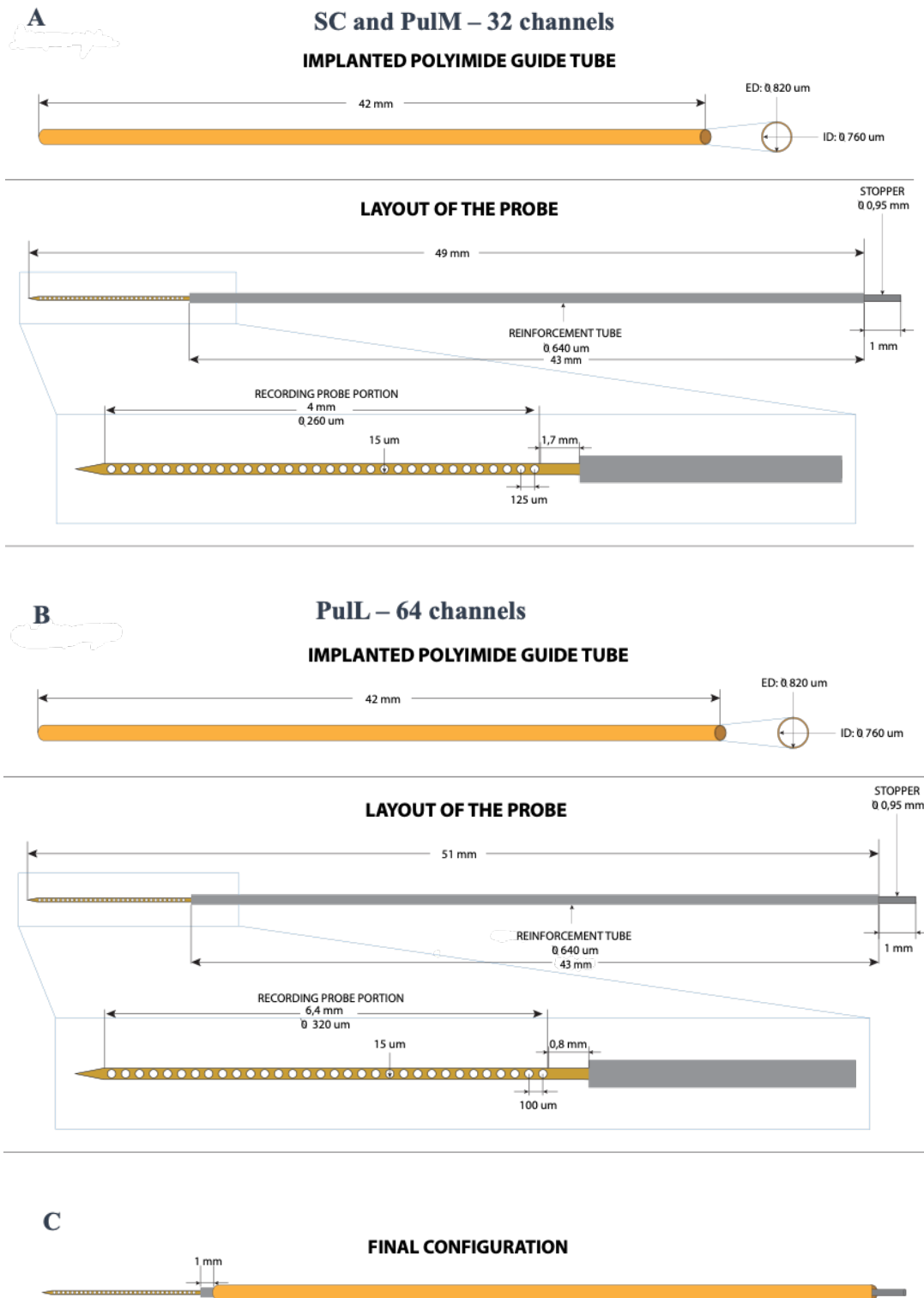
Figure 8. **(A)**. Guide tube and probe features in the superior colliculi and medial pulvinars sites. **(B)**. Guide tube and probe features in the lateral pulvinars sites. **(C)**. Final configuration of the probe insertion through the guide tube.

3.3. Experimental Setup

The monkeys were trained to sit in a primate chair with their heads fixed via a headpost, positioned 57 cm away from a monitor (KD55X85J Sony 55-Inch, refresh rate of 120 Hz, and resolution of 3840 x 2160 pixels), where visual stimuli were displayed (*Figure 9*). Scripts for the tasks were crafted using MATLAB (academic use version, R2023b, The MathWorks, Inc.), and MonkeyLogic (v2.2.24) was utilized for script execution and concurrent behavioral and physiological data collection.

Prior to the tasks, a calibration was performed to track the right eye's position with the aid of the eye tracker software Oculomatic (version 0.4, 600 Hz, spatial accuracy <0.5°, latency <1.8 ms (Zimmermann et al., 2016)). Inputs for Oculomatic were provided by a camera (Chameleon3 CM3-U3-13Y3M, frame rate 149 fps, resolution 1280 × 1024, pixel size 4.8 x 4.8 µm) processed by Flycapture (FlyCap2 2.11.3.121). To facilitate eye tracking, infrared lights illuminated the animals' faces, providing information about gaze position and movement.
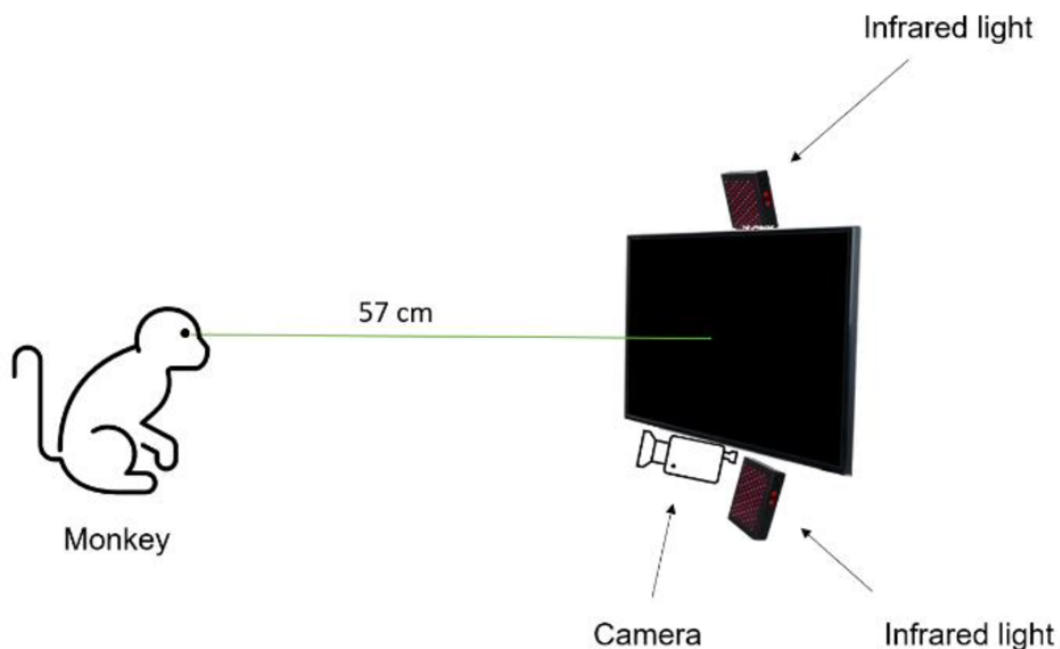

Figure 9. Illustration of the training session arrangement. The monkey was positioned in a primate chair, 57 cm away from a display. Positioned beneath the screen, a camera aimed at the monkey's right eye. For eye-tracking purposes, the monkey's face was illuminated by two infrared lights placed at the monitor's top and bottom, although these lights were invisible to the monkeys.

Materials and Methods

The eye position data, converted into voltage values via user-defined parameters, were then relayed to the behavioral control software, MonkeyLogic, through a National Instrument DAQ board (*Figure 10*). Within MonkeyLogic, trial errors or stimulus presentations triggered the generation of a transistor-transistor logic (TTL) pulse, subsequently transmitted to the Open Ephys Acquisition Board, facilitating the correlation of neuron firing with a specific task phase or event. A photodiode, placed at the lower right corner of the monitor, detected a white square that appeared concurrently with a particular phase of the task. The signal from the photodiode was later analyzed offline to accurately ascertain the stimulus presentation's onset and duration.
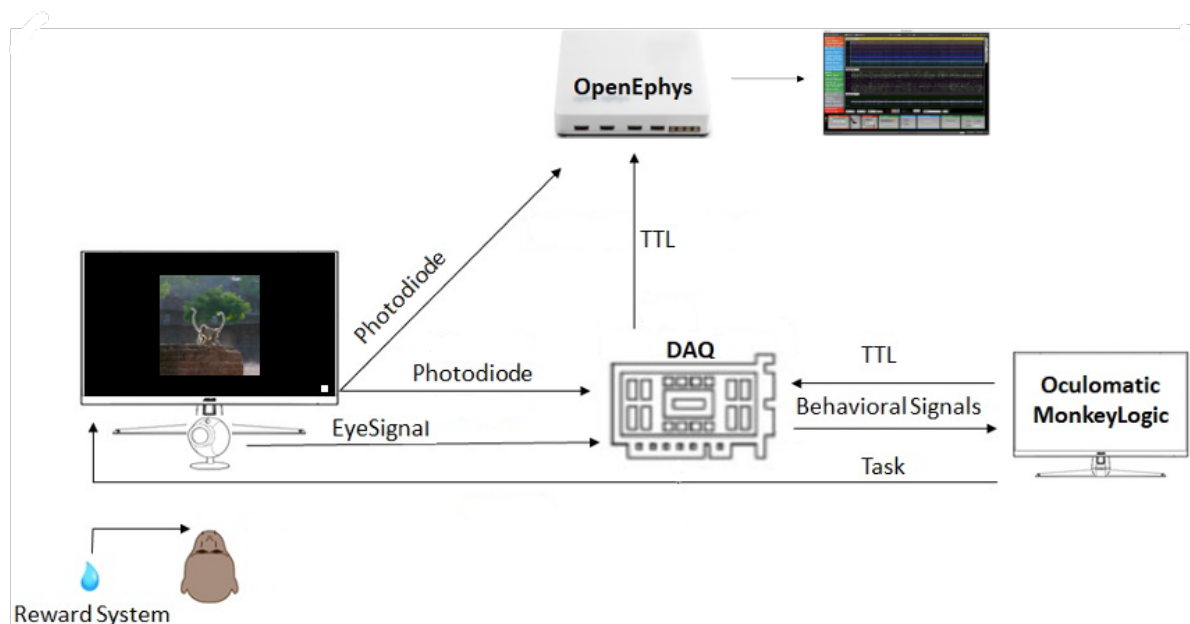

Figure 10. The experimental setup of the neural recording sessions.

3.4 Tasks

The monkeys were trained to perform the tasks using operant conditioning and a behavioral shaping approach, in which the desired final behavior was reached through single incremental steps using positive reinforcements, whereby the task difficulty gradually increased. Every successful trial was rewarded with a liquid reward, other behaviors were not rewarded nor punished. When a trial was not performed correctly, it was interrupted and a new trial was initiated.

3.4.1. Free-viewing paradigm

Before the complete execution of the paradigm, which included the presentation of five videos, the monkeys underwent a habituation process. In this phase, the number of videos shown was progressively increased each day until it reached the total of five. During each free-viewing session, seated approximately 57 cm from the LCD monitor in a primate chair with their heads restrained, the animals watched five movies. Each movie depicted natural scenes and lasted five minutes. Concurrently, an eye-tracking system captured and recorded the eye-position signals. Between the end of one video presentation and the beginning of the next, the monkeys received a liquid reward, following the protocol used in other free-viewing studies (Aboharb et al., 2023; Mosher et al., 2014; White, Berg, et al., 2017).

The video content was assembled from 30 five-seconds clips extracted from HD National Geographic documentaries and copyright-free videos on YouTube. These 30 clips were balanced across three main categories (*Figure 11*), depicting:

- Biological content featuring monkeys

- Naturalistic content depicting natural landscapes and/or non-primate animals

- Non-biological content

Figure 11. The clips were balanced across three main categories. **(A)**. Biological content featuring monkeys. **(B)**. Naturalistic content depicting natural landscapes or non-primate animals. **(C)**. Non-biological movement.

For simplicity, these categories will be referred to as *Monkey*, *Natural* and *Non-Bio* throughout the remainder of the text. Each clip was subsequently mirrored, resulting in a total of 60 clips. The final visual stimuli were obtained by randomly ordering the 60 clips, creating a 5-minute video. The randomization process was repeated 5 times, resulting in a total of 5 videos, each lasting 5 minutes (*Figure 12*). The movies were presented within a square frame of 40° wide and 40° high (640 x 640 pixel resolution) at 25 frames per second.
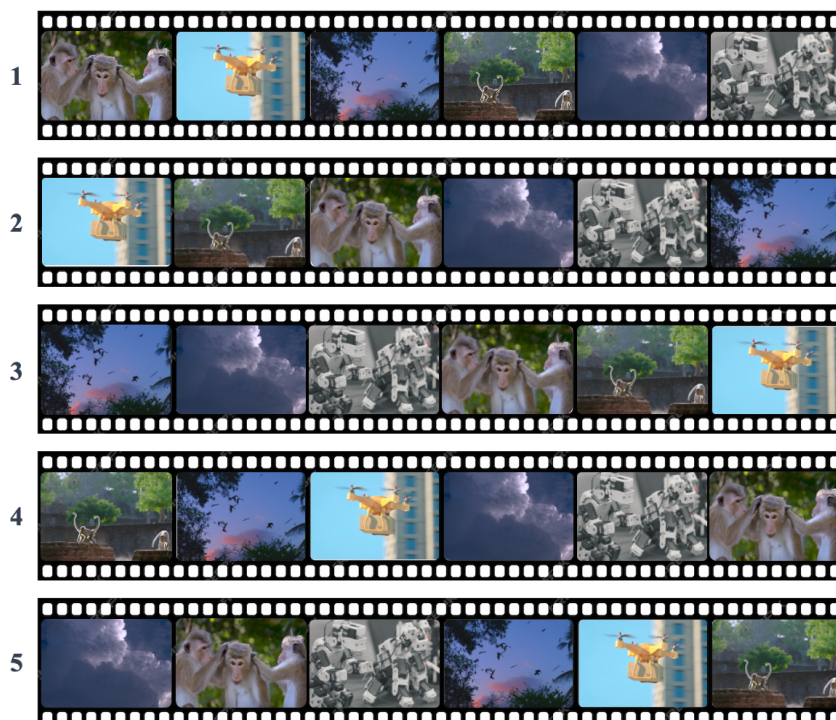

Figure 12. The order of the clips has been randomized five times to generate five different videos.

The video content used in the present study differs from that which will be presented during the neural acquisition phase to preserve the novelty of the stimuli. This is crucial because presenting videos that the subjects had already seen multiple times could diminish their interest and engagement, potentially affecting the neural responses. While the videos are different, it is important to note that the content of the clips remains within the same three categories. This approach ensures that novelty is maintained without altering the fundamental nature of the stimuli under investigation.

3.4.2. RF Mapping

Prior to each free-viewing session, monkeys performed a fixation task essential for mapping the receptive fields (RFs) of neurons in the superior colliculus (SC) and the pulvinar (Pul). To ensure the accurate performance of the fixation task, the monkeys underwent a specific training phase. Although both subjects were already proficient at a fixation task that involved the presentation of a single stimulus per trial, the task used in our study was more complex, involving the sequential presentation of four briefly flashed stimuli within each trial. The training was designed to acclimate the monkeys to this increased demand, ensuring they could maintain fixation throughout the rapid succession of stimuli necessary for the mapping of receptive fields. This preparatory RF mapping is crucial as it allows to correlate neural activity with the entire array of visual elements within the movie scene, not just those in foveal focus, ensuring a comprehensive understanding of how visual stimuli across the entire field of view contribute to neural responses. RF mapping followed the technique described by Papale and colleagues (Papale et al., 2018), utilizing the Berkeley Segmentation Dataset (Martin et al., 2001), a library of natural images in which human observers marked object boundaries. Utilizing natural images facilitates the extraction of distinct low-level features such as spatial

position, orientation, contrast, and luminance, among others, whose predictive influence on neural activity will be assessed to construct the model that better predicts neural responses.

Monkeys engaged in a fixation task to ensure fixation at the center of the display. In each trial, a central fixation point (FP), represented by a green circle with a radius of 0.5° in a fixation window with a radius of 1.5°, was presented. Monkeys were required to gaze at the FP within 1000 ms and maintain fixation for 300 ms. After this 300 ms hold period, four images were presented, each lasting 200 ms and separated by 200 ms intervals (see *Figurse 13*). Monkeys were required to maintain gaze within the fixation window for the entire duration of the trial to receive a liquid reward. In total, 2 sets of 750 images each were presented.
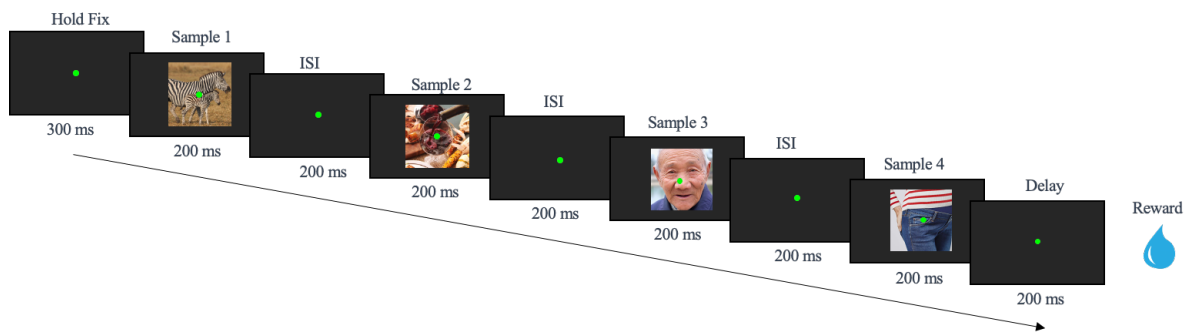

Figure 13. Sequence of events of the fixation task paradigm conducted with naturalistic images for receptive field mapping.

To validate the results obtained from the RF mapping conducted with natural images, a *locally sparse noise* was subsequently presented in the subregion of the visual field where the analysis of receptive fields revealed a higher degree of spatial tuning. This paradigm followed the same sequence as for natural images, with the distinction that the stimuli presented were displays of multiple white spots (see *Figure 14*). To map the receptive field structure, the cellular response was averaged over all trials when a given location is occupied by a spot. Although other spots were present in the stimulus for each trial, the arrangement of these spots was always different, ensuring that the impact of other spots on a cell's response would average out.
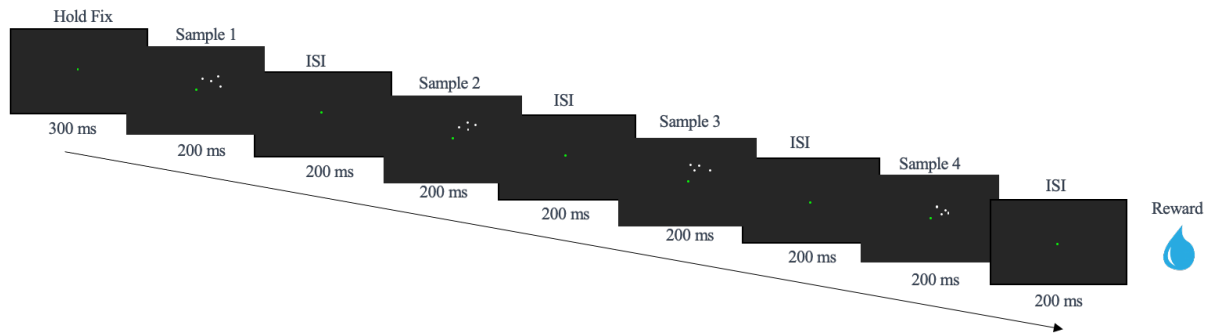
Figure 14. Sequence of events of the fixation task paradigm conducted with *locally sparse noise* for validating receptive field mapping with natural images.

3.5. Data Analysis

All analyses were conducted using the eye-tracking data, recorded in horizontal (x) and vertical (y) coordinates, during the free-viewing of the movies. The ocular data, recorded by MonkeyLogic at a sampling rate of 1000Hz, were downsampled to match the 149Hz sampling frequency of the video camera, resulting in a pair of coordinates approximately every 7 milliseconds. Furthermore, using the video start and end timestamps, all coordinate pairs recorded before or after the presentation of any video were excluded. All analyses were conducted using MATLAB (academic use version, R2023a, 9.14.0.2286388, The MathWorks, Inc.) on two different sessions: one with high attentive engagement (HAE) and one with low attentive engagement (LAE).

3.5.1. Average Looking Time Percentage

To quantify attentional engagement, we calculated the ratio of the number of eye position signals detected within the video presentation window to the total number of eye position signals recorded during each movie viewing. This ratio, representing the proportion of looking time on the video, was converted into a percentage for each of the five videos presented in a session. We then averaged these percentages to determine the average looking time per session.

3.5.2. Correlation Matrices and Heatmap Generation

To assess the consistency of ocular behavior across different clip presentations, a correlation analysis was conducted. Ocular data for each 5-second clip presentation were isolated and then methodically arranged into an array of 300 cells, corresponding to the total number of clips displayed in a session—60 clips per video over five videos. The array was structured in the following fashion: the first 100 cells were dedicated to *Monkey* category clips, followed by 100 cells for *Natural* category clips, and the final 100 for *Non-Bio* category clips. Within these category-specific segments, the ocular data from different presentations of the same clip were placed sequentially. Each category section comprised 20 sets of 5 cells, corresponding to the 20 clips presented per category and their five repetitions in a session. Subsequent to the data structuring, the dataset was refined to contain exclusively the eye position signals within the 40-degree square video frame, which was our primary area of interest. For each ocular data matrix within the array, heatmaps with an 8x8 bin resolution were generated. These heatmaps were then normalized and transformed into vectors. Kendall's tau correlation coefficient was calculated for each pair of vectorized heatmaps, constructing a comprehensive 300x300 correlation matrix. To determine the average level of correlation within and between categories, the mean correlation values were computed for each 100x100 cell block of the correlation matrix, resulting in a 3x3 correlation matrix.

3.5.3. Category-related attention analysis

To compare the level of attentive engagement between categories, a paired sample t-test was conducted using both raw data and fixation clusters. Due to its higher variability, the *Natural* category was excluded from the comparison. In the first case, the average looking time for the video clips within each category was computed. Specific blocks of cells in the dataset corresponding to the *Monkey* and *Non-Bio* categories (cells 1 to 100 and 201 to 300,

respectively) were utilized. Within each block, every group of 5 cells contained all 5 repetitions of each video clip. For each of these groups, the average number of eye position signals recorded was computed, representing the average looking time for each clip. These values were then combined into vectors for each category, facilitating t-test analysis. For the t-test, we adopted an alpha level of 0.05 to determine statistical significance.

To evaluate potential differences between the *Monkey* and *Non-Bio* categories in attentive engagement levels, fixation clusters were also utilized. To extract fixations from our eye data, we utilized the EyeMMV toolbox (Krassanakis et al., 2014), which employs an algorithm based on adjustable spatial and temporal constraints. This fixation detection process relies on three main parameters: two spatial criteria and one for minimum duration. The spatial parameters of the algorithm, t1 and t2, utilize Euclidean distance calculation respectively to differentiate between different clusters and to admit only the points that meet a certain degree of coherence, thus eliminating any potential instrument noise. We selected the same value for both spatial parameters (t1 = t2 = 0.5), equivalent to 1 degree of visual angle, as recommended in the literature (Blignaut, 2009; Krassanakis et al., 2016), along with a relatively restrictive temporal criterion, allowing only clusters with a minimum duration of 200ms, thus reflecting sustained attentional capture. The fixation detection algorithm was applied to each cell of the array to obtain the number of fixation clusters for each presentation of every clip across all categories. Subsequently, we averaged across the 5 presentations of each clip to obtain the average number of recorded fixation clusters per clip. Similarly to the average looking time per clip, the obtained values were aggregated to create a vector for each category of interest, upon which the paired t-test was applied, using the predetermined alpha level of 0.05 to assess statistical significance.

Materials and Methods

4. RESULTS

4.1. Average Looking Time Percentage

Two distinct sessions from one monkey were analyzed for this work, by selecting specifically those with the highest and lowest levels of attentional engagement (see **Materials and Methods 3.5.1.**). In the session characterized by high attentional engagement (HAE), the average looking time percentage was 79% (SD = 4.7), while in the session with low attentional engagement (LAE), it amounted to 58% (SD = 5.8) (*Figure 15*).
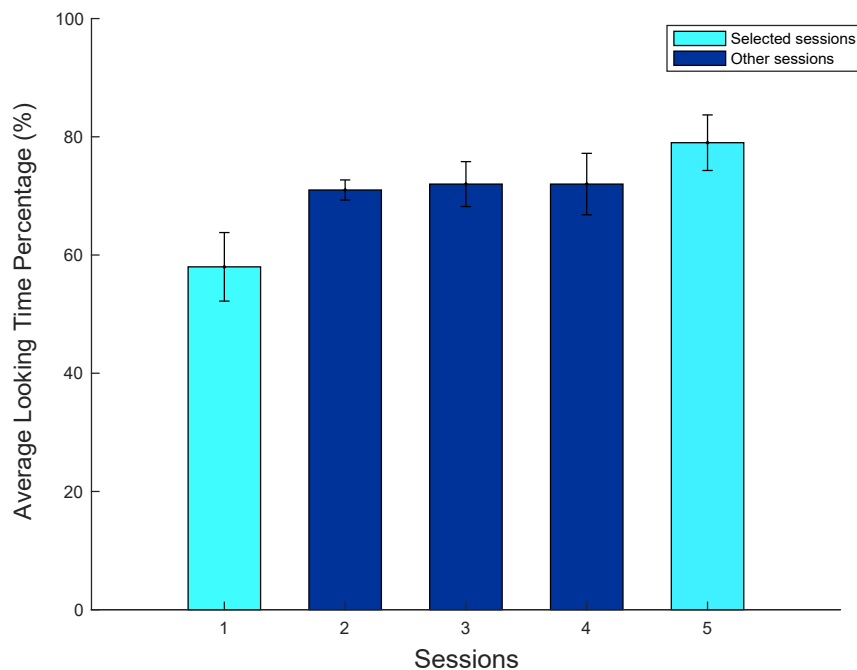

Figure 15. Average percentage of looking time during different free-viewing sessions. Sessions with the highest and lowest values are highlighted in light blue for our analysis.

These results offer an overview of the overall attentive engagement during each session, serving as valuable indicators for subsequent evaluation and quantification of potential variations in eye movement behavior when exploring the same stimuli under different levels of attentional engagement.

4.2. Correlation Analysis

To further analyze eye movement patterns, the consistency of ocular behavior for each session was examined. The analysis focused on determining and quantifying the possible presence of a correlation between the observed oculomotor pattern during the viewing of clips belonging to different categories, between different clips within the same category, and across multiple presentations of the same clip. In addition to this, conducting the same analyses on both sessions allowed us to examining whether these correlation values varied depending on high or low levels of attentive engagement (*Figure 16*). The first step in the investigation aimed to determine the average correlation among clips within the same category (intra-category) and between clips across different categories (inter-category) (see **Material and Methods 3.5.2.**).
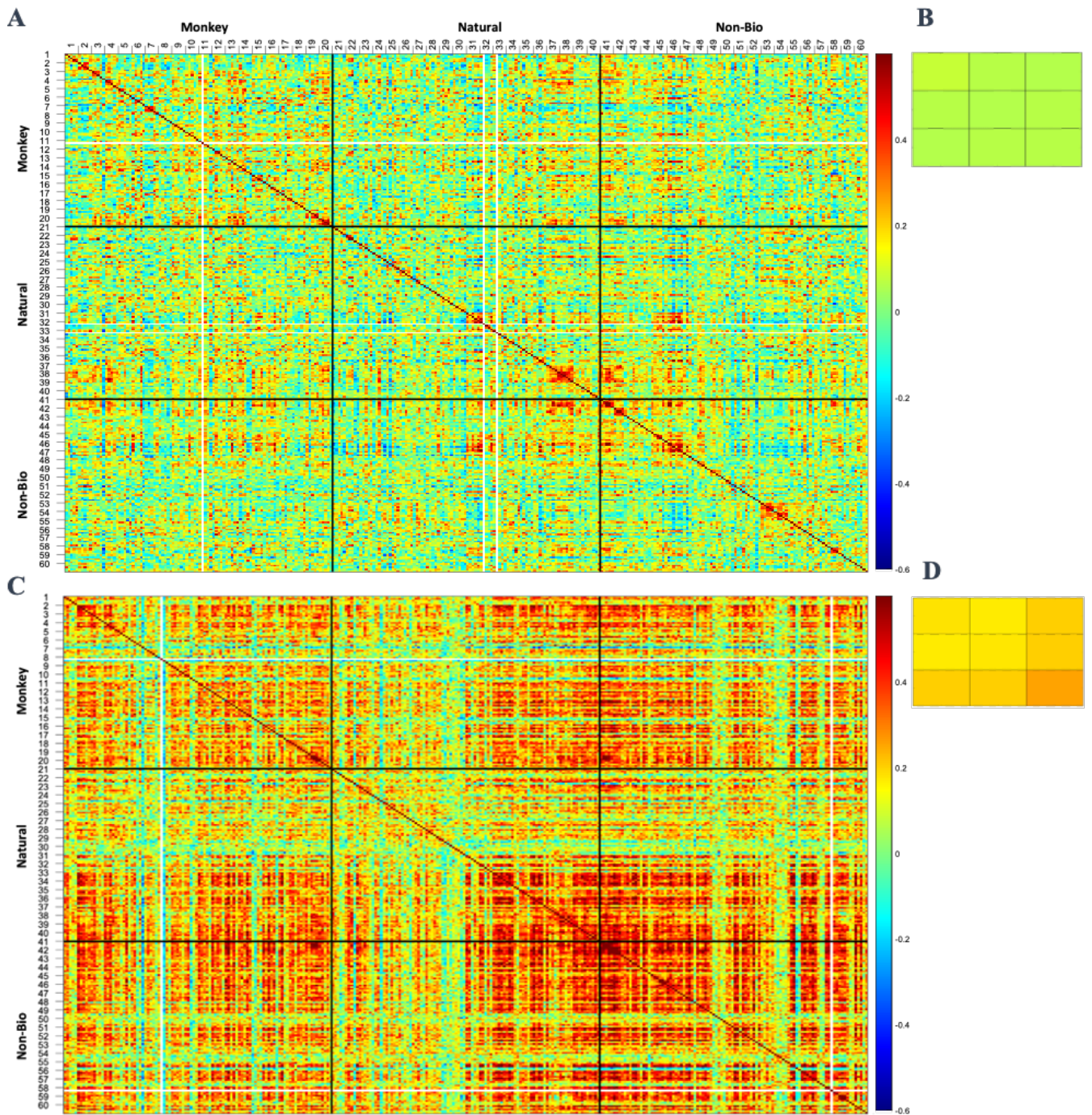
Figure 16. **(A)** Correlation matrix of the HAE session. **(B)** Matrix containing the average values of each 100 x 100 block in the HAE session. **(C)** Correlation matrix of the LAE session. **(D)** Matrix containing the average values of each 100 x 100 block in the LAE session.

In the HAE session, both correlation values within each category and those between different categories remained close to 0. The maximum correlation value, at 0.1, is observed for the block corresponding to presentations of clips in the *Monkey* category (*Figure 16* (**B**)). Conversely, during the LAE session, we observed generally higher values, around 0.2, with a peak of 0.26 observed for the *Non-Bio* category (*Figure 16* (**D**)). In the HAE session (*Figure 16* (**A**)), there appears to be no correlation between eye movement patterns belonging to different categories, except for few sets between the the *Natural* category and the *Non-Bio* category, where there seems to be a clustering of correlation values higher than 0.2. Additionally, there seems to be no correlation between patterns recorded during the viewing of different clips belonging to the same category. At the level of individual clips, we observe that the correlation values between eye behaviors across multiple presentations of the same clip are highly variable and inconsistent. In fact, this consistency appears to be clip-dependent; some clips exhibited notable reproducibility in visual exploration across viewings, unlike others. As anticipated by the matrix of average correlation values per block, in the LAE session, there appears to be a generally higher level of correlation, with a remarkable concentration of high values both among different presentations of the same clip and across different clips within the *Non-Bio* category (*Figure 16* (**C**)). To delve deeper into the nature of the correlations observed in both sessions, the heatmaps derived from eye behavior during the viewing of different presentations of the same clip were examined, analyzing each category separately.

Regarding the *Monkey* category, the heatmaps of clips characterized by higher correlation values share a common focus on areas where conspecifics are present (*Figure 17*).
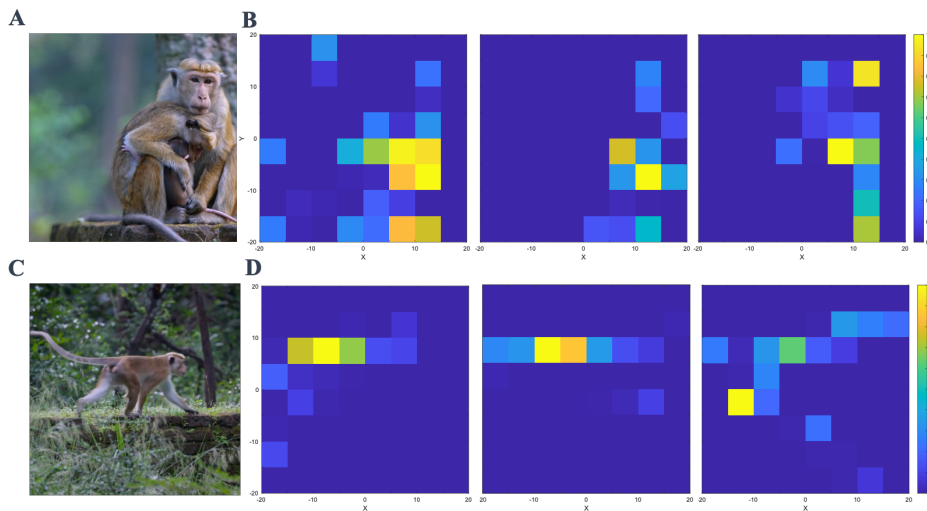
Figure 17. **(A)** Frame of clip 7 (average correlation value = 0.25). **(B)** Normalized heatmap of eye-tracking data obtained from three presentations of clip 7. **(C)**. Frame of clip 19 (average correlation value = 0.26) **(D)**. Normalized heatmap of eye-tracking data obtained from three presentations of clip 19.

Additionally, for most of these clips, this high level of correlation remains consistent across both direct and mirrored presentations of the same clips. However, high levels of correlation are not ubiquitous across all clips in the *Monkey* category, as indicated by the correlation matrix. As illustrated in the case of clip 12, visual exploration may vary significantly across different presentations, resulting in low correlation values between heatmaps (*Figure 18*).

Figure 18. **(A)**. Three different frames from clip 12 (average correlation value = 0.14). The yellow circle indicates potential areas of interest based on the respective heatmaps below. **(B)**. Normalized heatmaps of eye-tracking data obtained from three presentations of clip 12.

Unlike clips with higher correlation values, which feature relatively static scenes with limited elements, most of the other clips feature dynamic scenes and a high number of elements, which could compete and result in different attentional patterns with each presentation.

The *Natural* category also showed highly heterogeneous correlation values among the different clips. Some clips with high correlation values exhibited notably prolonged gaze times towards areas containing relevant content, often focusing more on peripheral areas of the stimulus, such as the edges or corners of the video presentation window, as exemplified in clip 31 (*Figure 19*). The raw ocular data was scrutinized to determine whether the peripheral concentration represents a genuine point of interest or merely serves as a transit zone during saccadic movements (*Figure 19* (**C**)). Consistent patterns were revealed across presentations, where the gaze frequently drifted above the video presentation window.

Figure 19. (**A**). Frame of clip 31 (average correlation value = 0.27) (**B**). Normalized heatmaps of eye-tracking data obtained from three presentations of clip 31 (**C**). Raw eye-tracking data corresponding to the above heatmaps. The blue line represents the eye-tracking trace, while the red rectangle denotes the video presentation window, which is 40° wide and 40° high.

In other cases, as evident from the example of clip 37, a higher gaze time is focused on an area of high visual salience, marked by the presence of a light flash (*Figure 20*).

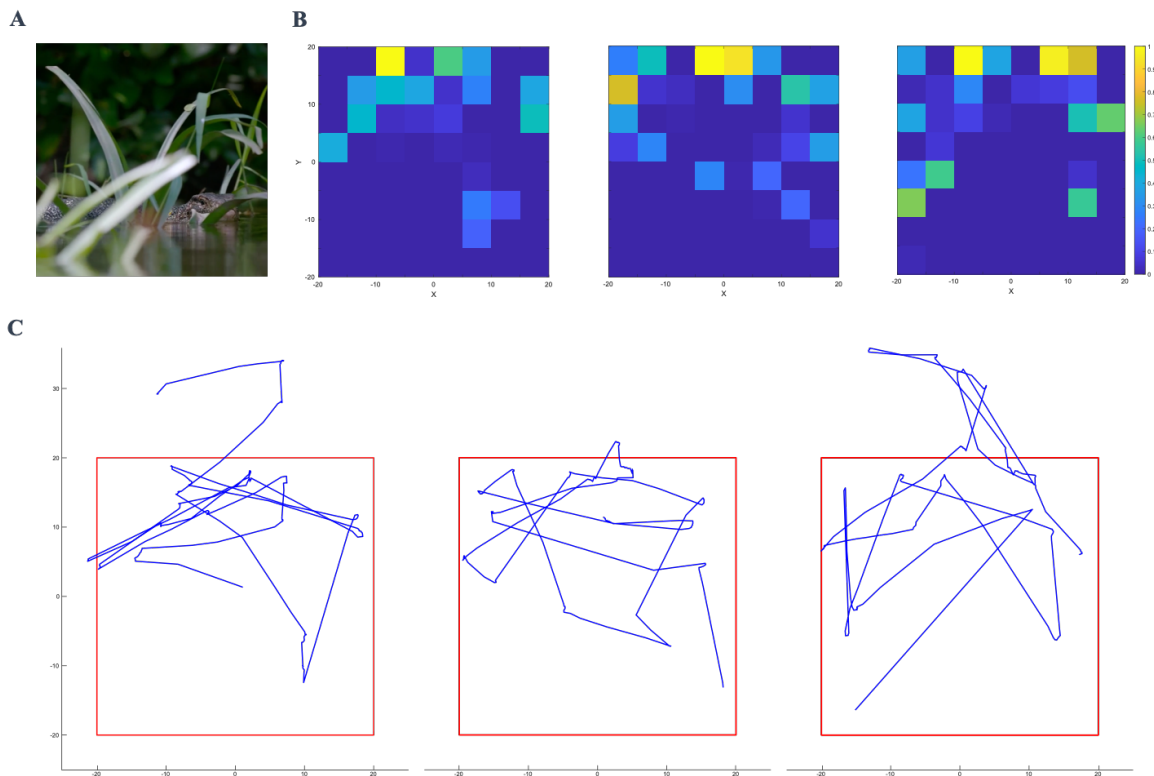*Figure 20.* **(A)**. Frame of clip 37 (average correlation value = 0.26). **(B)** Normalized heatmaps of eye-tracking data obtained from three presentations of clip 37.

In the *Non-Bio* category as well, several clips with high correlation values are observed alongside a high degree of heterogeneity. Similar to the previous category, the high correlation levels may arise from both a tendency to gaze stereotypically towards peripheral areas, which provide little information about the video presentation window, as exemplified by clip 41, and from a prolonged gaze time towards perceptually salient areas of the scene (*Figure 21*).
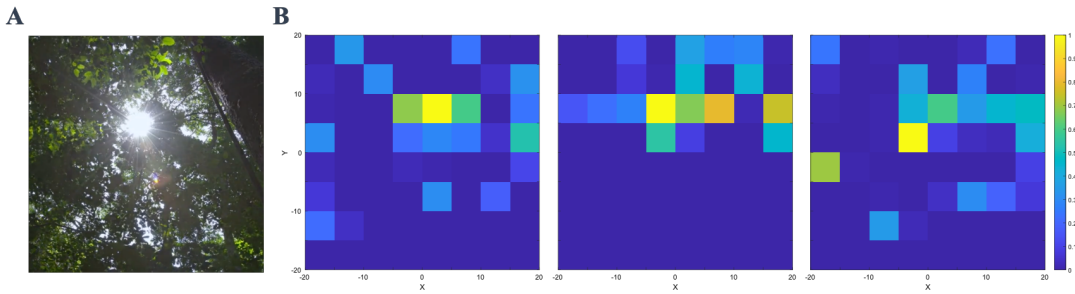
Figure 21. **(A)**. Frame of clip 41 (average correlation value = 0.38) **(B)**. Normalized heatmaps of eye-tracking data obtained from three presentations of clip 41. **(C)**. Raw eye-tracking data corresponding to the above heatmaps. The blue line represents the eye-tracking trace, while the red rectangle denotes the video presentation window, which is 40° wide and 40° high. **(D)**. Frame of clip 46 (average correlation value = 0.40) **(E)**. Normalized heatmaps 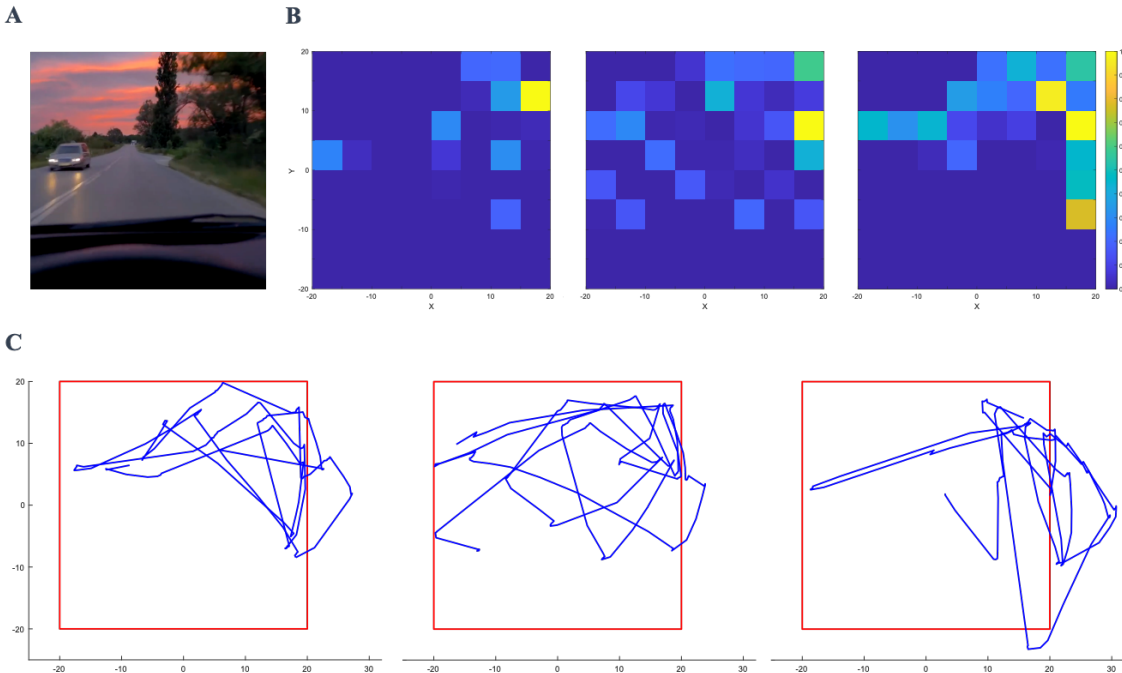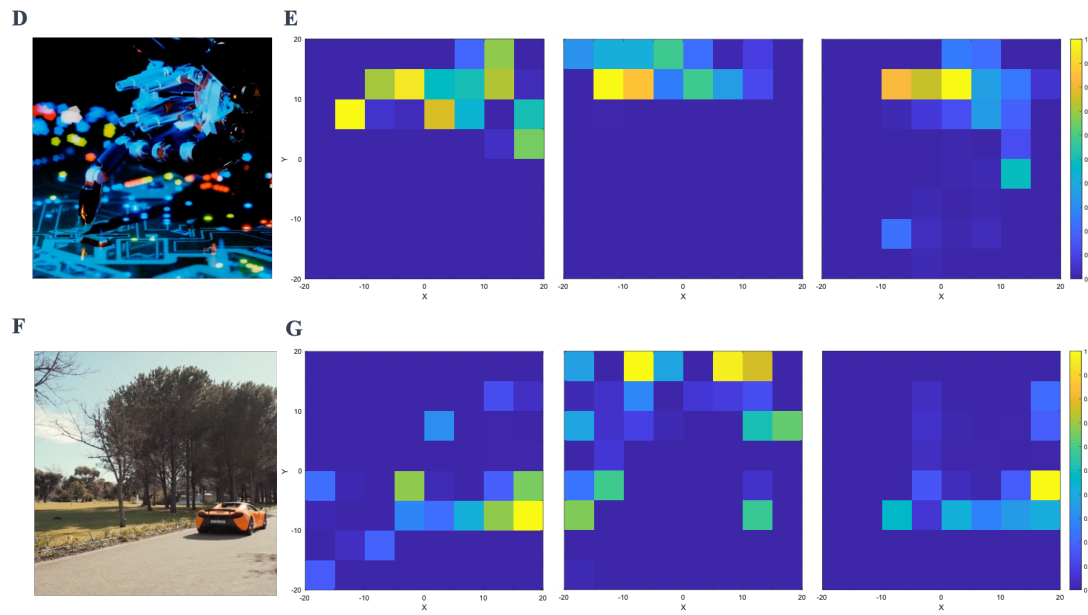of eye-tracking data obtained from three presentations of clip 46 **(F)**. Frame of clip 53 (average correlation value = 0.36) **(G)**. Normalized heatmaps of eye-tracking data obtained from three presentations of clip 53.

In summary, in the *Monkey* category, the eye behavior pattern tends to be captured by high-level content, such as the bodies of monkeys, and is particularly reproducible for clips characterized by a static scene and few elements. Conversely, in the *Natural* and *Non-Bio* categories, high correlations result either from visually salient features that consistently capture attention—such as bright lights or significant movement—or from intrinsic animal behaviors, such as looking away from the video, possibly reflecting a default resting state or distraction.

In the LAE session, as anticipated, there appear to be elevated correlation values across the blocks, affecting both intra- and inter-category comparisons. Regarding the *Monkey* category, it was noted that the visual focus is often drawn to areas rich in content—specifically where monkeys are present. However, a recurrent observation is the attention peaks in the upper right region of the screen—a zone seemingly lacking in both object density and perceptual salience (*Figure 22*).

Figure 22. **(A)**. Frame of clip 2 (average correlation value = 0.44) **(B).** Normalized heatmaps of eye-tracking data obtained from three presentations of clip 2. **(C)**. Frame of clip 4 (average correlation value = 0.35) **(D)**. Normalized heatmaps of eye-tracking data obtained from three presentations of clip 4.

In the *Natural* category, when analyzing clips with particularly high correlation levels, the gaze is predominantly focused on peripheral areas of the screen, notably the upper corners. As observable in the case of clip 33, the scene's most informative feature—a moving predator—largely goes unnoticed (*Figure 23* (**A**, **B**)). In other cases, as shown by clip 38, the same pattern, with significant attention allocated to the upper corners—especially the right one, go along with a fair amount of gaze directed towards the center of the image, a zone with pronounced perceptual salience (Figure 23 (**C**, **D**)).

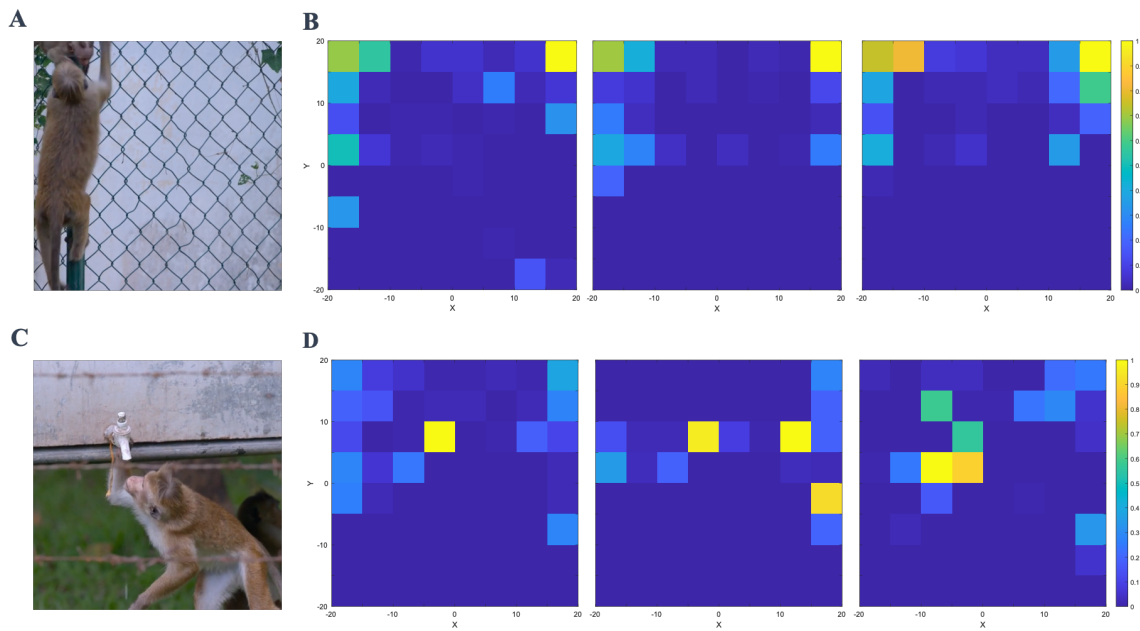*Figure 23.* **(A)**. Frame of clip 33 (average correlation value = 0.47) **(B)**. Normalized heatmaps of eye-tracking data obtained from three presentations of clip 33. **(C)**. Frame of clip 38 (average correlation value = 0.46) **(D)**. Normalized heatmaps of eye-tracking data obtained from three presentations of clip 38.

Conversely, when the correlation values are lower, as observed from the examples of clips 21 and 30, the visual exploration is more evenly spread throughout the entire video presentation window and appears to be at least partially directed towards high-level elements of the scene, such as animals (*Figure 24*).

Figure 24. **(A)**. Frame of clip 21 (average correlation value = 0.09) **(B)**. Normalized heatmaps of eye-tracking data obtained from three presentations of clip 21. **(C)**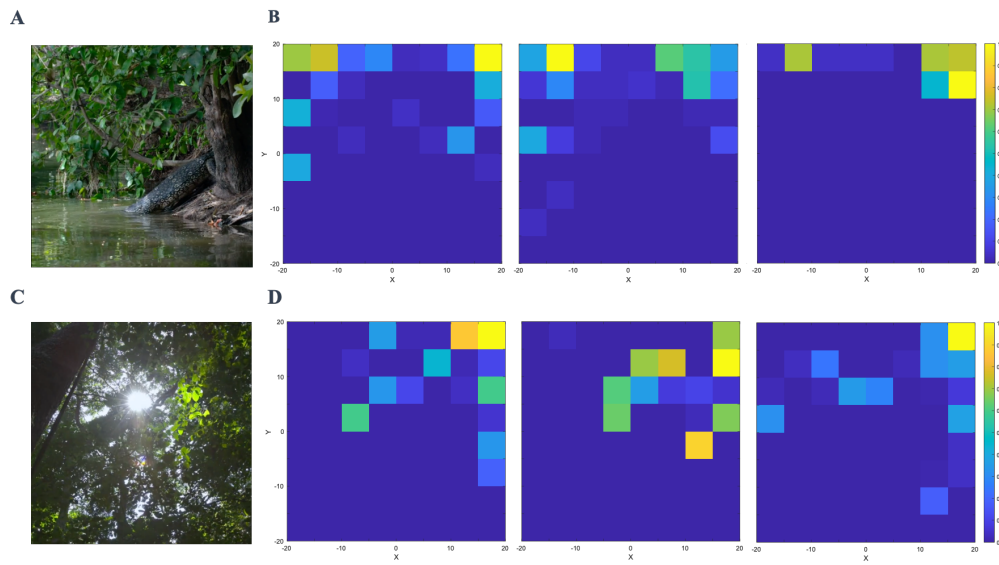. Frame of clip 30 (average correlation value = 0.16) **(D)**. Normalized heatmaps of eye-tracking data obtained from three presentations of clip 30.

In the last category, depicting various types of non-biological movement, the average level of correlation is the highest among all blocks. As observed for much of the *Natural* category, elevated correlation values are associated with high gaze time at the edges and corners of the video window, where neither relevance nor perceptual salience is high (*Figure 25* (**A**, **B**, **C**, **D**)). In the few cases where the correlation level is low, a notable dual pattern is observed: the gaze is still drawn to the periphery of the screen, but there's also a more distributed screen exploration. This includes attention to stimuli that, despite having different cinematic characteristics, mimic biological movement (*Figure 25* (**E**, **F**)). Similar to what was noted in the *Natural* category, this dual focus—both on the periphery and on the relevant stimuli—results in more moderate correlation values, likely due to a broader and more comprehensive visual exploration of the screen.
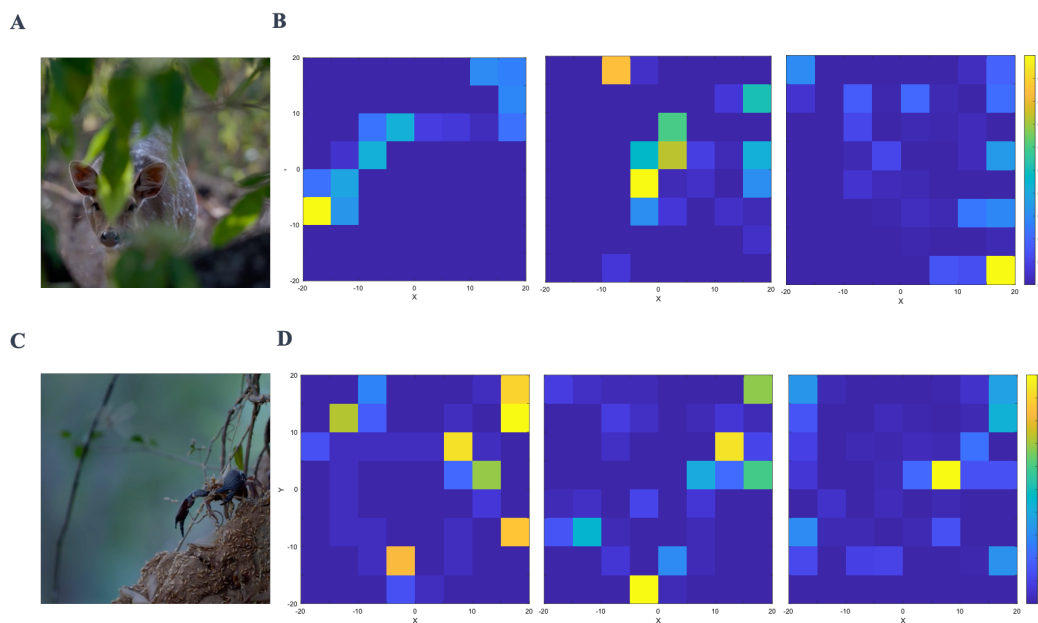

Figure 25. (**A**). Frame of clip 51 (average correlation value = 0.41) (**B**). Normalized heatmaps of eye-tracking data obtained from three presentations of clip 51. (**C**). Frame of clip 56 (average correlation value = 0.48) (**D**). Normalized heatmaps of eye-tracking data obtained from 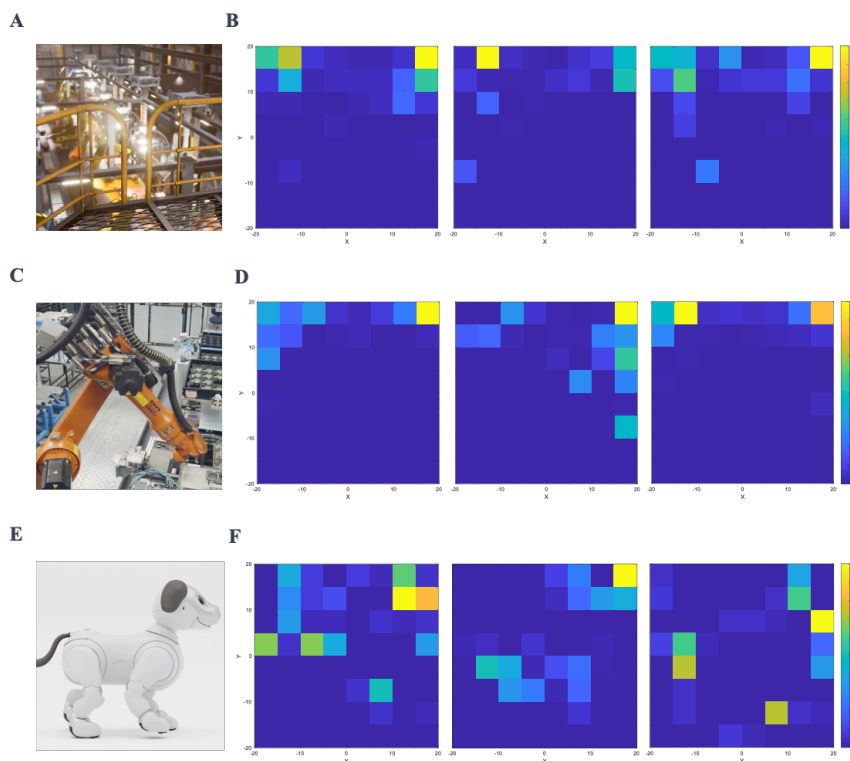three presentations of clip 56. (**E**). Frame of clip 49 (average correlation value = 0.13) (**F**). Normalized heatmaps of eye-tracking data obtained from three presentations of clip 49.

To summarize, in the LAE sessions, the ocular behavior appears to be only partially directed towards salient or relevant portions of the visual stimuli and exhibits a consistent behavior of averting the gaze from the central video content to the periphery, particularly the edges and corners of the screen. This behavior results in high correlation values both between repetitions of the same clip and between presentations of different clips within the same category and across different categories. Specifically, in the *Monkey* category, despite the gaze often being drawn to the screen's periphery, there is a significant amount of gaze time directed towards relevant scene elements like monkey bodies. In the *Natural* category, the gaze pattern varies— sometimes focusing on important scene elements or stimuli with high perceptual salience and other times fixating on less informative, peripheral areas. In contrast, the *Non-Bio* category shows a more pronounced tendency for the gaze to wander to the same peripheral screen areas, especially the top right corner, irrespective of the stimulus. The analysis of heatmaps derived from the LAE session suggests that even with low attentive engagement, high-level stimuli such as conspecific bodies or animals, as well as salient low-level features, still draw attention. This effect is particularly prominent in clips featuring conspecifics and, to a lesser degree, in natural-themed clips, resulting in a varied and stimulus-specific gaze pattern. Conversely, in the *Non-Bio* category, a stereotypical gaze pattern focusing on peripheral screen areas appears to occur, irrespective of the clip content. Taken together, these findings suggest that the observed differences may also be associated with differences in gaze duration within the video presentation window across categories, as it is likely that the numerous ocular signals recorded at the edges or corners may represent saccades directed outside the video window, as observed in some clips of the HAE session.

4.3. Category-related attention analysis

To assess potential differences in gaze duration, and thus in attentional engagement, across different categories, a paired-samples t-test was conducted on both the raw data recorded within the video window and on the fixation clusters extracted from them (see **Materials and Methods 3.5.3.**). Due to the higher variability observed in the LAE session, the *Natural* category was excluded from the comparison, which focused solely on the *Monkey* and *Non-Bio* categories, where more pronounced differences were observed.

4.3.1. High Attentional Engagement Session

In both session, the average looking time was determined by averaging the number of eye position signals recorded across the 5 presentations of each clip. In the HAE session, the average looking time for the *Monkey* category averaged 599 eye position signals (SD = 131), while for the *Non-Bio* category, it was 595 eye position signals (SD = 100), with a p-value of the t-test being 0.92654 (see *Figure 26* (**A**)). Additionally, the mean number of fixation clusters, calculated by averaging across the 5 presentations of each clip, was 5.1 (SD = 1.7) for the *Monkey* category, and 4.6 (SD = 1.8) for the *Non-Bio* category, with a p-value of the t-test being 0.30922 (see *Figure 26* (**B**)). These results suggest that under conditions of high attentive engagement, attention is equally directed towards clips belonging to the two different categories.

4.3.2. Low Attentional Engagement Session

In the LAE session, the average looking time for clips belonging to the *Monkey* category had a mean value of 475 eye position signals (SD = 83), while the *Non-Bio* category had a mean value of 413 eye position signals (SD = 72). The t-test yielded a p-value of 0.028781, indicating a significant difference between the average looking times of the two groups (see *Figure 26* (**A**)).

Since the p-value is below the established alpha level of 0.05, we reject the null hypothesis that there is no difference between the means, concluding that the difference observed in the average looking times between the *Monkey* and *Non-Bio* categories is statistically significant. Moving to the t-test conducted with the fixation clusters, the vector of values corresponding to the *Monkey* category had a mean of 2.5 fixation clusters per clip presentation (SD = 1.1), while that of the *Non-Bio* category had a mean of 1.5 fixation clusters (SD = 1.1). The paired-samples t-test produced a p-value of 0.017448, which is below the alpha threshold of 0.05, indicating a significant difference between the averages of the two groups (see *Figure 26* (**B**)). The results obtained from comparing the number of fixation clusters, which delineate the areas of focused attention and provide a concrete measure of visual engagement, further confirm the discrepancy already observed for the average looking time between the *Monkey* and *Non-Bio* categories, suggesting that clips depicting monkey-relevant content and those depicting non-biological content have a different influence on attentional deployment, which appears to be stronger for content depicting conspecifics, manifesting consistently in both the HAE and LAE sessions (*Figure 27*).
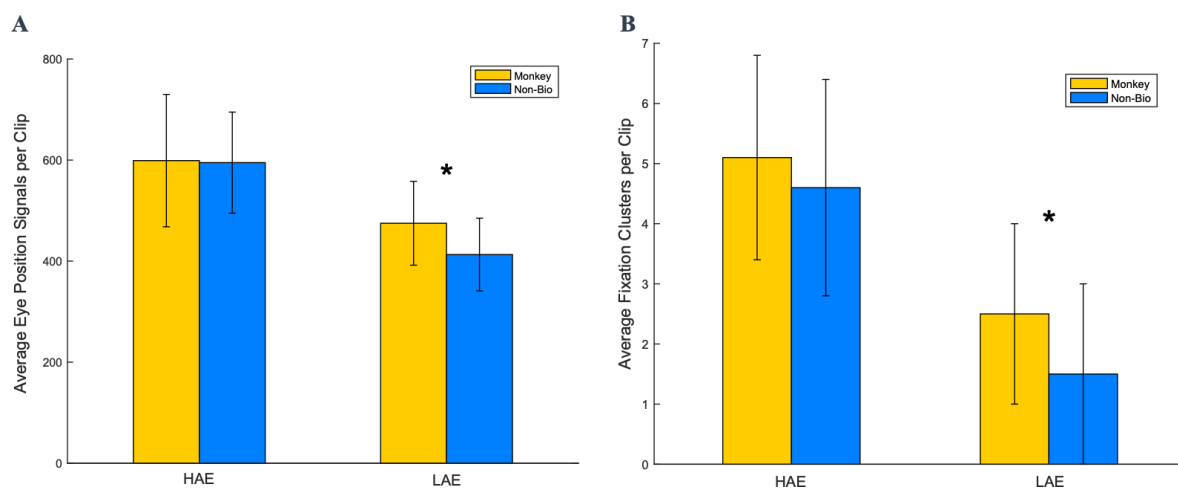
Figure 26. (**A**). Average eye position signals per clip for the Monkey category (in yellow) and Non-Bio category (in blue), recorded during sessions with high attentive engagement (HAE) and low attentive engagement (LAE), respectively. (**B**). Average fixation clusters per clip for the Monkey category (in yellow) and Non-Bio category (in blue), recorded during sessions with high attentive engagement (HAE) and low attentive engagement (LAE), respectively.

Figure 27. **(A)**. Clusters of fixations extracted from the 5 repetitions of clip 6 in the HAE session, superimposed on a frame of the clip (39 fixation clusters detected). **(B)**. Clusters of fixations extracted from the 5 repetitions of clip 55 in the HAE session, superimposed on a frame of the clip (31 fixation clusters detected). **(A)**. Clusters of fixations extracted from the 5 repetitions of clip 6 in the LAE session, superimposed on a frame of the clip (19 fixation clusters detected). **(B)**. Clusters of fixations extracted from the 5 repetitions of clip 55 in the LAE session, superimposed on a frame of the clip (10 fixation clusters detected).

5. DISCUSSION AND CONCLUSIONS

The analyses conducted clearly suggest that the ocular behavior of macaques during free viewing of natural videos is strongly influenced by both the content of the visual stimuli and the subject's endogenous level of attention. When the overall attentional engagement is high (79%), visual exploration is sustained and distributed, and the ocular pattern exhibited during the viewing of the clips appears to be reproducible only among different presentations of certain clips and does not correlate between different clips within the same category or across different categories. When viewing clips featuring monkeys, the gaze appears to be captured by conspecifics body parts, biologically and socially relevant stimuli whose power to attract attention and gaze has been observed in various species of non-human primates as well as in humans (Kano et al., 2018; Shepherd et al., 2010). The tendency to observe conspecifics, however, is not necessarily associated with reproducible ocular patterns, as these seem to depend on additional factors such as scene dynamism and the number of depicted elements. Significantly high correlation values were observed between different presentations of clips characterized by minimalistic scenes with fewer elements and little dynamism in their framing. Conversely, clips depicting dynamic scenes enriched with numerous relevant elements yielded unpredictable and diverse patterns of visual engagement, likely due to the competing attentional demands of these elements, resulting in a different focus of attention with each viewing. The viewing of scenes depicting natural content such as animals or natural landscapes is also associated with ocular and attentional patterns that are rarely reproducible. The more consistent oculomotor patterns appear to be due to the presence of heterospecific bodies, along with perceptual saliency factors like luminance contrast, movement, and color, which predominantly capture attention. These factors are widely recognized as influential in attentional capture across various primate species (Kano & Tomonaga, 2011; Parkhurst et al., 2002; Segraves et al., 2017;

Wilming et al., 2017). In line with the two categories analyzed so far, the non-biological category is also associated with ocular and attentional patterns whose reproducibility across different presentations of the same content is generally low, with some exceptions. In these cases, elements that consistently capture attention, eliciting correlated ocular patterns, include perceptual saliency factors such as movement, color, and luminance. It is plausible that within this category, lacking high-level elements like conspecifics or animals, saliency factors are the only bottom-up factors contributing to making eye movement reproducible. In summary, under conditions of high attentional engagement, visual exploration is thorough and is reproducible only among different presentations of clips that exhibit certain characteristics. Clips that display highly reproducible patterns are those containing few static high-level relevant elements, such as conspecific and heterospecific bodies, whose ability to capture attention during movie viewing has already been documented in the literature (Kano et al., 2018; Shepherd et al., 2010). Additionally, perceptually salient elements such as color, luminance, and movement, whose influence on macaque ocular behavior has been demonstrated in image search task (Segraves et al., 2017; Wilming et al., 2017), also contribute to these patterns. Moreover, reproducible patterns may occur, especially in the *Natural* and *Non-Bio* categories, due to stereotyped eye movements directed outside the screen, likely attributable to these categories' lesser ability to sustain attention.

In conditions of low attentive engagement, ocular behavior appears much more consistent, as suggested by the correlation matrix showing high values both among different presentations of the same clip and among presentations of different clips belonging to the same or different categories. Analysis of individual heatmaps suggests that these high values may be due to a stereotyped pattern, independent of the presented stimulus, directed towards peripheral portions of the screen. This could be due to a resting state or to the tendency for ocular signals

to concentrate around the screen edge when engagement is low. However, although this pattern is highly frequent across all three categories, the analysis of heatmaps suggests that the effect of certain factors in capturing attention persists. During the viewing of clips depicting monkeys, attention and gaze often appear to be captured by conspecifics. In the *Natural* category, the pattern is quite unpredictable and variable, sometimes showing attentive capture by salient or relevant factors, while in other cases, these factors are entirely ignored. In the non-biological content category, the pattern is more homogeneous, as evidenced by consistently high correlation values across the entire category. In this category more than in others, the tendency to drift gaze towards the periphery is prominent, presumably due to lower attentive capture elicited by the depicted stimuli, which, being non-biological in nature, likely hold less relevance, and the salient elements, if they capture attention, do so in a less sustained manner (Donk & Van Zoest, 2008). The analysis of heatmaps, on one hand, highlights that even with low attentive engagement, high-level stimuli such as conspecific bodies or animals, and salient low-level features still draw attention, confirming findings from existing literature (Kano et al., 2018; Segraves et al., 2017; Shepherd et al., 2010; Wilming et al., 2017). On the other hand, it suggests that high correlation values may be due to low attentional engagement, manifesting in a stereotyped "inattentive" pattern whose characteristics are not related to the spatial configuration of the presented stimulus. The fact that these extreme correlation values are mostly present for presentations of clips depicting non-biological content and in some clips of the natural category, suggests that these stimuli are those that, during the LAE session, were also the least attended and watched. The subsequent comparison of both the average gaze time and the number of detected clusters, conducted for both the HAE session and the LAE session, seems to be consistent with what was inferred from the analysis of correlations and heatmaps. The statistical analyses, in which the *natural* category was excluded due to its variability shown

Discussion and Conclusions

in the LAE session, revealed that while in the HAE session, both the gaze time and the number of fixations recorded for monkey content are overlapping with those obtained from the *non-bio* category, in the LAE session they are significantly different, and higher in the *monkey* category. These results suggest that, while under conditions of high attentive engagement, the level of attention was mostly the same between the two categories, albeit likely attracted by different factors, under conditions of low attentive engagement, scenes depicting conspecifics were looked at for a longer time and more attentively, generating correlation values that are generally lower than those obtained in the *non-bio* category, due to stimulus-driven and therefore more varied visual exploration.

In conclusion, the findings of this study are consistent with numerous pieces of evidence in the literature showing that both high-level factors, such as conspecific bodies, and low-level perceptual factors, such as salience, can contribute to attentional deployment in primates (Kano et al., 2018; Kano & Tomonaga, 2011; Segraves et al., 2017; Wilming et al., 2017). Notably, these factors can elicit highly reproducible ocular patterns, particularly in static, simple scenes or when salience levels are high. Moreover, by considering the level of attentional engagement, our study sheds light on the differential exploration of non-biological visual stimuli compared to monkey-relevant contents. This suggests a prioritized role for conspecifics and socially relevant stimuli in directing attention and eye movements, over low-level visual factors. These findings underscore the critical importance of paradigm selection in visual attention research. A free-viewing paradigm, as demonstrated, unveils category-level attentional differences that remain obfuscated in controlled settings, where trial-specific rewards drive attention. This method offers clear insights into attentional mechanisms, contrasting with the indirect inferences drawn from error or reaction time analyses in more constrained paradigms. Furthermore, the choice of naturalistic, complex stimuli allows for ecologically relevant

exploration of scenes, shedding light on attentional mechanisms in a more realistic context. The presentation of diverse stimuli, encompassing various colors, luminance, movements, and contrasts, mirrors the complexity of real-world visual environments and contributes to a comprehensive understanding of the visuo-oculomotor system's functioning. Conversely, the use of simple artificial stimuli may lead to impoverished and stereotypical responses at both the behavioral and neuronal levels.

However, the results obtained from this study have limitations. Firstly, it is important to mention the use of only two sessions, which, although involving numerous repetitions of the same clips presented in a randomized order between each video, may constitute a limited sample size for fully understanding the effect that different types of stimuli can have on attentional orientation and eye behavior. Additionally, it should be noted that the two sessions were analyzed separately to compare different levels of attentional engagement. To conduct more robust analyses, it would therefore be advisable to select a larger number of sessions both with HAE and LAE. Moreover, given that this study served as a preliminary, habituation phase for upcoming neural acquisition sessions, which will employ varied and new video stimuli, the repeated exposure to identical content across days might have induced familiarity. This could influence not only attentive engagement but also the intricacy and persistence of visual exploration.

Considering that the present research anticipates the investigation of neural responses during free viewing of natural movies, the next step is to integrate all the evidence collected here into future analyses, both behavioral and neural. Given that increasing the dataset is a necessary step, the subsequent phases of this study should first involve characterizing both the high-level profile, concerning animals and, more generally, objects, and the low-level profile, quantifying at each pixel and at each time instant, the levels of luminance, color, motion,

contrast, and direction, drawing from object-based and saliency-based models already present in the literature. All these features can be used to explain attentional deployment and eye behavior, allowing for a deeper understanding of the findings in the present study, also through the execution of unsupervised clustering. In addition to this, we could incorporate a distinction between sessions featuring novel content—where videos have been presented fewer times—and those with familiar content—where multiple sessions have already taken place. Considering the combined effects of attentive engagement, the characteristics of high and low-level features, and the degree of scene familiarity could be crucial for neuronal responses within structures of interest such as the superior colliculus and the pulvinar. While the superior colliculus has been examined in free-viewing contexts, its activities have been characterized based on the salience responsiveness of its receptive fields (White, Berg, et al., 2017). By integrating higher-level, object-based factors along with metrics of overall attention and familiarity, a more nuanced understanding of this structure's contributions to visual processing and attentional deployment can be achieved. Regarding the pulvinar, while it's widely agreed that it represents a crucial hub in attentional allocation, our understanding of this thalamic nucleus remains limited. Therefore, investigating the pulvinar through free-viewing paradigms is not just a novel approach but a necessary one to reveal its complex response properties, which have never been assessed through natural vision paradigms.

REFERENCES

Aboharb, F., Serene, S., Sliwa, J., & Freiwald, W. A. (2023). Mixed Multi-Level Visual, Reward, and Motor Signals in Dorsomedial Frontal Cortex Area F7 during Active Naturalistic Video Exploration. *bioRxiv*, 2023.2009. 2025.559420.

Allen, K. M., Lawlor, J., Salles, A., & Moss, C. F. (2021). Orienting our view of the superior colliculus: specializations and general functions. *Current opinion in neurobiology*, *71*, 119-126.

Awh, E., Armstrong, K. M., & Moore, T. (2006). Visual and oculomotor selection: links, causes and implications for spatial attention. *Trends in cognitive sciences*, *10*(3), 124-130.

Azadi, R., Lopez, E., Taubert, J., Patterson, A., & Afraz, A. (2024). Inactivation of face-selective neurons alters eye movements when free viewing faces. *Proceedings of the National Academy of Sciences*, *121*(3), e2309906121.

Berg, D. J., Boehnke, S. E., Marino, R. A., Munoz, D. P., & Itti, L. (2009). Free viewing of dynamic stimuli by humans and monkeys. *Journal of vision*, *9*(5), 19-19.

Bisley, J. W., & Goldberg, M. E. (2006). Neural correlates of attention and distractibility in the lateral intraparietal area. *Journal of neurophysiology*, *95*(3), 1696-1717.

Bisley, J. W., & Mirpour, K. (2019). The neural instantiation of a priority map. *Current opinion in psychology*, *29*, 108-112.

Bisley, J. W., Mirpour, K., Arcizet, F., & Ong, W. S. (2011). The role of the lateral intraparietal area in orienting attention and its implications for visual search. *European Journal of Neuroscience*, *33*(11), 1982-1990.

Blatt, G. J., Andersen, R. A., & Stoner, G. R. (1990). Visual receptive field organization and cortico-cortical connections of the lateral intraparietal area (area LIP) in the macaque. *Journal of Comparative Neurology*, *299*(4), 421-445.

Blignaut, P. (2009). Fixation identification: The optimum threshold for a dispersion algorithm. *Attention, Perception, & Psychophysics*, *71*, 881-895.

Borji, A., Sihite, D. N., & Itti, L. (2012). Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study. *IEEE Transactions on Image Processing*, *22*(1), 55-69.

Borji, A., Sihite, D. N., & Itti, L. (2013). What stands out in a scene? A study of human explicit saliency judgment. *Vision research*, *91*, 62-77.

Bringmann, A., Syrbe, S., Görner, K., Kacza, J., Francke, M., Wiedemann, P., & Reichenbach, A. (2018). The primate fovea: structure, function and development. *Progress in retinal and eye research*, *66*, 49-84.

Bruce, C. J., & Goldberg, M. E. (1985). Primate frontal eye fields. I. Single neurons discharging before saccades. *Journal of neurophysiology*, *53*(3), 603-635.

Caggiano, V., Fogassi, L., Rizzolatti, G., Pomper, J. K., Thier, P., Giese, M. A., & Casile, A. (2011). View-based encoding of actions in mirror neurons of area f5 in macaque premotor cortex. *Current Biology*, *21*(2), 144-148.

Carpenter, R. H. (2000). The neural control of looking. *Current Biology*, *10*(8), R291-R293.

Cohen, B., & Henn, V. (1972). Unit activity in the pontine reticular formation associated with eye movements. *Brain research*, *46*, 403-410.

Colby, C. L., Duhamel, J.-R., & Goldberg, M. E. (1996). Visual, presaccadic, and cognitive activation of single neurons in monkey lateral intraparietal area. *Journal of neurophysiology*, *76*(5), 2841-2852.

Connor, C. E., Egeth, H. E., & Yantis, S. (2004). Visual attention: bottom-up versus top-down. *Current biology*, *14*(19), R850-R852.

Corbetta, M. (1998). Frontoparietal cortical networks for directing attention and the eye to visual locations: identical, independent, or overlapping neural systems? *Proceedings of the National Academy of Sciences*, *95*(3), 831-838.

Corbetta, M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Ollinger, J. M., Drury, H. A., . . . Van Essen, D. C. (1998). A common network of functional areas for attention and eye movements. *Neuron*, *21*(4), 761-773.

Cortes, N., Ladret, H. J., Abbas-Farishta, R., & Casanova, C. (2023). The pulvinar as a hub of visual processing and cortical integration. *Trends in Neurosciences*.

Curtis, C. E., & Connolly, J. D. (2008). Saccade preparation signals in the human frontal and parietal cortices. *Journal of Neurophysiology*, *99*(1), 133-145.

Cynader, M., & Berman, N. (1972). Receptive-field organization of monkey superior colliculus. *Journal of Neurophysiology*, *35*(2), 187-201.

Daroff, R. B., & Aminoff, M. J. (2014). *Encyclopedia of the neurological sciences*. Academic press.

David, S. V., Vinje, W. E., & Gallant, J. L. (2004). Natural stimulus statistics alter the receptive field structure of v1 neurons. *Journal of Neuroscience*, *24*(31), 6991-7006.

DiCarlo, J. J., & Maunsell, J. H. (2000). Form representation in monkey inferotemporal cortex is virtually unaltered by free viewing. *Nature neuroscience*, *3*(8), 814-821.

Donk, M., & Van Zoest, W. (2008). Effects of salience are short-lived. *Psychological Science*, *19*(7), 733-739.

Dragoi, V., & Sur, M. (2006). Image structure at the center of gaze during free viewing. *Journal of Cognitive Neuroscience*, *18*(5), 737-748.

Einhäuser, W., Spain, M., & Perona, P. (2008). Objects predict fixations better than early saliency. *Journal of vision*, *8*(14), 18-18.

End, A., & Gamer, M. (2017). Preferential processing of social features and their interplay with physical saliency in complex naturalistic scenes. *Frontiers in Psychology*, *8*, 418.

Evans, K. K., Horowitz, T. S., Howe, P., Pedersini, R., Reijnen, E., Pinto, Y., . . . Wolfe, J. M. (2011). Visual attention. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(5), 503-514.

Everling, S., Paré, M., Dorris, M. C., & Munoz, D. P. (1998). Comparison of the discharge characteristics of brain stem omnipause neurons and superior colliculus fixation neurons in monkey: implications for control of fixation and saccade behavior. *Journal of neurophysiology*, *79*(2), 511-528.

Felsen, G., & Dan, Y. (2005). A natural approach to studying vision. *Nature neuroscience*, *8*(12), 1643-1646.

Felsen, G., Touryan, J., Han, F., & Dan, Y. (2005). Cortical sensitivity to visual features in natural scenes. *PLoS biology*, *3*(10), e342.

Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of vision*, *8*(2), 6-6.

Froesel, M., Cappe, C., & Hamed, S. B. (2021). A multisensory perspective onto primate pulvinar functions. *Neuroscience & Biobehavioral Reviews*, *125*, 231-243.

Gallant, J. L., Connor, C. E., & Van Essen, D. C. (1998). Neural activity in areas V1, V2 and V4 during free viewing of natural scenes compared to controlled viewing. *Neuroreport*, *9*(7), 1673-1678.

Gandhi, N. J., & Katnani, H. A. (2011). Motor functions of the superior colliculus. *Annual review of neuroscience*, *34*, 205-231.

Girard, B., & Berthoz, A. (2005). From brainstem to cortex: computational models of saccade generation circuitry. *Progress in neurobiology*, *77*(4), 215-251.

Goffart, L., Hafed, Z. M., & Krauzlis, R. J. (2012). Visual fixation as equilibrium: evidence from superior colliculus inactivation. *Journal of Neuroscience*, *32*(31), 10627-10636.

Goldberg, M. E., & Wurtz, R. H. (1972). Activity of superior colliculus in behaving monkey. II. Effect of attention on neuronal responses. *Journal of neurophysiology*, *35*(4), 560-574.

Hikosaka, O., Sakamoto, M., & Usui, S. (1989). Functional properties of monkey caudate neurons. I. Activities related to saccadic eye movements. *Journal of neurophysiology*, *61*(4), 780-798.

Hikosaka, O., Takikawa, Y., & Kawagoe, R. (2000). Role of the basal ganglia in the control of purposive saccadic eye movements. *Physiological reviews*, *80*(3), 953-978.

Hikosaka, O., & Wurtz, R. H. (1983). Visual and oculomotor functions of monkey substantia nigra pars reticulata. IV. Relation of substantia nigra to superior colliculus. *Journal of neurophysiology*, *49*(5), 1285-1301.

Ibbotson, M., & Krekelberg, B. (2011). Visual perception and saccadic eye movements. *Current opinion in neurobiology*, *21*(4), 553-558.

Im, M., & Fried, S. I. (2016). Directionally selective retinal ganglion cells suppress luminance responses during natural viewing. *Scientific reports*, *6*(1), 35708.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature reviews neuroscience*, *2*(3), 194-203.

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.

Jonikaitis, D., & Moore, T. (2019). The interdependence of attention, working memory and gaze control: behavior and neural circuitry. *Current opinion in psychology*, *29*, 126-134.

Kandel, E. R., Koester, J. D., Mack, S. H., & Siegelbaum, S. A. (2021). *Principles of Neural Science, Sixth Edition*. McGraw Hill LLC.

Kano, F., Shepherd, S. V., Hirata, S., & Call, J. (2018). Primate social attention: Species differences and effects of individual experience in humans, great apes, and macaques. *PloS one*, *13*(2), e0193283.

Kano, F., & Tomonaga, M. (2011). Perceptual mechanism underlying gaze guidance in chimpanzees and humans. *Animal cognition*, *14*, 377-386.

Karamanlis, D., Schreyer, H. M., & Gollisch, T. (2022). Retinal encoding of natural scenes. *Annual Review of Vision Science*, *8*, 171-193.

Keller, E. (1974). Participation of medial pontine reticular formation in eye movement generation in monkey. *Journal of neurophysiology*, *37*(2), 316-332.

Killian, N. J., Jutras, M. J., & Buffalo, E. A. (2012). A map of visual space in the primate entorhinal cortex. *Nature*, *491*(7426), 761-764.

King, A. J. (2004). The superior colliculus. *Current Biology*, *14*(9), R335-R338.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human neurobiology*, *4*(4), 219-227.

Koehler, K., Guo, F., Zhang, S., & Eckstein, M. P. (2014). What do saliency models predict? *Journal of vision*, *14*(3), 14-14.

Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision research*, *35*(13), 1897-1916.

Krassanakis, V., Filippakopoulou, V., & Nakos, B. (2014). EyeMMV toolbox: An eye movement post-analysis tool based on a two-step spatial dispersion threshold for fixation identification. *Journal of Eye Movement Research*, *7*(1).

Krassanakis, V., Filippakopoulou, V., & Nakos, B. (2016). Detection of moving point symbols on cartographic backgrounds. *Journal of eye movement research*, *9*(2).

Krauzlis, R. J., Goffart, L., & Hafed, Z. M. (2017). Neuronal control of fixation and fixational eye movements. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1718), 20160205.

Krauzlis, R. J., Lovejoy, L. P., & Zénon, A. (2013). Superior colliculus and visual spatial attention. *Annual review of neuroscience*, *36*, 165-182.

Kummerer, M., Wallis, T. S., Gatys, L. A., & Bethge, M. (2017). Understanding low-and high-level contributions to fixation prediction. Proceedings of the IEEE international conference on computer vision,

Kustov, A. A., & Lee Robinson, D. (1996). Shared neural control of attentional shifts and eye movements. *Nature*, *384*(6604), 74-77.

Kümmerer, M., & Bethge, M. (2023). Predicting visual fixations. *Annual Review of Vision Science*, *9*, 269-291.

Land, M. (2019). Eye movements in man and other animals. *Vision research*, *162*, 1-7.

Land, M., & Tatler, B. (2009). *Looking and acting: Vision and eye movements in natural behaviour*. Oxford University Press.

Land, M. F. (1999). Motion and vision: why animals move their eyes. *Journal of Comparative Physiology A*, *185*, 341-352.

Land, M. F. (2011). 2 Oculomotor behaviour in vertebrates and invertebrates. In S. P. Liversedge, I. Gilchrist, & S. Everling (Eds.), *The Oxford Handbook of Eye Movements* (pp. 0). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199539789.013.0001

Leopold, D. A., & Park, S. H. (2020). Studying the visual brain in its natural rhythm. *Neuroimage*, *216*, 116790.

Martin, D., Fowlkes, C., Tal, D., & Malik, J. (2001). A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001,

May, P. J. (2006). The mammalian superior colliculus: laminar structure and connections. *Progress in brain research*, *151*, 321-378.

McFarland, R., Roebuck, H., Yan, Y., Majolo, B., Li, W., & Guo, K. (2013). Social interactions through the eyes of macaques and humans. *PLoS One*, *8*(2), e56437.

McMahon, D. B., Russ, B. E., Elnaiem, H. D., Kurnikova, A. I., & Leopold, D. A. (2015). Single-unit activity during natural vision: diversity, consistency, and spatial sensitivity among AF face patch neurons. *Journal of neuroscience*, *35*(14), 5537-5548.

Moore, T., Armstrong, K. M., & Fallah, M. (2003). Visuomotor origins of covert spatial attention. *Neuron*, *40*(4), 671-683.

Moore, T., & Fallah, M. (2001). Control of eye movements and spatial attention. *Proceedings of the National Academy of Sciences*, *98*(3), 1273-1276.

Mosher, C. P., Zimmerman, P. E., & Gothard, K. M. (2014). Neurons in the monkey amygdala detect eye contact during naturalistic social interactions. *Current Biology*, *24*(20), 2459-2464.

Munoz, D. P., & Everling, S. (2004). Look away: the anti-saccade task and the voluntary control of eye movement. *Nature Reviews Neuroscience*, *5*(3), 218-228.

Munoz, D. P., & Wurtz, R. H. (1993). Fixation cells in monkey superior colliculus. I. Characteristics of cell discharge. *Journal of neurophysiology*, *70*(2), 559-575.

Noton, D., & Stark, L. (1971). Eye movements and visual perception. *Scientific American*, *224*(6), 34-43.

Papale, P., Leo, A., Cecchetti, L., Handjaras, G., Kay, K. N., Pietrini, P., & Ricciardi, E. (2018). Foreground-background segmentation revealed during natural image viewing. *eneuro*, *5*(3).

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision research*, *42*(1), 107-123.

Peters, R. J., Iyer, A., Itti, L., & Koch, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision research*, *45*(18), 2397-2416.

Powell, K. D., & Goldberg, M. E. (2000). Response of neurons in the lateral intraparietal area to a distractor flashed during the delay period of a memory-guided saccade. *Journal of Neurophysiology*, *84*(1), 301-310.

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), 31-40.

Robinson, D. L., & Petersen, S. E. (1992). The pulvinar and visual salience. *Trends in neurosciences*, *15*(4), 127-132.

Rubo, M., & Gamer, M. (2018). Social content and emotional valence modulate gaze fixations in dynamic scenes. *Scientific reports*, *8*(1), 3804.

Russ, B. E., & Leopold, D. A. (2015). Functional MRI mapping of dynamic visual features during natural viewing in the macaque. *Neuroimage*, *109*, 84-94.

Sakon, J. J., & Suzuki, W. A. (2021). Neural evidence for recognition of naturalistic videos in monkey hippocampus. *Hippocampus*, *31*(8), 916-932.

Schiller, P. H., & Stryker, M. (1972). Single-unit recording and stimulation in superior colliculus of the alert rhesus monkey. *Journal of neurophysiology*, *35*(6), 915-924.

Scudder, C. A., Kaneko, C. R., & Fuchs, A. F. (2002). The brainstem burst generator for saccadic eye movements: a modern synthesis. *Experimental brain research*, *142*, 439-462.

Segraves, M. A., & Goldberg, M. E. (1987). Functional properties of corticotectal neurons in the monkey's frontal eye field. *Journal of Neurophysiology*, *58*(6), 1387-1419.

Segraves, M. A., Kuo, E., Caddigan, S., Berthiaume, E. A., & Kording, K. P. (2017). Predicting rhesus monkey eye movements during natural-image search. *Journal of Vision*, *17*(3), 12-12.

Shepherd, S. V., Steckenfinger, S. A., Hasson, U., & Ghazanfar, A. A. (2010). Human-monkey gaze correlations reveal convergent and divergent patterns of movie viewing. *Current Biology*, *20*(7), 649-656.

Shipp, S. (2004). The brain circuitry of attention. *Trends in cognitive sciences*, *8*(5), 223-230.

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual review of neuroscience*, *24*(1), 1193-1216.

Skripkauskaite, S., Mihai, I., & Koldewyn, K. (2023). Attentional bias towards social interactions during viewing of naturalistic scenes. *Quarterly Journal of Experimental Psychology*, *76*(10), 2303-2311.

Snowden, R., Snowden, R. J., Thompson, P., & Troscianko, T. (2012). *Basic vision: an introduction to visual perception*. Oxford University Press.

Sparks, D. L. (2002). The brainstem control of saccadic eye movements. *Nature Reviews Neuroscience*, *3*(12), 952-964.

Sparks, D. L., Holland, R., & Guthrie, B. L. (1976). Size and distribution of movement fields in the monkey superior colliculus. *Brain research*, *113*(1), 21-34.

Stanton, G. B., Goldberg, M., & Bruce, C. (1988). Frontal eye field efferents in the macaque monkey: I. Subcortical pathways and topography of striatal and thalamic terminal fields. *Journal of Comparative Neurology, 271*(4), 473-492.

Steinmetz, N. A., & Moore, T. (2012). Lumping and splitting the neural circuitry of visual attention. *Neuron, 73*(3), 410-412.

Tanaka, M., & Kunimatsu, J. (2011). Thalamic roles in eye movements. In S. P. Liversedge, I. Gilchrist, & S. Everling (Eds.), *The Oxford Handbook of Eye Movements* (pp. 0). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199539789.013.0013

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of vision, 11*(5), 5-5.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology, 12*(1), 97-136.

Van der Linden, L. (2018). *The influence of low-level visuomotor factors versus high-level cognitive factors on object viewing* AIX-MARSEILLE UNIVERSITÉ].

Veale, R., Hafed, Z. M., & Yoshida, M. (2017). How is visual salience computed in the brain? Insights from behaviour, neurobiology and modelling. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372*(1714), 20160113.

Wang, L., Zhou, X., Zeng, F., Cao, M., Zuo, S., Yang, J., . . . Kwok, S. C. (2023). Mixed selectivity coding of content-temporal detail by dorsomedial posterior parietal neurons. *Journal of Neuroscience*.

Wardak, C., Olivier, E., & Duhamel, J. R. (2011). The relationship between spatial attention and saccades in the frontoparietal network of the monkey. *European Journal of Neuroscience, 33*(11), 1973-1981.

White, B. J., Berg, D. J., Kan, J. Y., Marino, R. A., Itti, L., & Munoz, D. P. (2017). Superior colliculus neurons encode a visual saliency map during free viewing of natural dynamic video. *Nature communications, 8*(1), 14263.

White, B. J., Itti, L., & Munoz, D. P. (2021). Superior colliculus encodes visual saliency during smooth pursuit eye movements. *European Journal of Neuroscience, 54*(1), 4258-4268.

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, 114*(35), 9451-9456.

Wilke, M., Turchi, J., Smith, K., Mishkin, M., & Leopold, D. A. (2010). Pulvinar inactivation disrupts selection of movement plans. *Journal of Neuroscience, 30*(25), 8650-8659.

Wilming, N., Kietzmann, T. C., Jutras, M., Xue, C., Treue, S., Buffalo, E. A., & König, P. (2017). Differential contribution of low-and high-level image content to eye movements in monkeys and humans. *Cerebral Cortex, 27*(1), 279-293.

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature reviews neuroscience, 5*(6), 495-501.

Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour, 1*(3), 0058.

Wurtz, R. H., & Goldberg, M. E. (1972). Activity of superior colliculus in behaving monkey. 3. Cells discharging before eye movements. *Journal of Neurophysiology, 35*(4), 575-586.

Wurtz, R. H., Goldberg, M. E., & Robinson, D. L. (1982). Brain mechanisms of visual attention. *Scientific American, 246*(6), 124-135.

Yamane, Y., Ito, J., Joana, C., Fujita, I., Tamura, H., Maldonado, P. E., . . . Grün, S. (2023). Neuronal population activity in macaque visual cortices dynamically changes through repeated fixations in active free viewing. *Eneuro, 10*(10).