



UNIVERSITÀ DEGLI STUDI DI MILANO  
Facoltà di Scienze e Tecnologie  
*Corso di Laurea Magistrale in Informatica*

# AUDIO-VISUAL ATTENTION MODELING VIA REINFORCEMENT LEARNING

**Relatore:** Prof. Giuseppe BOCCIGNONE

**Correlatore:** Dr. Alessandro D'AMELIO

Tesi di:  
Daniele BOCCHINO  
Matricola: 991031

Anno Accademico 2022-2023

# Abstract

One of the essential components of effective human-human and human-computer interaction is the comprehension of human behavioral signals. Nonverbal communication, consisting of various modalities such as gesture, facial expression, prosody, and gaze behavior, plays a critical role in conveying a common message, on par with spoken words. Although many computational models of visual attention allocation have been proposed in recent decades, attention modeling has experienced a renaissance, driven by the increasing number of potential applications in areas such as autonomous driving, image/video compression, and robotics. However, some key features of eye movements, particularly their dynamics, have been overlooked in most of the proposed models, which can only handle static stimuli (e.g., images), with few exceptions. This thesis presents a new computational model of attentive eye guidance that formulates attention deployment as a stochastic foraging problem, deriving gaze dynamics in a principled manner. By treating a virtual observer attending to a video as a stochastic composite forager searching for valuable patches in a multi-modal landscape, the model produces simulated gaze trajectories that are statistically indistinguishable from those performed by humans while free-viewing the same scene. The proposed model is evaluated on a publicly available dataset of eye-tracked subjects engaged in conversations and social interactions between humans.

# Acknowledgments

È giunto il momento di scrivere la parte più complicata della mia tesi, in contrasto con i capitoli successivi, dove tutto è definito da formule, teoremi ed esperimenti. In questa sezione, devo esprimere emozioni e sentimenti provati negli anni che mi hanno portato ad essere qui oggi. Non è facile racchiudere anni di riconoscimenti in qualche centinaio di parole, ma farò del mio meglio.

Prima di procedere, voglio fare una premessa: chiedo scusa fin da subito a tutte quelle persone che non sono citate direttamente all'interno di questi ringraziamenti. Se leggerete queste parole, molto probabilmente siete parte integrante della mia vita e, come tali, desidero esprimere la mia gratitudine per aver contribuito, in un modo o nell'altro, al raggiungimento di questo grande traguardo.

Desidero innanzitutto ringraziare il mio relatore, il Professore Giuseppe Boccignone, per avermi dato l'opportunità di lavorare su una tesi tanto complessa quanto interessante, oltre ad avermi fornito supporto durante tutto il suo sviluppo.

Voglio anche ringraziare il mio correlatore, il Dottor Alessandro D'Amelio, per la grande disponibilità e professionalità dimostrate durante il periodo di tirocinio, oltre all'enorme aiuto che mi ha dato durante lo sviluppo del progetto e la stesura della tesi.

Inoltre, ci tengo a ringraziare tutti i membri del laboratorio PhuseLab, con cui ho trascorso questo periodo indimenticabile in cui ho potuto apprendere moltissimi insegnamenti, come la guida di barche a vela, le tecniche di degustazione vinicole, relativi esami pratici e la costruzione di zattere (ovviamente, ho imparato anche qualcosa su Machine Learning, computer vision, ecc...).

Rimanendo in ambito universitario, voglio ringraziare tutti i miei compagni e amici: Adri, Benia, Bozzo, Marcolino, Capu e tutti gli altri che mi hanno accompagnato e supportato in questi anni di università durante tornei di calcetto, partite di briscola chiamata e, qualche volta, seppure raramente, durante lezioni ed esami.

Desidero ringraziare Teo Limo per le ore spese a progettare la conquista del mondo tra certificazioni, statistiche aziendali ed eventi in cui fare networking, anche con il portinaio pur di ottenere un nuovo collegamento su LinkedIn.

Un grazie particolare va ad Andre per le innumerevoli giornate spese ad ascoltare e assecondare ogni mia idea, presentata come se potesse diventare la nuova scoperta del secolo!

Per quanto riguarda gli amici di vecchia data, ci tengo a ringraziare coloro che mi hanno visto crescere e sono cresciuti con me, sopportando tutte le mie stravaganze e passioni temporanee e non! In particolare, voglio ringraziare Teo, Leo, Enri, Luca, Tommy, Ale, Mara e tutti gli altri che da 15/20 anni mi conoscono e sopportano le mie più strane fissazioni, accompagnandomi in ogni momento ed in ogni contesto, dalle passeggiate in montagna, allo snowboard, dai computer alla Svizzera e via dicendo.

Ora, voglio ringraziare le persone che mi sono state più vicine nel quotidiano, che mi hanno appoggiato e supportato, non senza difficoltà, durante tutti questi anni.

In primis, devo ringraziare le persone che più di tutti mi hanno permesso di arrivare fino a qui. Mamma e Papà, che nonostante gli infiniti problemi che hanno dovuto affrontare a causa mia, hanno sempre creduto in me, dandomi la possibilità di conseguire questo titolo. Sono eternamente grato per avermi permesso di compiere errori ed imparare dai miei sbagli stando sempre pronti a intervenire e porre rimedio. Spero di essere riuscito, con questo mio traguardo, a renderli orgogliosi e fieri di quanto il duro lavoro fatto con me abbia portato alla fine ottimi frutti.

Vorrei ringraziare anche Marco per l'aiuto che mi ha fornito ininterrottamente negli ultimi anni, dalle "custom tactics" su Fifa agli esercizi di matematica. Desidero esprimere un ringraziamento particolare per la sua costante presenza e per avermi stimolato a riflettere quando non riuscivo a capire qualcosa. Senza che lui lo sapesse, è stato una grande fonte di ispirazione che mi ha spinto a migliorarmi e a dare sempre il massimo sotto il profilo accademico e non, fino al raggiungimento di questo importante traguardo.

Voglio ringraziare Nonna Piera, Nonno Pompeo, Nonno Girolamo e Nonna Antonietta perché, in qualità di nonni, ci sono sempre stati e in qualsiasi modo sempre ci saranno per me!

Voglio ringraziare Zia Ale perché, con il suo modo defilato, è sempre riuscita a fare il massimo per aiutarmi in tutte le mie necessità.

Desidero esprimere un ringraziamento speciale a Nietzsche, per le innumerevoli giornate passate ad ascoltarmi durante lo studio, rimanendo seduto e guardandomi perplesso mentre ripetevo a macchinetta una cascata di parole. Inoltre, voglio ringraziarlo per avermi insegnato il valore della perseveranza, trascorrendo ore e ore a scodinzolare solo per ottenere qualche "biscottino".

Voglio ringraziare Zia Rita, Zia Elisa, Zio Giacomo, Zio Michele e tutti gli zii che in questi anni mi hanno visto crescere e mi hanno incoraggiato ed ascoltato nelle mie scelte, gioendo orgogliosi per tutti i successi raggiunti.

Infine, ma non meno importante, voglio ringraziare la mia compagna Nicol, per il tempo dedicato all'ascolto di ogni mio problema, dubbio, piano o strategia. Per avermi rassicurato, e incoraggiato a dare sempre il massimo, confrontandomi ogni volta che le cose non andavano come sperato. Ma soprattutto un ringraziamento speciale per aver creduto in me in ogni momento, rimanendo sempre al mio fianco ed accompagnandomi in ogni mia scelta e decisione.

Concludendo, voglio ringraziare Daniele, per la forza d'animo, il coraggio e la determinazione che ha messo in ogni sfida e difficoltà che gli si è presentata. Solo lui sa quanto spesso è stata dura e quanto forte è stata la tentazione di mollare. Proprio per questo e per ciò che sei riuscito a raggiungere, goditi questo momento e continua così, perché oggi ce l'abbiamo fatta!

Questo importante traguardo è stato possibile solo grazie all'aiuto delle persone che mi hanno affiancato in questi anni. Non riesco a esprimere pienamente con queste parole l'enorme gratitudine che provo per tutti coloro che hanno contribuito a farmi diventare la persona che sono oggi, tuttavia spero che ciò basti a far comprendere quanto ne sia grato.

**Grazie a tutti dal profondo del cuore!**

# Indice

<b>Abstract</b>	ii
<b>Acknowledgments</b>	iii
<b>1 Introduction</b>	1
<b>2 Gaze Modeling</b>	5
2.1 Psychological and Computational Models of Visual Attention . . . . .	6
2.2 Computational Modeling of Visual Attention . . . . .	9
2.2.1 The Evolution of Saliency Models . . . . .	9
2.3 Overview of Saccadic Models for Gaze Shift Prediction on Static Stimuli	10
2.4 Summary . . . . .	15
<b>3 Theoretical Background</b>	17
3.1 Machine Learning . . . . .	17
3.2 Reinforcement Learning . . . . .	19
3.2.1 Multi-Armed Bandits . . . . .	21
3.2.2 Contextual Multi-Armed Bandits . . . . .	24
3.2.3 Algorithm for Contextual Bandits . . . . .	25
3.2.4 Comparison of Contextual Multi-Armed Bandit Algorithms . .	28
3.3 Stochastic models of eye movement . . . . .	28
3.3.1 Stochastic Processes . . . . .	30
3.4 Lévy Flights . . . . .	31
3.4.1 Fixational eye movement with Brownian motion . . . . .	33
3.4.2 Saccades as Lèvy Flights . . . . .	33
3.5 Foraging . . . . .	35
3.5.1 The Lévy Flight foraging hypothesis . . . . .	37
3.6 The Ornstein-Uhlenbeck model . . . . .	38
3.7 Marginal Value Theorem . . . . .	41
3.8 Summary . . . . .	43

<b>4 The OUTS-CMAB Gaze Model</b>	<b>44</b>
4.1 Model input: an environment of multimodal patches . . . . .	45
4.2 The spatial behaviour of the attentive forager: how to walk within and between patches . . . . .	46
4.2.1 Decision-making of the forager . . . . .	48
4.2.2 Optimal Foraging Strategies and the Marginal Value Theorem	51
4.3 Gaze model operationalisation: Context and Reward . . . . .	55
4.3.1 The Context Vector . . . . .	56
4.3.2 The reward function . . . . .	57
4.3.3 The OUTS-CMAB Gaze Model at the algorithmic level . . . . .	58
4.4 Summary . . . . .	61
<b>5 Results</b>	<b>62</b>
5.1 Dataset Description . . . . .	62
5.2 Qualitative Analysis . . . . .	63
5.2.1 Duration Distribution . . . . .	64
5.2.2 Multi Duration Distribution . . . . .	66
5.2.3 Saccade Amplitudes . . . . .	67
5.2.4 Saccade Direction . . . . .	69
5.3 Quantitative analysis . . . . .	70
5.3.1 Comparative Analysis of Simulated and Human-Recorded Scan Paths . . . . .	71
5.4 Summary . . . . .	73
<b>6 Main findings and conclusions</b>	<b>75</b>
6.1 Summary of Findings . . . . .	75
6.2 Implications and Significance . . . . .	76
6.3 Limitations and Challenges . . . . .	76
6.4 Future Research Directions . . . . .	77
6.5 Conclusion . . . . .	78

# Capitolo 1

## Introduction

In our daily lives we are immersed in a multitude of stimuli. These stimuli are generated by the environment in which humans live and can be originated from different sources. In particular, they can be visual stimuli, auditory stimuli, or other signals.

In this thesis we contend with the subtleties and hindrances behind the strategies that we, immersed in the Jamesian blooming, buzzing confusion [James, 1890] of our sensations, adopt to forage for those stimuli that are most relevant to our survival, conscious or unconscious goals, or just simply to keep up with the immaterial threads of our dreams.

Whenever a human being undergoes a stimulus that bears some importance for the human/world interaction context, his or her attention is deployed to such stimulus. This type of attention is called *selective attention*. Selective attention is a crucial mechanism that allows us to deal with the overwhelming amount of sensory information we encounter in our daily lives. It helps us to process and prioritize the most relevant information, while ignoring the rest. This selection ability exists for all our senses, markedly, the auditory, visual, and tactile sensing modalities.

The cocktail party effect is a well-known example of selective attention in auditory attention. In a noisy environment, we can selectively attend to a particular voice or sound and filter out the others. This ability to focus on a specific source of sound while ignoring others is known as auditory attention.

As to visual attention, a pivotal role is played by the fovea, the central region of the retina that has the highest resolution (i.e., density of cone photoreceptors) permitting us to see the scene details. Our oculomotor system allows us to move our eyes rapidly to keep the most important part of the visual landscape in the fovea while ignoring the rest of the stimuli. For example, if we are looking at a picture of a

flower in a garden, we may move our eyes to focus on the flower and ignore the other parts of the garden. This selective attention to the most relevant part of the visual scene is crucial for processing and interpreting visual information. Yet, experiments on change blindness have shown that humans can miss significant changes in a visual scene if they are not paying attention to such, sometimes conspicuous changes . This further underscores the nuances of selective attention in visual perception.

To create a rich representation of the surrounding world, humans use a "saccade and fixate" strategy: we rapidly move our eyes to new areas of interest and then hold our gaze steady. This pattern of eye movements allows us to build a comprehensive and detailed representation of the world around us. In addition, covert attention - the ability to attend to a particular location in the visual field without moving our eyes - also plays a role in complex vision tasks.

Research has shown that overt and covert attention mechanisms work together to process sensory information and perform complex cognitive tasks.

In particular, the concept of foraging play a fundamental role for the develop of Gaze Modeling. Foraging behavior is a fundamental aspect of an organism's survival and reproductive success, and understanding the decision-making processes and strategies employed in patchy environments is crucial for comprehending foraging efficiency and adaptation.

In this study, thus, we delve into the spatio-temporal foraging behavior of individuals in patchy (audio and visual) environments and investigate the decision-making process for leaving a patch based on the perceived reward rate. Such behavior is the essence of the model we devise in this Thesis, the OUTS-CMAB (Ornstein-Uhlenbeck Thompson Sampling-Contextual Multi Armed Bandits) gaze model.

The key research questions that we will address here can be crudely summarised as follows:

- *What constitutes a valuable audio-visual information patch? (RQ1)*
- *How is gaze guided within and between these patches? (RQ2)*

By incorporating the concept of the expected reward rate, we aim to gain valuable insights into how individuals optimize their foraging efficiency and adapt their behavior in response to changing environmental conditions.

The decision to leave a patch and explore new opportunities is influenced by various factors, including environmental richness, patch distance, and predation risk. To capture the essence of these considerations, we introduce the concept of the expected reward rate denoted as  $g(u, t)$ , which estimates the average reward a forager anticipates to gain within a specific time interval. This stochastic representation of the

continuous energy intake rate allows us to model the decision-making process more accurately, accounting for the inherent probabilistic nature of foraging.

To calculate the expected reward rate, we adopt the Bayesian foraging approach, taking into account the experiential state of the forager represented as  $u$ . In the case of a patch with a discrete number of items, we can express  $u$  as  $(n, t)$ , where  $n$  represents the number of items consumed within the time period  $t$ . By incorporating the remaining items,  $k = m - n$ , and the elapsed time,  $t_{W_p}$ , we can develop concrete methods to determine the optimal time for leaving a patch and seeking new foraging opportunities.

In our modeling approach, we also consider the time required to find an item, which follows an exponential distribution. This probability distribution allows us to account for the forager's searching efficiency, where the rate parameter  $\lambda = A_k$  depends on the forager's searching ability, denoted as  $A$ . Through our research, we gain insights into the relationship between  $A$  and the forager's behavior, revealing that individuals tend to concentrate their efforts in areas with higher rewards.

Furthermore, we incorporate the concept of handling time and its impact on the expected time until encountering the next item within a patch. By setting  $A$  as a function of the patch's value,  $v_p(t)$ , and a positive constant,  $\varphi$ , we consider the influence of patch quality on the forager's decision-making process. Additionally, we introduce a weighting factor,  $e^{-kdp}$ , which accounts for the cost of relocating between patches. Here,  $dp$  represents the distance from the current point of gaze to patch  $p$ . By integrating these factors, we gain a more comprehensive understanding of the determinants influencing the forager's choice to leave a patch.

To validate and assess the effectiveness of our model, we conduct experiments using a dataset comprising 78 videos capturing various foraging scenarios in patchy environments. The training phase involved 10 observers, while the remaining 29 observers were used as a test group for the experiments.

Notably, the proposed OUTS-CMAB model demonstrates the ability to learn and predict the focus of observers based on specific instructions, such as focusing on a particular part of the video, such as hands. This task-driven behavior analysis exemplifies the adaptability of our model and showcases its potential for real-world applications. In addition to the task-driven experiments, we also explored the model's adaptability to dynamic patches by allowing the addition and dropping of arms (patches) during the foraging process. While our model showcased its capacity to adjust its foraging strategies in response to changes in patch availability, it is important to note that retraining is necessary to learn and adapt to new patch configurations. We acknowledge this limitation and recognize the need for future research to enhance the model's flexibility in handling dynamic patches without compromising its performance. Overall, this study provides valuable insights into the foraging behavior of individuals in patchy environments. By incorporating the concept of the expected

reward rate, leveraging the Bayesian foraging approach, and considering task-driven behavior, we contribute to a deeper understanding of how individuals optimize their foraging efficiency. The experiments conducted on a diverse dataset, along with the exploration of adding and dropping patches, highlight the adaptability of our model and its potential for real-world applications. The limitations identified pave the way for future research endeavors to further enhance the model's capabilities and address dynamic environmental changes. This thesis unfolds as follows:

- Chapter 2 provides an overview of gaze modeling, covering the computational modeling of visual attention and gaze prediction discussing psychological, computational models, saliency models, saccadic models and so on.
- Chapter 3 unfolds the thread of the theoretical approach to predictive gaze modeling via reinforcement learning. It covers some algorithms such as contextual multi armed bandits, stochastic modeling of eye movements and foraging behaviour.
- Chapter 4 deals with modeling attention deployment and gaze behaviour while viewing social interactions. The OUTS-CMAB gaze model is proposed, by drawing upon concepts from foraging theory and the Marginal Value Theorem laid down in the previous chapter.
- Results are presented in Chapter 5 related to the analysis performed on a very large dataset on foraging behavior in inhomogeneous environments; qualitative and quantitative analyses are reported concerning model performance using contextual bandits and specifically Thompson's sampling algorithms.
- Eventually, Chapter 6 draws some conclusions on results so far achieved and discusses how the approach undertaken in this Thesis can improve our understanding of foraging behaviour in inhomogeneous environments. Also, some insights are outlined into optimization strategies and possible future research.

# Capitolo 2

## Gaze Modeling

Attention can be understood as the mental capacity to choose stimuli, responses, memories, or thoughts that are behaviorally relevant among many others that are behaviorally irrelevant. Importantly, this definition emphasizes that in the context of visual attention, the concept of relevance of cues becomes crucial, which depends on the specific tasks or goals of the observer (behavioral relevance) [Corbetta, 1998]. In the realm of visual attention, the process of selecting information is accomplished through a sequence of eye movements, known as overt attention. However, covert attention, which refers to the underlying mechanisms of attentional selection based on task knowledge and observer goals, also plays a significant role in enhancing the perception process [Hoffman, 1998]. Covert attention influences and guides overt attention, thus eye movements, towards specific regions of interest [Hoffman, 1998]. There is evidence to suggest that shifts of attention to a particular location often precede explicit eye movements towards that location [Hoffman, 1998]. Importantly, this coupling between covert and overt attention persists regardless of whether the eye movement is triggered by bottom-up factors, such as sudden movements occurring within the observed scene, or top-down influences, such as endogenous control, instructions, or expectations [Hoffman, 1998].

While overt attention involves the explicit selection of a specific region of interest, the determination of which regions of the stimuli are considered interesting is entrusted to covert attention mechanisms mediated by the observer's goals, whether internal or external [Hoffman, 1998]. The allocation of attentional control mechanisms jointly determines which parts of the stimuli are deemed important, thus guiding where to look and in what order [Corbetta, 1998]. Computational models of visual attention aim to provide an explicit description of this complex machinery, encompassing covert and overt attention, bottom-up versus top-down control, spatial salience, and the temporal unfolding of eye movements.

## 2.1 Psychological and Computational Models of Visual Attention

The modeling of visual attention initially focused on psychological perspectives, with a diverse range of theories and models aimed at understanding human perception [Frintrop et al., 2010].

One influential model in this domain is Treisman's Feature Integration Theory (FIT) [Treisman and Gelade, 1980], which suggests that different features are selected in parallel across the visual field at an early stage, while object identification occurs later and separately. FIT represents these features using separate feature maps that are later fused to create a master map, providing a topographical representation of all the features.

Similarly, [Wolfe, 1994] proposed the Guided Search Model, which shares many concepts and architectural designs with FIT. Several other notable psychophysical models of visual attention do exist, such as the Biased Competition Model by Koch and Ullman's model [Koch and Ullman, 1985], and Tsotsos' Selective Tuning model [Tsotsos et al., 1995].

In addition to psychological approaches, connectionist models have also been propo-

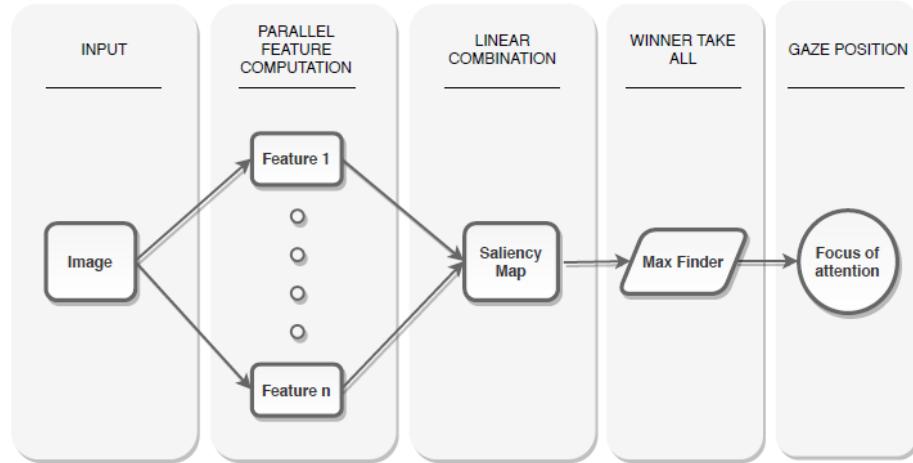


Figura 1: General structure of models of visual attention.

sed, including MORSEL [Mozer, 1987], SLAM [Phaf et al., 1990], SERR, and SAIM, which was subsequently refined into the Visual Search SAIM (VS-SAIM). While these psychological models primarily deal with simple stimuli, such as synthetic images, the field of computational vision aims to develop models capable of handling more complex stimuli, such as natural scenes. As a result, numerous models of visual attention

have been proposed in the past 15-20 years [Frintrop et al., 2010]. The general structure of all such models can be described as shown in Figure 1; although each system may vary in detail, most of them share a similar structure. A notable example is the model by [Itti et al., 1998], which is considered one of the most popular computational models of visual attention. Since its publication, the field has seen significant growth. This model serves as the computational counterpart of the Koch and Ullman [Koch and Ullman, 1985] and Treisman's FIT models [Treisman and Gelade, 1980], relying on the computation of several features in parallel. These features are then collected in conspicuity maps, represented as grayscale images with brightness proportional to the intensity of the feature (e.g., color, intensity, orientation). The computation of these features is achieved through a set of linear center-surround operations that mimic visual receptive fields in the primary visual cortex. Prior to feature computation, the input image is resampled at multiple scales using Gaussian pyramids, enabling the detection of salient locations that stand out from their surroundings. The conspicuity maps are then fused into a single saliency map, often referred to as Treisman's master map of location, through a linear combination. The sequence of



Figura 2: *Example output of the model on a sample image.*

fixation locations on the input image is determined by selecting the local maxima of the saliency map using a winner-take-all (WTA) network. Inhibition of return (IOR) is incorporated to transiently inhibit previously fixated locations, preventing the model from becoming fixated on the same portion of the stimuli. Figure 2 illustrates the output of this procedure on a sample image. This biologically motivated approach provides insights into how such a mechanism might operate in the human brain [Koch and Ullman, 1985]. Importantly, the sequential WTA selection mechanism allows the model to produce a sequence of fixations and saccades that mimic the overt attentional control of the human visual system. The oculomotor control mechanism in humans determines which portion of the scene is worth attending to

at any given time. Thus, a computational model of attentive eye guidance aims to answer the question: "*Where to look next?*"

By abstracting the structure of the model, one can practically address this question by mapping visual data of a natural scene, represented by raw image data  $I$  (either a static picture or a stream of images), to a sequence of time-stamped gaze locations

$$(r_{F1}, t_1), (r_{F2}, t_2), \dots \quad (1)$$

denoted as [Boccignone, 2016].

$$I \rightarrow r_{F1}; t_1; r_{F2}; t_2; \dots \quad (2)$$

From a modeling perspective, when given a stimulus  $I$ , whether it is a static image or a dynamic video, the only available observations are the sequence of locations in the scene visited by the observer. In the case of static stimuli, the continuous sequence of eye movements can be classified into corresponding fixations and saccades. However, when dealing with videos, smooth pursuits should also be considered. For simplicity, we will use the term "gaze shift" to describe a sequence of pursuits, saccades, or fixations. Henceforth, we denote the time series of gaze shifts as  $(r_{F(1)}, t_1), (r_{F(2)}, t_2), \dots$ , compactly represented as  $(r_{Fn}, t_n) = r_F(n)$ . The typical approach to establish the mapping is a two-step process:

- Compute the perceptual representation:

$$I \rightarrow W \quad (3)$$

- use  $W$  to generate the scanpath:

$$W \rightarrow r_{F(1)}, r_{F(2)}, \dots \quad (4)$$

In the recent literature, the focus of computational modeling has been primarily set on the first step, which involves deriving a perceptual representation, usually in the form of a saliency map. This approach has enabled the development of models that can determine *where* to look on a given stimulus, disregarding the temporal aspect of gaze deployment. Consequently, the second step, which addresses *how* to look at the stimulus, has often been neglected. However, in the upcoming sections, we will delve into a more comprehensive exploration of both questions.

## 2.2 Computational Modeling of Visual Attention

As previously mentioned, constructing the perceptual representation of stimuli involves determining where to focus attention, such as on features, objects, or actions, and their respective locations within the scene. The computational modeling of visual attention has predominantly been dealing with this specific aspect [Borji et al., 2013, Itti and Koch, 2001]. Numerous saliency models have been developed with the primary goal of deriving a representation  $W$ , which captures the salient regions of the stimuli.

### 2.2.1 The Evolution of Saliency Models

Visual attention models, particularly saliency models, have gained significant popularity in various fields such as computer vision, image and video processing, compression, and quality assessment [Nguyen et al., 2018]. These models primarily derive a representation of the visual stimuli that emphasizes important cues within the image or video [Bylinskii et al., 2015]. This emphasis on cues allows these models to not only serve as tools for understanding human attentive behavior but also as means to process and analyze images themselves. However, it should be noted that modern deep saliency models often lack explanatory capabilities due to their black box nature [Bylinskii et al., 2015].

The influential model proposed by [Itti et al., 1998] follows a bottom-up approach by computing conspicuity maps based on simple features such as color, intensity, and orientation [Itti et al., 1998]. These maps are then fused into a saliency map, which represents the likelihood of fixating at each location in the image. This approach primarily focuses on low-level visual cues and does not consider top-down information [Itti et al., 1998]. While there has been a long-standing debate regarding the bottom-up versus top-down nature of eye guidance control recent studies suggest that factors such as context, spatial biases, affect, personality, and dynamics of attention deployment play significant roles in the effectiveness and performance of saliency models. Despite these considerations, many computational models have predominantly focused on low-level visual conspicuity without incorporating semantic content into the scene [Foulsham and Underwood, 2008, Einhäuser et al., 2008, Tatler and Vincent, 2011, Borji and Itti, 2013, Itti et al., 1998, Bruce and Tsotsos, 2015]. However, to overcome this limitation, researchers have explored approaches that incorporate top-down information to improve fixation prediction when dealing with objects, faces, text regions, or contextual cues.

It is important to note that historically, the term "saliency" has been used to describe the topographic representation of bottom-up features. However, even with

the addition of high-level processing capabilities, such models are still referred to as saliency models in the modern jargon. The combination of bottom-up features and top-down information, along with the use of machine learning techniques, has enabled the development of saliency models as predictors, employing various learning techniques such as Support Vector Machines and sparse representation [Kienzle et al., 2006, Jiang et al., 2015, Yan et al., 2010]. While this approach allows for data-driven assessment of visual features and optimal predictors, it can also lead to models that are data-dependent, slow, and lack explanatory power.

## 2.3 Overview of Saccadic Models for Gaze Shift Prediction on Static Stimuli

Within the thesis, various models were developed to simulate gaze shifting or saccadic eye movements. These models involve processes that take a perceptual representation of stimuli (referred to as  $W$ ) as input and generate a sequence of fixation locations. In the following sections, we provide a concise overview of some of the saccadic models proposed in the literature, specifically highlighting the distinction between models designed for static or dynamic stimuli. It is worth noting that there are only a limited number of models available for predicting the dynamics of gaze displacement, and the majority of these models are primarily focused on processing static images.

When considering static stimuli, the[Itti et al., 1998] model stands out as a prominent example, being the pioneering work in the field. However, this model lacks certain crucial aspects of gaze shifts, such as variability and the modeling of systematic tendencies. Some recent studies have addressed these aspects. For instance,[Le Meur and Liu, 2015] proposed a model for predicting observers' scan paths on static images by incorporating bottom-up saliency and considering oculomotor biases through sampling from empirical distributions of saccade amplitudes and orientations derived from publicly available datasets.

[Wang and Ji, 2011] presented a computational model for scanpath prediction based on information maximization. This model integrates three factors, including reference sensory responses, fovea-periphery resolution discrepancy, and visual working memory, as driving forces for guiding sequential eye movements. [Sun et al., 2014] introduced a statistical framework that models both saccadic eye movements and visual saliency using super-Gaussian component (SGC) analysis. More recently, deep learning approaches have been employed. [Assens et al., 2018] introduced PathGAN,

a deep neural network trained on adversarial examples for visual scanpath prediction. They also presented SaltiNet [Assens et al., 2018], a deep model for predicting scanpaths and saliency in 360-degree images. [Xia et al., 2019] addressed the problem of saccadic scanpath prediction by introducing an iterative representation learning framework, where eye movements are outcomes based on the current representation, which is updated according to the gaze shift. Fixation selection relies on a perceptual residual computed using an auto-encoder network.[Xia et al., 2019] applied a similar iterative representation model to predict human scanpaths on web pages. [Bao and Chen, 2020] proposed a deep convolutional saccadic model that simultaneously predicts foveal saliency maps and fixation durations using convolutional neural networks (CNNs). [Sun et al., 2019] proposed a recurrent mixture density network-based framework for predicting human-like scanpaths on static images. This model predicts both the sequence of fixations and their durations by leveraging both bottom-up saliency and semantic features extracted by CNNs. Of particular interest for the subsequent sections of this chapter is the Constrained Lévy Exploration (CLE) model proposed by [Boccignone and Ferraro, 2004]. The CLE model considers gaze motion as the stochastic dynamics of a Lévy forager influenced by an external force that depends on a salience or attention potential field. Namely, at time  $t$  the transition from the current position  $r(t)$  to a new position  $r_{new}(t)$ ,  $r(t) \rightarrow r_{new}(t)$ , is given by:

$$r_{new}(t) = r(t) + g(W(r(t))) + \eta \quad (5)$$

The movement of the variable  $r$  follows a trajectory that is influenced by two components. The first component, denoted as  $g$  represents the deterministic part of the trajectory and relies on salience or fixation density. The second component, denoted  $\eta$  represents the stochastic part of the trajectory,  $\eta \sim P(\eta)$ . This component is a random vector sampled from a heavy-tailed distribution  $P(\eta)$  and takes into account motor biases (e.g., small gaze displacements, occasionally followed by long 'jumps'). The dynamics of the Lévy forager can be mathematically expressed as

$$r_{new}(t) = r(t) - \Delta V + \eta. \quad (6)$$

The new gaze position is determined by two factors: a) the gradient of the external force field, denoted  $V$ , which is shaped by the perceptual landscape, and b) the scalar field  $V(\cdot; t)$ , which represents the time-varying nature of the force field.

$$V(x; y; t) = \exp(-\tau_V W(x; y; t)); \quad (7)$$

The stochastic vector  $\eta$  with components

$$\eta_x = l \cos(\theta), \eta_y = l \sin(\theta) \quad (8)$$

In this context, the angle  $\theta$  represents the flight direction, while  $l$  denotes the jump length. The flight direction is sampled from a uniform distribution, and the jump length is sampled from  $\alpha$ -stable distribution:

$$\theta \sim Unif(0, 2\pi), \quad (9)$$

$$l \sim \varphi(W)f(l; \alpha, \beta, \omega, \delta) \quad (10)$$

During the extensive stage, the flight direction  $\theta$  and jump length  $l$  capture the internal decision-making process of the forager. The function  $W(\cdot)$  introduces modifications to the pure Levy flight, as it incorporates the probability of transitioning from one location to another based on the strength of a bond.

$$\varphi(W) = \frac{\exp(-\beta_P(W(r(t)) - W(r_{new(t)}))}{\sum_{r'_{new}} \exp(-\beta_p(s(r(t)) - W(r'_{new}(t)))})} \quad (11)$$

For models addressing gaze shifts in dynamic stimuli, only a few have been proposed. [Boccignone and Ferraro, 2004] introduced Ecological Sampling, a stochastic model of eye guidance in videos. This model assumes that gaze sequences are generated by a stochastic process. The dynamics of gaze shifts are implemented using a stochastic differential equation driven by  $\alpha$ -stable noise. The motivation for this approach stems from the concept of Lévy flights in foraging displacements [Viswanathan et al., 2008]. The perceptual representation in this model is formalized in terms of proto-objects, which are units of visual information accessed through selective attention and validated as actual objects. The eye guidance strategy involves selecting where to look next by sampling appropriate motor behaviors (such as fixating, pursuing, or saccading) based on the perceived world and previous actions. The overall control strategy relies on a complexity measure of the time-varying scene, and the behavioral state is determined by a composite sampling strategy that considers the complexity of the perceived scene at a given time. Complexity is computed using stochastically sampled interest points from the proto-object representation.

*The problem of evaluation.* Similar to saliency models, evaluating gaze shift models poses the challenge of defining a performance metric that can effectively capture the complexities associated with a fixation sequence. While standard metrics exist and are widely adopted for saliency estimation, there is a lack of consensus regarding the most suitable evaluation metrics for gaze shift models [Le Meur and Liu, 2015, Anderson, 2013]. In recent years, several measures have been proposed, each designed to address specific aspects of scan path similarity. Consequently, the choice of an appropriate scanpath metric depends on the specific feature one intends to measure. While a brief summary of some widely used metrics and their qualitative behavior is

provided here, for a comprehensive review and discussion, refer to [Anderson, 2013].

### **Comparative Analysis of ScanMatch & Multimatch**

ScanMatch is a successful method for comparing scanpaths, which is based on the string edit distance commonly used for comparing character sequences. It involves applying a series of edits (insertions, deletions, substitutions) to transform one string into another, and the similarity between the two sequences is determined by the number of editing steps required for the transformation. This approach has been employed in scanpath comparison [?] (Brandt and Stark, 1997), where the image is divided into cells and assigned unique characters, treating fixation sequences as sequences of characters. A refined version of this method, known as ScanMatch. ScanMatch utilizes the Needleman-Wunsch algorithm from bioinformatics to compare DNA sequences. In this approach, the two fixation sequences are first spatially and temporally binned and then recoded to obtain sequences of letters representing spatial position, temporal duration, and order information. The resulting letter sequences are compared by maximizing a similarity score computed from a substitution matrix that assigns scores to pair substitutions of letters and penalties for gaps. The algorithm generates a score in the range [0, 1] for each pair of scanpaths. ScanMatch offers the advantage of considering spatial, temporal, and sequential similarity between scanpaths, providing an overall measure of resemblance. However, it does not provide a detailed description of performance across different dimensions of gaze dynamics, and it may suffer from quantization issues associated with spatial and temporal binning processes.

Another notable metric is Multimatch, which is also used for comparing scanpaths. .

### **MultiMatch**

The MultiMatch metric [Jarodzka et al., 2010, Dewhurst et al., 2012] is a vector-based method designed to measure scanpath similarity across multiple dimensions. The algorithm takes two scanpath sequences as input, which can have varying lengths. The scanpaths are first simplified by combining successive fixations that are within a specified distance or directional threshold of each other. Then, the scanpaths are aligned using a dynamic programming approach. The output of the method is a 5-dimensional vector that quantifies the similarity between the scanpaths in different aspects:

- *Vector* : Measures the overall similarity in shape between the aligned saccade pairs, considering spatial differences in fixation position.
- *Direction*: Represents the angular distance between aligned saccade vectors, providing insights into the similarity of saccade directions.

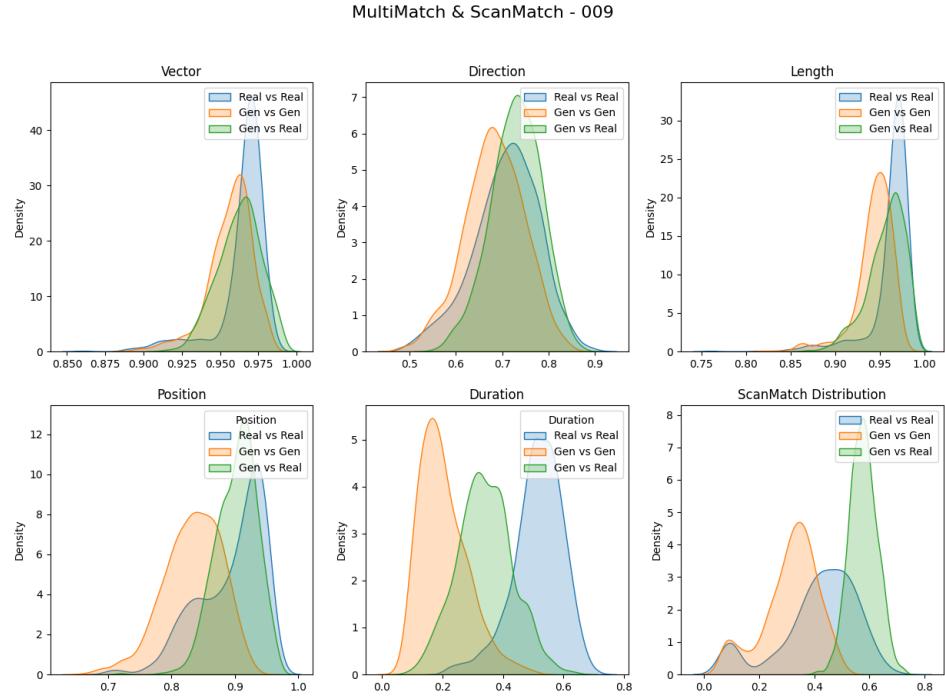


Figura 3: *Comparison of Kernel Density Estimation (KDE) plots: Real vs Real, Generated vs Generated, Real vs Generated.*

- *Length*: Quantifies the absolute difference in length between the endpoints of aligned saccade vectors, focusing specifically on saccade amplitudes.
- *Position*: Calculates the Euclidean distance between aligned fixations, taking into account both saccade amplitudes and directions.
- *Duration*: Computes the absolute difference in fixation duration between aligned fixations, normalizing it by the maximum duration.

Each dimension is normalized to have values in the range [0, 1], with higher values indicating higher similarity. The MultiMatch method offers multiple measures to evaluate scanpath similarity, capturing different aspects of resemblance. However, the scanpath simplification procedure raises questions about the metric's sensitivity to variations [Anderson, 2013]. [Jarodzka et al., 2010] Multimodal information preserves perceptual performance in visual search for expert radiologists. Using ScanMatch to compare fixations across groups: A tutorial and a database. Journal of

Eye Movement Research. The MultiMatch method normalizes each dimension to a range of [0, 1], with higher values indicating higher similarity ( $1 - distance$ ). Its main advantage is that it offers multiple measures to evaluate scanpath similarity, with each measure capturing a distinct aspect of scanpath resemblance. However, the scanpath simplification process raises questions about the metric's sensitivity to variations [Anderson, 2013].

## 2.4 Summary

This section provides an overview of the field of gaze modeling, focusing on the computational modeling of visual attention and the prediction of gaze behavior. It begins by discussing the importance of attention in visual perception and the role of eye movements in selecting relevant information. The coupling between covert and overt attention is highlighted, emphasizing their interplay in guiding visual exploration. The section then explores the evolution of psychological and computational models of visual attention. It covers influential models such as Treisman's Feature Integration Theory (FIT) and the Guided Search Model, along with connectionist models like MORSEL and SLAM. Additionally, it introduces the popular Itti et al. model, which computes saliency maps to determine where to focus attention. The focus then shifts to the computational modeling of visual attention, specifically saliency models. These models aim to derive perceptual representations that emphasize salient regions of stimuli. Various saliency models, including [Itti et al., 1998] , are discussed, highlighting the incorporation of bottom-up features and the recent integration of top-down information.

Next, the section provides an overview of saccadic models for gaze shift prediction on static stimuli. It discusses notable models such as [Le Meur and Liu, 2015] model, which incorporates oculomotor biases, and [Wang and Ji, 2011] model, which integrates reference sensory responses, fovea-periphery resolution discrepancy, and visual working memory. Deep learning approaches, such as PathGAN and SaltiNet, are also mentioned. For gaze shift prediction on dynamic stimuli, the section mentions models like Ecological Sampling, which simulates gaze sequences based on stochastic processes, and G-Eymol, which incorporates gravitational fields and inhibition of return. However, it notes that few models exist for predicting gaze behavior in dynamic stimuli.

The section concludes by discussing evaluation metrics for gaze shift models, highlighting ScanMatch and MultiMatch as popular methods for comparing scanpaths. It emphasizes that the choice of metric depends on the specific aspects of gaze dynamics one intends to measure.

Overall, this section provides a comprehensive overview of gaze modeling, encompassing psychological and computational models of visual attention, saliency models, saccadic models for static and dynamic stimuli, and evaluation metrics. This foundation sets the stage for further exploration and development of gaze modeling techniques in the subsequent sections of the thesis. 5

# **Capitolo 3**

## **Theoretical Background**

In this section, we will delve into the theoretical approach adopted for the project. Building upon the concepts and scope discussed in the previous chapter, our goal is to develop a predictive gaze modeling system utilizing foraging principles. To achieve this, we employ a machine learning approach, with a particular focus on reinforcement learning. Reinforcement learning is a subfield of machine learning that involves training an agent to interact with an environment and make decisions based on rewards and penalties. It is well-suited for modeling gaze behavior as it allows the agent to learn optimal strategies for selecting gaze targets based on the perceived environment. Throughout this chapter, we explore and compare various reinforcement learning algorithms. These algorithms were carefully selected and evaluated to determine the most suitable approach for our gaze modeling objectives. By assessing their performance and effectiveness, we aim to identify the algorithm that provides the best results in terms of predicting gaze behavior accurately. We will discuss the underlying principles and mechanics of each algorithm, highlighting their strengths and limitations. Additionally, we will provide insights into the implementation details and considerations involved in training the models using these algorithms. By thoroughly examining the different approaches, we strive to find the algorithm that offers the highest predictive accuracy and efficiency for our gaze modeling project.

### **3.1 Machine Learning**

Machine learning is a branch of artificial intelligence that focuses on the development of algorithms and models capable of learning and making predictions or decisions without being explicitly programmed. It involves the study of computer algorithms and statistical models that allow systems to automatically learn and improve from experience or data.

Machine learning encompasses a wide range of techniques and algorithms that can be applied to various problem domains, including gaze modeling. These techniques can be broadly categorized into three main types:

- **Supervised Learning:** In supervised learning, a model learns from labeled training data, where each input sample is associated with a corresponding target or output label. The model aims to learn the underlying patterns and relationships between the input features and the output labels. Common supervised learning algorithms include linear regression, decision trees, support vector machines (SVM), and neural networks.
- **Unsupervised Learning:** Unsupervised learning involves learning from unlabeled data, where the model aims to discover hidden patterns or structures within the data. Unlike supervised learning, there are no predefined output labels or targets. Clustering algorithms, such as k-means and hierarchical clustering, and dimensionality reduction techniques like principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE), are examples of unsupervised learning algorithms.
- **Reinforcement Learning:** Reinforcement learning is concerned with training an agent to interact with an environment and learn optimal actions based on rewards and penalties. The agent learns through trial and error, receiving feedback in the form of rewards or punishments based on its actions. Q-learning, deep Q-networks (DQN), and policy gradient methods are popular reinforcement learning algorithms.

Machine learning techniques have proven to be highly effective in gaze modeling tasks, allowing for the prediction and understanding of human eye movements. By leveraging the power of large datasets and advanced algorithms, machine learning models can capture complex relationships between various factors influencing gaze behavior. In gaze modeling, machine learning algorithms can be used to learn from eye-tracking data, environmental stimuli, and other relevant features to predict the gaze target or trajectory. By training models on annotated eye-tracking datasets, they can learn patterns and associations that lead to accurate gaze predictions.

Several machine learning models have been applied in gaze modeling research, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and deep reinforcement learning architectures. These models have demonstrated significant advancements in gaze prediction accuracy and have been instrumental in understanding visual attention and gaze behavior.

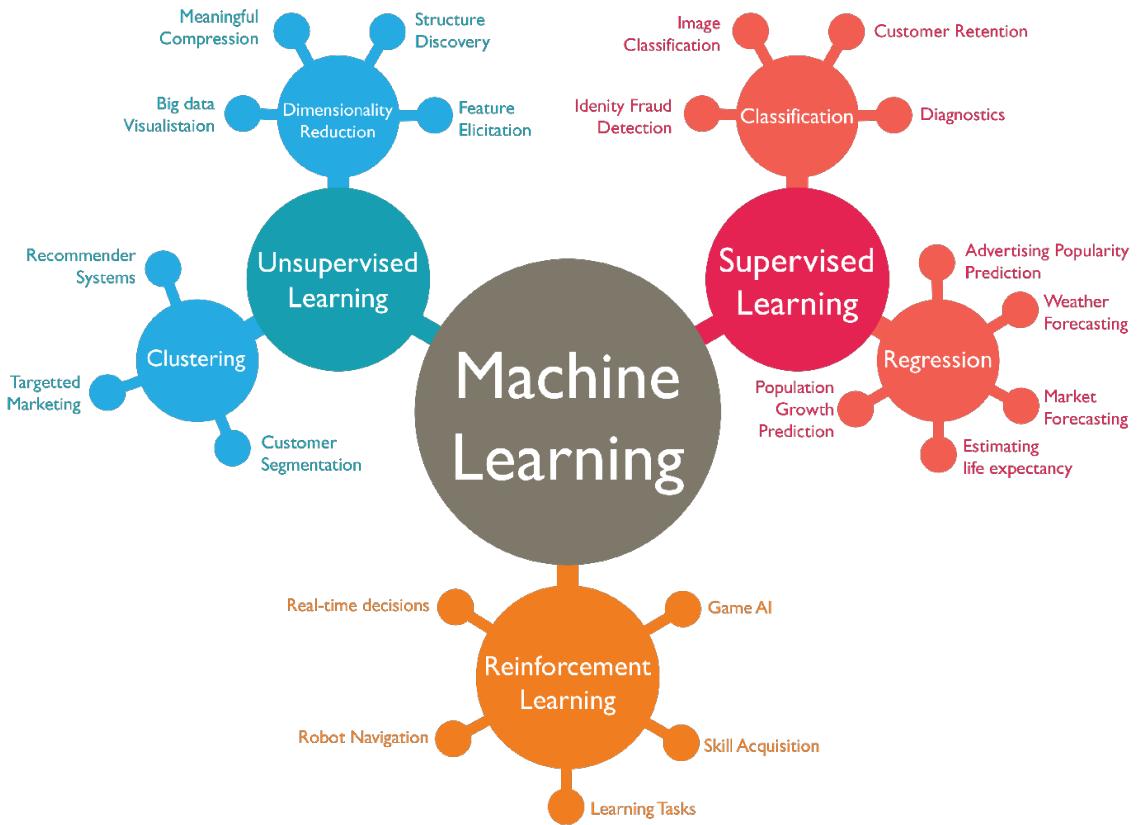


Figura 4: *The three main types of machine learning techniques—supervised learning, unsupervised learning, and reinforcement learning—are represented. Each variant has its distinct characteristics and applications in understanding and predicting gaze behavior*

### 3.2 Reinforcement Learning

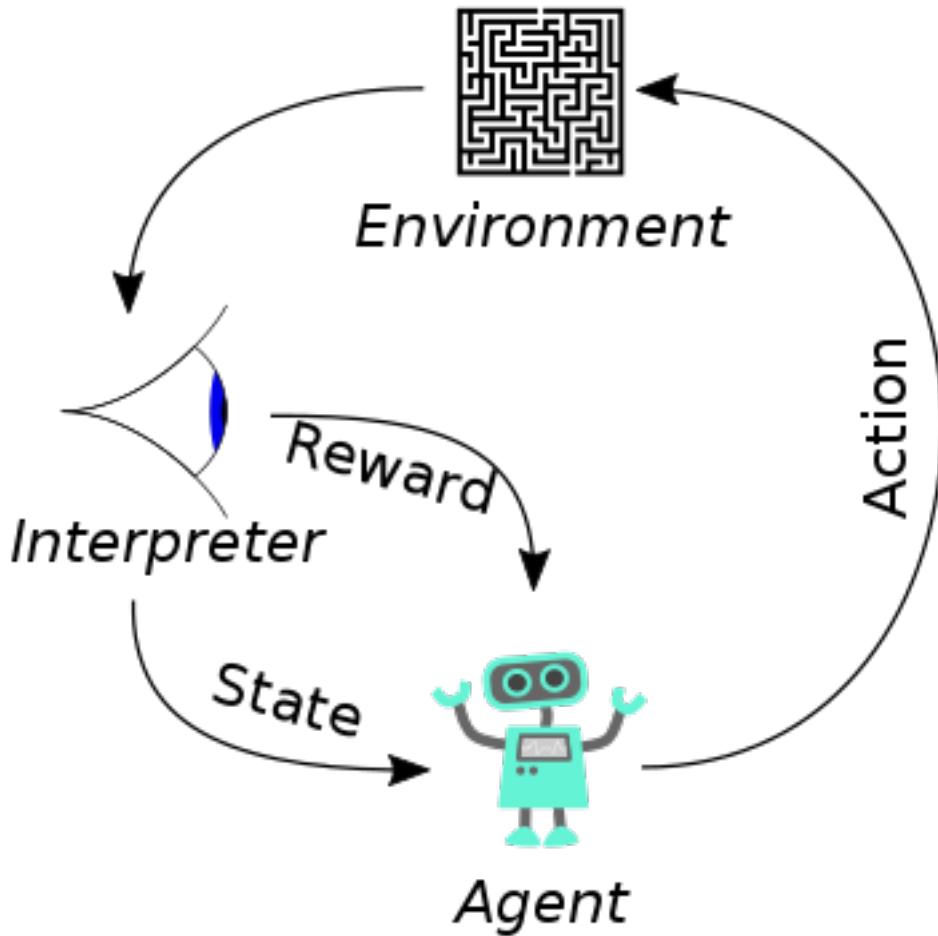
Reinforcement learning (RL) is a field of machine learning that deals with how intelligent agents can make decisions in an environment to maximize their cumulative reward. Unlike supervised learning, reinforcement learning does not require labelled input/output pairs or explicit correction of suboptimal actions. Rather, it focuses on achieving a balance between exploring new areas and exploiting current knowledge.

The environment is often expressed in the form of a Markov decision process (MDP) because many reinforcement learning algorithms for this context use dynamic programming techniques. The main difference between classical dynamic programming

methods and reinforcement learning algorithms is that the latter do not assume knowledge of an exact mathematical model of the MDP, and they target large MDPs where exact methods are impractical.

Basic reinforcement learning is modeled as a Markov decision process (MDP):

- a set of environment and agent states,  $S$ ;
- a set of actions,  $A$ , of the agent;
- $P_a(s, s') = \Pr(s_{t+1} = s' | s_t = s, a_t = a)$  is the probability of transition (at time  $t$ ) from state  $s$  to state  $s'$  under action  $a$
- $R_a(s, s')$  is the immediate reward after transition from  $s$  to  $s'$  with action  $a$ .



A basic reinforcement learning agent AI interacts with its environment in discrete

time steps. At each time  $t$ , the agent receives the current state  $s_t$  and reward  $r_t$ . It then chooses an action  $a_t$  from the set of available actions, which is subsequently sent to the environment. The environment moves to a new state  $s_{t+1}$  and the reward  $r_{t+1}$  associated with the transition  $(s_t, a_t, s_{t+1})$  is determined. The goal of a reinforcement learning agent is to learn a policy:

$$\pi : A \times S \rightarrow [0, 1], \pi(a, s) = \Pr(a_t = a \mid s_t = s)$$

which maximizes the expected cumulative reward.

In reinforcement learning, the problem is formulated as a Markov decision process (MDP) when the agent has full observability of the current environmental state. However, if the agent only has access to a subset of states or the observed states are corrupted by noise, the problem must be formulated as a Partially observable Markov decision process (POMDP). In both cases, the set of actions available to the agent can be limited. For example, the state of an account balance could be restricted to be positive, preventing transitions that would reduce the balance below zero.

To act near optimally, the agent must reason about the long-term consequences of its actions and maximize future rewards, even if the immediate reward is negative. When comparing the agent's performance to that of an optimally acting agent, the difference in performance is referred to as regret. Thus reinforcement learning is a suitable approach for problems that require a balance between long-term and short-term rewards. This methodology has been effectively implemented in various domains, such as robot control, elevator scheduling, telecommunications, and games like backgammon, checkers, and Go (AlphaGo). One of the most common problem that use reinforcement learning is the Multi-armed Bandits.

### 3.2.1 Multi-Armed Bandits

The multi-armed bandit, also called MAB, is a classic problem in the field of sequential decision-making under uncertainty. The name "bandit" is derived from the analogy with a slot machine (one-armed bandit) that has multiple levers or arms. Each arm of the bandit corresponds to a choice or action that can be taken, and the objective is to find the arm that yields the highest reward or payoff.

It's also known as the K- or N-armed bandit problem, is a situation in probability theory and machine learning where a fixed set of resources must be divided between competing choices to maximize expected gain. However, the properties of each choice are only partially known at the time of allocation and may become better understood with time or by allocating resources to that choice. This problem highlights

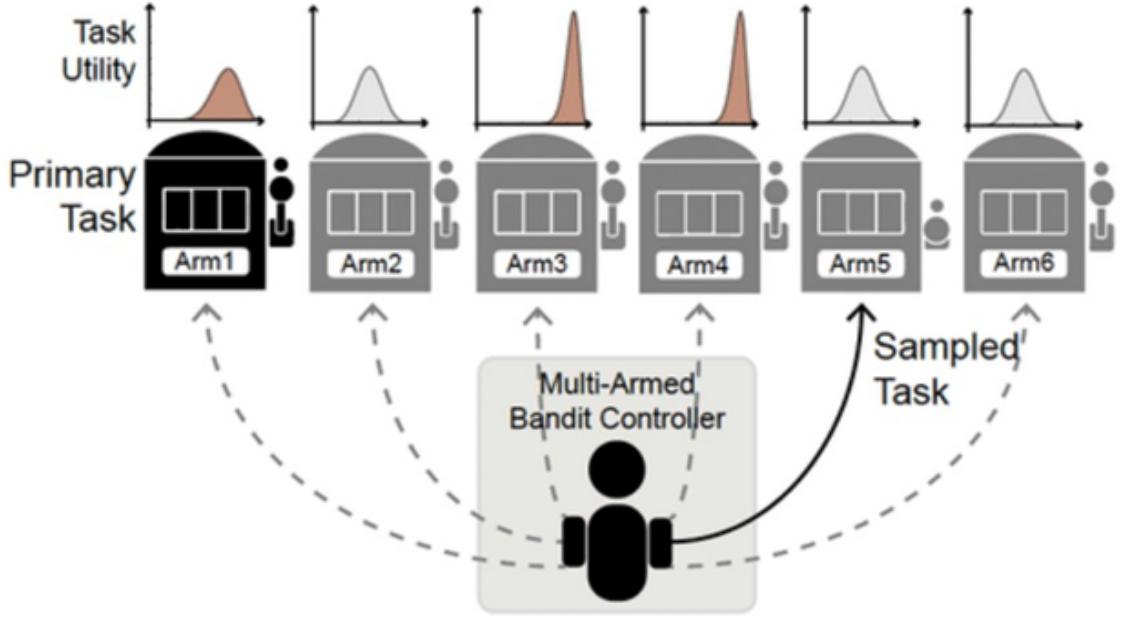


Figura 5: Example of Multi-Armed-Bandits

the exploration-exploitation tradeoff dilemma and falls under the category of stochastic scheduling. The challenge is often exemplified by a gambler playing multiple slot machines and trying to determine which one to play to maximize their winnings. The gambler must decide whether to exploit the machine with the highest expected payoff or explore other machines to gather more information. Multi-armed bandit problems have been used to model real-life scenarios such as managing research projects in large organizations.

### Multi-Armed Bandits Model

The multi-armed bandit, can be seen as a set of real distributions  $B = R_1, \dots, R_k$  each distribution being associated with the rewards delivered by one of the  $K \in \mathbb{N}^+$  levers. Let  $\mu_1, \dots, \mu_K$  be the mean values associated with these reward distributions. The gambler iteratively plays one lever per round and observes the associated reward. The objective is to maximize the sum of the collected rewards. The horizon  $H$  is the number of rounds that remain to be played. The bandit problem is formally equivalent to a one-state Markov decision process. The regret  $\rho$  after  $T$  rounds is defined as the expected difference between the reward sum associated with an optimal strategy and

the sum of the collected rewards:  $\rho = T\mu^* - \sum_{t=1}^T \hat{r}_t$ , where  $\mu^*$  is the maximal reward mean,  $\mu^* = \max_k \{\mu_k\}$  and  $\hat{r}_t$  is the reward in round  $t$ .

### Multi-Armed Bandits Variations

In the field of machine learning, the multi-armed bandit problem is a well-known concept, in which an agent has to decide which arm of a set of slot machines to play, in order to maximize the cumulative reward. There are several variations of Multi-Armed bandit and each them it uses for specific scope.

- **Binary multi-armed bandits** also known as the Bernoulli multi-armed bandit, is a popular form of the multi-armed bandit problem. It provides a reward of either one or zero, with a probability of  $p$  for the former and  $(1 - p)$  for the latter.
- **Markov Machine multi-armed bandits** each arm representing an independent Markov machine, with the reward dependent on the current state of the machine. In the restless bandit problem, non-played arms may also evolve over time, while some systems may increase the number of choices over time. Researchers in computer science have explored the multi-armed bandit problem under worst-case assumptions, developing algorithms to minimize regret in both finite and infinite time horizons, for both stochastic and non-stochastic arm payoffs.
- **Combinatorial multi-armed bandits** Combinatorial bandits involve scenarios where the agent selects a combination of arms or actions rather than a single arm. The agent faces a large combinatorial action space, and the goal is to discover the best combination of arms to maximize the overall reward. Combinatorial bandits have applications in online advertising, where the agent needs to select a set of ads to display that maximize user engagement or click-through rates.
- **Contextual multi-armed bandits** In contextual bandits, each arm's reward depends not only on the action chosen but also on additional contextual information or features. The agent receives a context vector before making each decision, and the goal is to learn a policy that maximizes the expected reward given the context. Contextual bandits find applications in personalized recommendation systems, where the context may include user demographics or browsing history.

This last one variants it was been used in this project.

### 3.2.2 Contextual Multi-Armed Bandits

The contextual multi-armed bandit, also called CMAB, extend the traditional multi-armed bandit framework by incorporating contextual information into the decision-making process. In contextual multi-armed bandits, each arm's reward depends not only on the action chosen but also on additional contextual information or features. This variation allows for more personalized and adaptive decision-making, making it applicable in various domains, such as personalized recommendation systems, online advertising, and healthcare.

Formally:

- **Context:** At each time step  $t$ , the agent receives a context vector denoted as  $x_t$ . This vector contains relevant features or information about the current state or context of the decision problem.
- **Action:** The action space consists of  $K$  arms or actions, where  $K$  is the total number of arms. At each time step  $t$ , the agent selects an arm  $k_t$  based on the observed context  $x_t$ .
- **Reward:** The reward  $r_t$  obtained by pulling arm  $k_t$  at time step  $t$  is a random variable. The reward depends on the context  $x_t$  and the chosen arm  $k_t$ . We denote the reward function  $asr(x_t, k_t)$ , which represents the expected reward associated with arm  $k_t$  given context  $x_t$ .

The objective in contextual multi-armed bandits is to find a policy that maximizes the expected cumulative reward over time. The agent aims to learn a mapping from the context space to the action space, i.e.,  $\pi(x_t) \rightarrow k_t$ , to make informed decisions based on the observed contexts.

Basically, the contextual multi-armed bandits problem is just like the Multi-Armed bandit problem but now the true expected reward parameter  $\theta_k$  depends on external variables. Therefore, we add the notion of context or state to support our decision. Thus, we're going to suppose that the probability of reward now is of the form:

$$\begin{aligned}\theta_k(x) &= \frac{1}{1+exp^{-f(x)}} \\ f(x) &= \beta_0 + \beta_1 \cdot x + \epsilon \\ \epsilon &\sim N(0, \sigma^2).\end{aligned}$$

In other words, the expected reward parameters for each bandit linearly depends of an external variable  $x$  with logistic link.

### 3.2.3 Algorithm for Contextual Bandits

In the context of contextual bandits, adaptive algorithms play a crucial role in balancing exploration and exploitation to maximize cumulative rewards. These algorithms dynamically adjust their decision-making strategy based on past observations and feedback. In this section, we will introduce the concept of adaptive algorithms for contextual bandits and discuss their advantages and applications.

Adaptive algorithms for contextual bandits take into account the contextual information associated with each arm and use it to make informed decisions. These algorithms continuously update their beliefs about the reward distributions of the arms based on the observed context and rewards. By adapting their exploration and exploitation strategies based on this updated information, they aim to efficiently explore promising arms and exploit the arms with higher expected rewards.

In the following subsections, we will delve into specific adaptive algorithms commonly used in the context of contextual bandits. These algorithms include Logistic UCB, Logistic UCB, Epsilon Greedy, Adaptive Greedy and Thompson Sampling. Each algorithm has its own characteristics and strategies for balancing exploration and exploitation. We will explore the details of each algorithm and discuss their performance in different scenarios.

#### Logistic UCB

The **Logistic UCB** is an adaptive algorithm for contextual bandits that combines the principles of UCB and logistic regression [Agrawal and Goyal, 2012]. It leverages contextual information to estimate the expected rewards for each arm, taking into account the specific context in which the bandit problem is being solved.

The algorithm utilizes logistic regression models to capture the relationship between the context and the expected rewards [Langford and Zhang, 2008]. Logistic regression is a widely used statistical method for modeling binary outcomes, which makes it suitable for estimating the probability of success for each arm in a contextual bandit setting. The logistic regression model estimates the model parameters based on observed context-reward pairs, allowing it to learn and adapt to the underlying reward distribution.

To guide the exploration-exploitation trade-off, Logistic UCB employs upper confidence bounds. These bounds are calculated based on the estimated parameters of the logistic regression model and provide a measure of uncertainty in the expected rewards for each arm [Jamieson and Nowak, 2014]. The upper confidence bounds determine the arm selection probabilities, with arms having higher bounds being more likely to be selected for exploration.

The Logistic UCB algorithm updates its estimates of the model parameters as new data becomes available, allowing it to adapt to changing contexts and potentially

non-stationary reward distributions [Dudik et al., 2011]. By continuously updating its estimates, Logistic UCB can make more informed decisions about arm selection, striking a balance between exploring new arms to gather information and exploiting the arms with higher expected rewards.

Research studies have shown that the Logistic UCB algorithm performs well in various contextual bandit scenarios, particularly when the underlying reward distributions are non-linear and exhibit complex dependencies on the context variables [Chapelle and Li, 2011]. It has demonstrated strong performance in real-world applications, such as personalized recommendation systems, online advertising, and healthcare interventions.[Engbert et al., 2011].

### Epsilon Greedy

The **Epsilon Greedy** algorithm is a classic and straightforward adaptive algorithm for contextual bandits. It aims to balance the exploration of different arms and the exploitation of the arm with the highest estimated reward. The algorithm achieves this balance by allocating a portion of the decision-making process to random exploration and the remaining portion to selecting the arm with the highest estimated reward. In the Epsilon Greedy algorithm, a parameter called epsilon ( $\epsilon$ ) determines the exploration rate. With a probability of  $\epsilon$ , the algorithm selects an arm randomly, exploring the available options. This random exploration helps in discovering potentially better-performing arms. Conversely, with a probability of  $(1-\epsilon)$ , the algorithm selects the arm with the highest estimated reward, exploiting the arms that have shown promising performance based on past observations. Epsilon Greedy is a simple yet effective adaptive algorithm that is easy to implement and understand. However, it may struggle in scenarios where fine-grained exploration is required, as it assigns equal exploration probability to all arms [Makarava et al., 2012], [Audibert and Bubeck, 2009], [Gittins, 1979].

The Epsilon Greedy algorithm boils down to the following:

$$A_t = \begin{cases} \text{random arm, with probability } \epsilon \\ \arg \max_i Q_t(i), \text{ with probability } 1 - \epsilon \end{cases} \quad (12)$$

where  $A_t$  represents the arm selected at time step  $t$ ,  $Q^i$  denotes the estimated reward for arm  $i$  at time step  $t$ , and  $\arg \max_i Q^i$  is the arm with the highest estimated reward. Thus, Eq. 12 describes the decision-making process of the Epsilon Greedy algorithm, where with probability  $\epsilon$  a random arm is chosen for exploration, and with probability  $1 - \epsilon$  the arm with the highest estimated reward is selected for exploitation.

The cited articles provide further insights into the exploration-exploitation trade-off in contextual bandit algorithms and offer additional analysis and evaluations of the Epsilon Greedy algorithm in different contexts.

### Adaptive Greedy

The **Adaptive Greedy** algorithm is an adaptive approach for contextual bandits that dynamically adjusts its exploration and exploitation strategies based on observed rewards. It aims to find a balance between exploring new options and exploiting the arms with higher expected rewards. The algorithm gradually decreases the exploration rate over time, allowing for a transition from thorough exploration to focused exploitation.

The exploration rate in the Adaptive Greedy algorithm can be expressed as follows:

$$\epsilon_t = 1/t^\alpha \quad (13)$$

where  $\epsilon_t$  is the exploration rate at time step  $t$ , and  $\alpha$  is a parameter that controls the rate of exploration decay. As time progresses, the exploration rate decreases, leading to a higher emphasis on exploitation.

The Adaptive Greedy algorithm is particularly effective when the context-reward relationship is complex and requires an extensive exploration phase to discover the optimal arms. It provides a flexible framework that can be adapted to different problem settings.

One study that discusses the Adaptive Greedy algorithm in the context of contextual bandits is the work by [Brockmann and Geisel, 2000a]. They provide empirical support for the algorithm and its effectiveness in scenarios where the context-reward relationship is intricate.

### Thompson Sampling

**Thompson Sampling (TS)** is a heuristic for multi-armed bandit problems that dates back to [Thompson, 1933], but has been rediscovered independently multiple times in the context of reinforcement learning, several studies (e.g., [Granmo, 2010]; [Scott, 2010] [Graepel et al., 2010] [Chapelle and Li, 2011] [Ortega and Braun, 2010]). TS belongs to the family of randomized probability matching algorithms, and assumes a simple prior distribution  $P(\tilde{\mu})$  on the parameters  $\tilde{\mu}$  of the reward distribution of each arm. At every time step, TS plays an arm according to its posterior probability of being the best arm, which is computed based on past observations  $D$  consisting of context  $b$  and reward  $r$ . The likelihood function  $P(r|b, \tilde{\mu})$  gives the probability of reward given a context  $b$  and a parameter  $\tilde{\mu}$ . The posterior distribution  $P(\tilde{\mu}|D)$  is proportional to the likelihood function  $P(D|\tilde{\mu})P(\tilde{\mu})$ , where  $P(D|\tilde{\mu})$  is the likelihood function.

In each round, the Thompson Sampling (TS) algorithm selects an arm based on its posterior probability of having the best parameter. A common approach is to produce a sample parameter for each arm using the posterior distributions, and then play the arm that produces the best sample. In this paper, we propose and analyze a natural extension of TS for contextual bandits that follows the general structure mentioned above and uses a Gaussian prior and Gaussian likelihood function. Although TS is a Bayesian approach, our algorithm description and analysis apply to the prior-free stochastic MAB model, and our regret bounds hold regardless of whether the actual reward distribution matches the Gaussian likelihood function used to derive this heuristic. Thus, our bounds for the TS algorithm are directly comparable to the UCB family of algorithms, which take a frequentist approach to the same problem. One could interpret the priors used by TS as a way of capturing the current knowledge about the arms.

### 3.2.4 Comparison of Contextual Multi-Armed Bandit Algorithms

By using the contextual multi-armed bandit library developed by Cortes [Cortes, 2018], a comparison was conducted to evaluate the performance of the previously mentioned algorithms in terms of cumulative reward. The results are presented in Figure 6, showcasing the superiority of the Thompson sampling algorithm in this specific contextual bandit scenario. Thompson sampling achieved an average accuracy of 67% across all videos, demonstrating the best performance among the algorithms tested. The adaptive nature of Thompson sampling, which utilizes Bayesian inference to update the distribution of rewards based on observed contextual information, allowed it to effectively explore and exploit actions in the given context.

## 3.3 Stochastic models of eye movement

Upon a preliminary review of the literature, it becomes evident that when studying eye movements, the stochastic processes best suited for modeling exhibit violations of the classical Central Limit Theorem (CLT). This observation suggests that the traditional assumptions underlying the CLT may not adequately capture the complex dynamics and inherent characteristics of eye movements. Consequently, researchers have sought alternative stochastic models that better capture the non-Gaussian and non-stationary nature of eye movement data. These models often incorporate features such as heavy-tailed distributions, long-range dependencies, and nonlinear dynamics

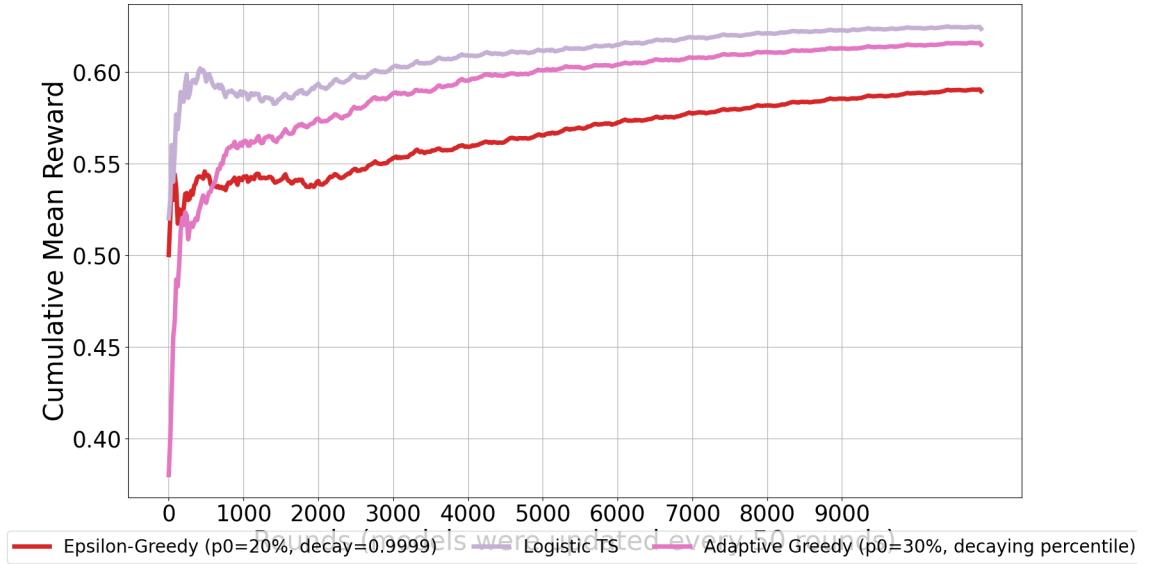


Figura 6: Comparison of cumulative rewards for different contextual multi-armed bandit algorithms. Thompson sampling outperformed other algorithms with an average accuracy of 67% across all videos.

to account for the intricate patterns and irregularities observed in eye movement behavior. By deviating from the assumptions of the CLT, these alternative stochastic processes offer more accurate representations and deeper insights into the underlying mechanisms governing eye movements. Among others, two interesting examples are the modelling of fixational eye movements by [Engbert, 2006] [Engbert et al., 2011] [Makarava et al., 2012], [Engbert and Kliegl, 2004] and the study of saccades amplitude by [Brockmann and Geisel, 2000a]. The foundations of stochastic modeling can be traced back to Albert Einstein's groundbreaking work on Brownian Motion in 1905 [Einstein, 1905]. Brownian Motion refers to the irregular motion observed in small pollen grains suspended in water, first investigated by botanist Robert Brown in 1827. Einstein's mathematical description of Brownian Motion laid the groundwork for several fundamental concepts in the theory of stochastic processes. This includes the Markov assumption, the Chapman-Kolmogorov Equation (which is the central dynamical equation for all Markov processes), and the diffusion equation that describes the behavior of particle ensembles. Subsequently, Paul Langevin expanded on Einstein's work and developed the Langevin equation, which serves as a prominent example of a Stochastic Differential Equation (SDE). Ornstein and Uhlenbeck further enhanced the understanding of stochastic processes with their work on the Ornstein-Uhlenbeck process, which describes the velocity of a Brownian particle

[Einstein, 1905, Gardiner, 2010]. Stochastic processes have proven to be invaluable not only in describing physical phenomena but also in various other fields. They find applications in analyzing electrical circuits, modeling radio wave propagation, and predicting stock market behavior in finance. Moreover, stochastic models have been successfully employed in studying biological systems, such as gene expression and fluctuations in protein concentration within bacteria. Additionally, they have been used to investigate animal movements, including eye movements, revealing intriguing connections between eye movements and ecological factors [Gardiner, 2010]. One fascinating aspect of stochastic processes is the occurrence of super-diffusive processes, which exhibit behavior that deviates from normal diffusion. The concept of anomalous diffusion, characterized by deviations from Gaussian distributions, was first brought to attention by Richardson in 1926. Since then, it has been observed in diverse fields such as finance, biology, and ecology. Lèvy flights, mathematically described as random walks with heavy-tailed jump lengths, have been employed to capture this anomalous diffusion. Notably, Brockmann and Geisel utilized Lèvy flights to describe the oculomotor behavior of humans viewing images, establishing an intriguing connection between eye movements and ecological patterns [Gardiner, 2010]. In the context of this thesis, the focus is on eye movements recorded from human subjects observing stimuli. These eye movements can be viewed as manifestations of fluctuating phenomena resulting from the allocation of visual attention. Stochastic models offer a suitable framework for describing these eye movements, where each sequence of gaze positions is associated with a different realization of a stochastic process [Gardiner, 2010]. The subsequent sections of this chapter aim to delve deeper into the concepts introduced. We will begin with a rigorous definition and description of stochastic processes, exploring their properties and mathematical formulation. We will then explore the adoption of stochastic processes as models for eye movements and their utility in understanding visual attention allocation. Finally, we will introduce the foraging perspective, which provides valuable insights into the underlying mechanisms and ecological significance of eye movements.

### 3.3.1 Stochastic Processes

A stochastic process can be described as a collection of random variables, denoted as  $\mathbf{X}(t)$ , where the index  $t$  typically represents time. If the variable  $t$  takes values from the set of real numbers, the stochastic process is referred to as a continuous time process. Conversely, if  $t$  belongs to the set of natural numbers, it is known as a discrete time process. In the context of eye movements, the trajectory can be viewed as a realization of a stochastic process  $\mathbf{X}(t)$ . When considering a realization of a stochastic process  $x(t)$ , certain measurements such as the mean and variance are readily available at any given time  $t$ . However, computing standard summary

statistics assumes a specific distribution shape for the noise, such as Gaussian, Lévy, Poisson, and so on. These quantities do not provide insights into the dynamics of the process, specifically regarding the influence of current measurements on future ones. The extent of dependence or memory in the measured signal can be characterized by the autocorrelation function.

$$c_{xx}(\Delta) = \frac{1}{N} \sum_{n=0}^{N-|\Delta|-1} x(n)x(n + \Delta) \quad (14)$$

The autocorrelation function quantifies the relationship between the values measured at different points in time within the stochastic process. It provides information about how the values at the current time  $t$  relate to the values at future times. By analyzing the autocorrelation function, one can gain insights into the temporal dependencies and patterns present in the process.

## 3.4 Lévy Flights

The **Lévy flights** are a type of random walk characterized by long, infrequent jumps that follow a power-law distribution. They were first introduced by Lévy in 1939 as a mathematical concept to describe the behavior of certain physical phenomena [Lévy, 1939]. Since then, Lévy flights have been studied and applied in various fields, including physics, ecology, finance, and computer science [Shlesinger et al., 1986]. In this thesis, Lévy flights are used to define eye movements.

### Definition and Properties

A Lévy flight is a stochastic process in which the step lengths between consecutive positions are drawn from a heavy-tailed probability distribution. It's characterized by their heavy-tailed step length distribution, which is often described by a Lévy distribution. The Lévy distribution has the property of infinite variance, meaning that it allows for the possibility of extremely long jumps. This heavy-tailed behavior distinguishes Lévy flights from traditional random walks, where step lengths typically follow a Gaussian distribution with finite variance [Shlesinger et al., 1986].

The probability density function (PDF) of the Lévy distribution can be expressed as:

$$p(x) = 1/\pi \int_0^{\infty} \exp(-k|x|^u) \cos(sx) ds \quad (15)$$

where  $x$  is the step length,  $k$  is a scale parameter, and  $\nu$  is the stability index. The stability index determines the tail behavior of the distribution. For  $0 < u < 2$  the distribution exhibits heavy tails, allowing for the occurrence of long jumps. Lévy flights exhibit several important properties that make them interesting for modeling various phenomena:

- **Scale invariance:** Lévy flights are scale-invariant, meaning that their statistical properties remain the same regardless of the scale or time window considered. This property makes them suitable for modeling systems with fractal or self-similar characteristics [Shlesinger et al., 1986].
- **Long-range dependence:** Lévy flights incorporate occasional long jumps, which introduce long-range dependence in the process. This long-range dependence is often characterized by power-law correlations, where the probability of a jump of length  $x$  is proportional to  $x^{-(1+u)}$ . The presence of long-range dependence can have significant implications for the behavior of the system [Shlesinger et al., 1986, Viswanathan et al., 2008].
- **Anomalous diffusion:** Lévy flights exhibit anomalous diffusion, where the mean squared displacement grows faster or slower than linearly with time. The diffusion behavior of a Lévy flight is determined by the stability index  $u$ . For  $0 < u < 2$  the mean squared displacement diverges, indicating superdiffusive behavior. For  $u = 2$ , the mean squared displacement grows linearly with time, corresponding to normal diffusion. For  $u > 2$ , the mean squared displacement converges, indicating subdiffusive behavior.

Lévy flights have been applied in various fields to model and understand complex phenomena. In physics, they have been used to describe the motion of particles in turbulent fluids, the behavior of cosmic rays, and the dynamics of complex systems. In ecology, Lévy flights have been employed to model animal foraging patterns and search strategies. In finance, they have been used to model asset price fluctuations and to study market efficiency. In computer science, Lévy flights have found applications in optimization algorithms, data clustering, and network traffic modeling. It is worth mentioning that Lévy flights are closely related to another important stochastic process known as Brownian motion or random walks. Brownian motion is a special case of a Lévy flight where the step lengths follow a Gaussian distribution with finite variance. Brownian motion exhibits different statistical properties compared to Lévy flights, with diffusive behavior characterized by a linear growth of the mean squared displacement with time. However, both Lévy flights and Brownian motion play important roles in understanding stochastic processes and have applications in various fields like eye movement. [Shlesinger et al., 1993]

### 3.4.1 Fixational eye movement with Brownian motion

It was been make an analysis on the random walk behavior of fixational eye movements (FEMs) using the mathematical framework of fractional Brownian motion. FEMs refer to small involuntary eye movements characterized by an erratic trajectory or random walk. Within FEMs, two significant types are distinguished: physiological drift or tremor, which involves slow movements at a low velocity, and microsaccades, which are rapid, small-amplitude movements occurring at a rate of approximately one to two per second. [Engbert and Kliegl, 2004] investigated the temporal correlation and persistence of fixational eye movements (FEMs) by calculating the empirical Mean Squared Displacement and estimating the Hurst exponent  $H$ . Similarly, [Makarava et al., 2012] conducted a more comprehensive analysis using Bayesian methods to estimate  $H$ . In this study and observed that, within a short time scale of 2 to 20 ms, fixational eye movements (FEMs) exhibit persistent behavior ( $H > 0.5$ ), indicating a positive correlation between successive increments. However, on longer time scales ranging from 100 ms to 400 ms, FEMs exhibit non-persistent behavior ( $H < 0.5$ ), indicating a negative correlation between successive increments. This shift in behavior is primarily attributed to microsaccades, which act as error-correcting movements to counterbalance the diffusive drift.

The findings align with psychological plausibility, as persistent behavior contributes to retinal image shifts that help prevent perceptual fading. However, a superdiffusive behavior ( $H > 0.5$ ) over a prolonged period could lead to a loss of focus of attention (FoA). Crucially, the negatively correlated increments observed on the longer time scale serve the purpose of preventing misalignment of the FoA [Engbert and Kliegl, 2004]. In a more recent work, [Engbert et al., 2011] proposed a model of FEMs that incorporates self-avoidance. This model considers FEMs as realizations of Self Avoiding Random Walks, where previous locations visited by the eye are avoided.

### 3.4.2 Saccades as Lèvy Flights

[Brockmann and Geisel, 2000a] provided empirical evidence supporting the relationship between eye movements and Lévy flights. They observed a power-law dependence in the tail of the saccade amplitude distribution, which aligned with the characteristics of Lévy flights. To investigate this further, they conducted simulations of artificial scanpaths using random walks with step lengths sampled from either a Gaussian or Cauchy distribution.

In order to simulate eye movements similar to those of humans during free viewing

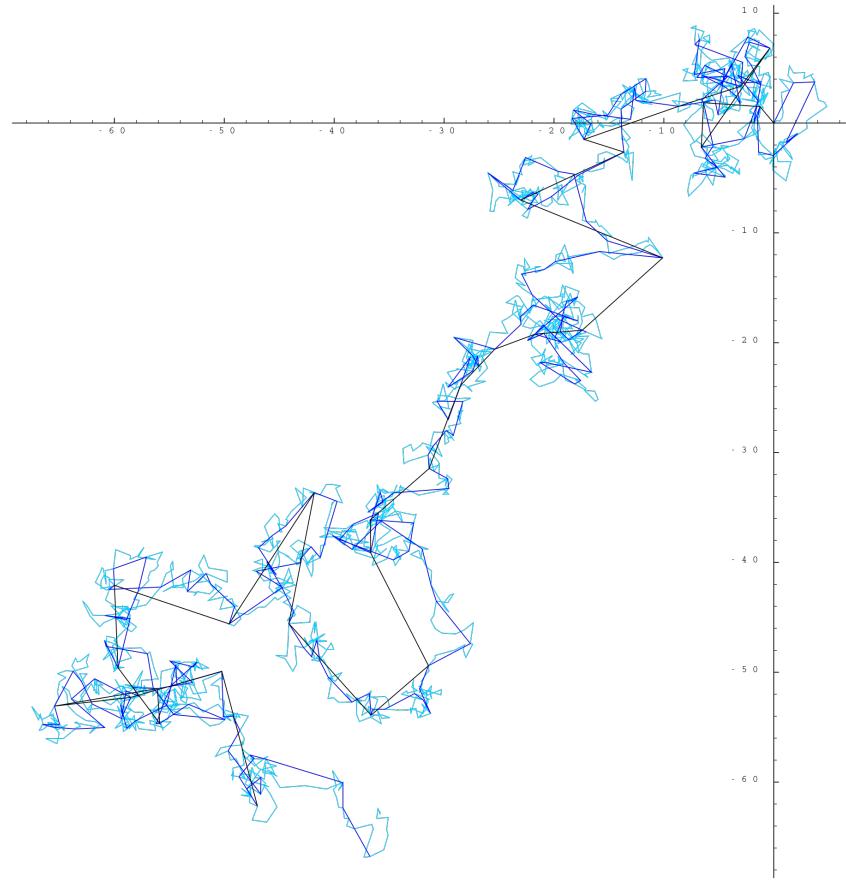


Figura 7: *Example of Brownian walk*

of images, they computed an empirical salience field based on the spatial distribution of fixations made by observers across the scene. This salience field was then used to guide the random walk, either following a Gaussian or Cauchy distribution. To validate their assumption that scanpath step lengths follow a power-law distribution, they collected saccadic magnitudes measured from real subjects while scanning natural scenes. Their analysis showed that the tail of the distribution, when plotted on a log-log scale, followed a straight line, providing support for the concept of saccades as Lévy flights. Furthermore, they demonstrated that using a Cauchy Flight as a model for saccades would result in a much shorter scanning time compared to a random process with Gaussian increments. From a mathematical standpoint, this is not surprising as Lévy flights exhibit super-diffusive behavior with the mean square displacement (MSD) growing super-linearly. However, from a broader perspective,

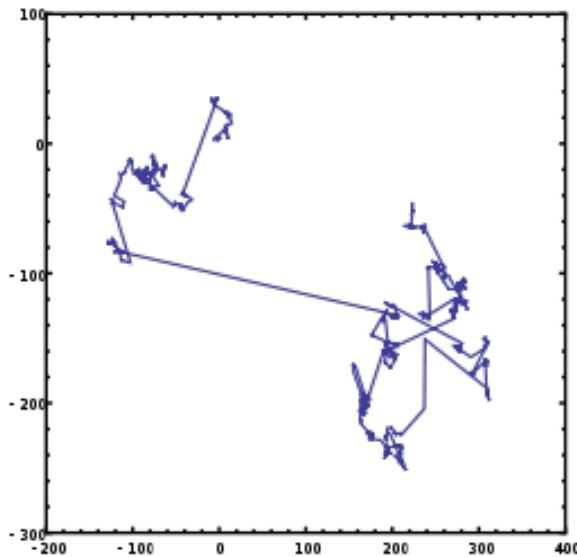


Figura 8: *Example of Lévy Flights*

this raises new questions regarding eye guidance and the implications of Lévy flights in the context of eye movements. Importantly, this perspective enables a connection between eye guidance modeling and theories of foraging animals found in ecological literature, specifically through the concept of the **Lévy Flight foraging hypothesis**. In other words, one can conceptualize the eye (or the brain modules controlling eye behavior) as a forager searching for valuable information (prey) within a given scene (foraging landscape), which may vary over time. This analogy highlights the similarities in the search strategies employed by animals and the mechanisms underlying eye movements and visual exploration.

### 3.5 Foraging

Foraging theory is a branch of behavioral ecology that focuses on studying the foraging behavior of animals, specifically how they search for and acquire food resources in the wild [Stephens and Krebs, 1986]. The movement ecology field within foraging theory aims to understand the movement patterns and statistical properties of animal trajectories during their search for resources. It encompasses various research areas, including stochastic processes and anomalous diffusion [Nathan et al., 2008]. One fundamental concept in behavioral ecology is the Optimal Foraging Theory (OFT), which suggests that the mechanisms driving foraging behavior in organisms have been naturally selected over time to maximize energy intake [Emlen, 1966,

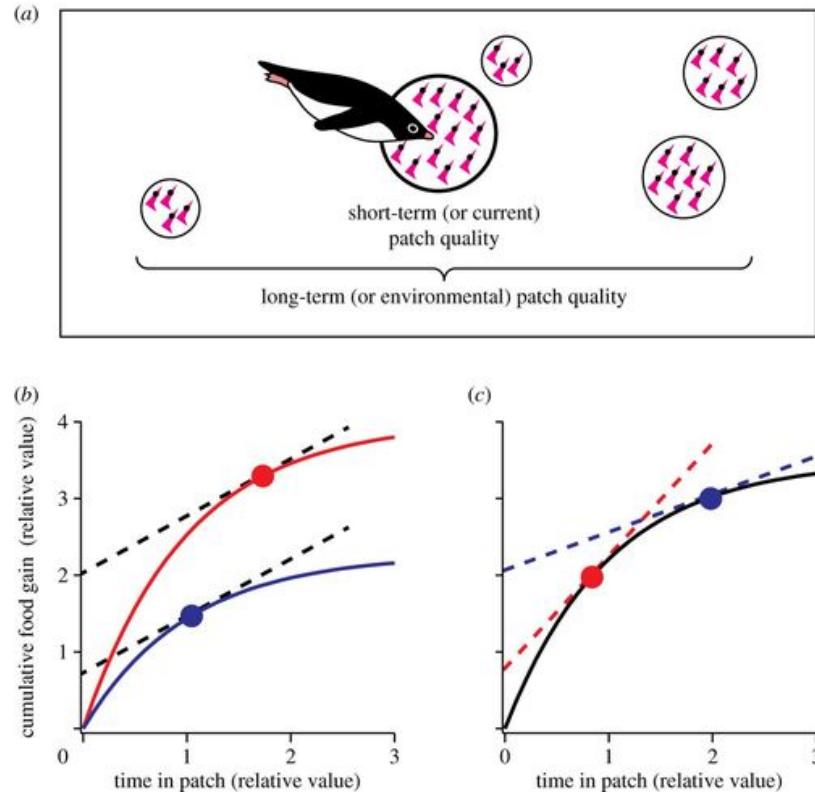


Figura 9: *Illustration of Foraging Theory and visual representation showcasing the duration of an animal's stay within a patch*

MacArthur and Pianka, 1966]. This theory has led to a significant amount of theoretical work aimed at identifying the relevant parameters and factors involved in optimization. Two notable examples in foraging theory are the Lévy Flight Foraging (LFF) Hypothesis and the Marginal Value Theorem (MVT) [Viswanathan et al., 1999] [Charnov, 1976]. Interestingly, many concepts from foraging theory and movement ecology also apply to human movement behavior. Humans engage in various forms of foraging by interacting with their surrounding environment. For example, studies by Gonzalez et al. [Gonzalez et al., 2008] and Brockmann et al. [Brockmann et al., 2006] analyzed human movements using GPS tracks and the circulation of dollar bills, respectively. They found that human movement patterns exhibited characteristics of Lévy flights, indicating super-diffusive behavior. Furthermore, the analysis of eye movements has also been approached from a foraging perspective. In addition to the use of Lévy flights to describe saccadic eye movements [Brockmann and Geisel, 2000b], researchers like Wolfe [Wolfe, 2013] and Cain et al. [Cain et al., 2012] have examined

human visual behavior in the context of foraging theory. They conducted experiments to test the predictive capabilities of the Marginal Value Theorem for visual search tasks. In the subsequent paragraphs of this chapter, we will provide a brief description of the Lévy Flight Foraging Hypothesis, the Marginal Value Theorem, and their applicability to the modeling of eye movements.

### 3.5.1 The Lévy Flight foraging hypothesis

The **Lévy Flight foraging hypothesis** proposes that organisms optimize their random searches by employing Lévy flights, as these flights maximize the coverage of space within a fixed time period. The hypothesis suggests that natural selection should favor adaptations for Lévy flight foraging due to the super-diffusive nature arising from the power-law decay of the tails of the step length distribution [Viswanathan et al., 1999].

The LFF hypothesis holds under specific but common circumstances. It assumes that the forager is engaged in a non-destructive random search, where targets are randomly distributed and regenerate after a short time, and that these targets are sparsely distributed in patches throughout the environment.

In their seminal work, Viswanathan et al. [Viswanathan et al., 1999] analyzed the efficiency of a Lévy forager for different values of the characteristic exponent  $\alpha$  ( $0 \leq \alpha \leq 2$ ) associated with the  $\alpha$ -stable distribution. They demonstrated that in extreme cases ( $\alpha = 0$  and  $\alpha = 2$ ), the forager acted sub-optimally for different reasons. In the case of  $\alpha = 0$ , the forager exhibited ballistic movements, moving in a straight line until encountering prey. This behavior resulted in the discarding of closer prey. Conversely, for  $\alpha = 2$ , the forager performed Brownian motion with Gaussian innovations, biasing the search toward closer targets. However, the linear growth of the mean squared displacement (MSD) in Brownian motion made the Gaussian propagator inefficient, and it led to oversampling of already visited sites. The optimal compromise lies in the range of  $0 < \alpha < 2$  for Lévy flights. It has been shown that the optimal value of the Lévy flight exponent that maximizes search efficiency is  $\alpha = 1$  (Cauchy flight) [Viswanathan et al., 1999]. Empirical findings align with this result, and it is theoretically justified by the fact that for  $\alpha = 1$ , Lévy flights reach the largest possible Hurst exponent ( $H = 1$ ), maximizing the super-diffusivity of the walk [Viswanathan et al., 2011]. However, it is important to note that such optimal behavior holds true only under the specified conditions. In cases of destructive random search, the ballistic Lévy searcher has been shown to act optimally [Viswanathan et al., 2011]. Similarly, in high-density-prey environments where targets are not patchily distributed, Lévy flight searches are indistinguishable from Brownian motion in terms of foraging efficiency [Viswanathan et al., 2011].

### Criticism to the LFF Hypotheses

Despite the considerable success of the Lévy flight foraging (LFF) hypothesis, there is still controversy surrounding its general applicability [Benhamou, 2007, Edwards et al., 2007]. In fact, many of the observed movement patterns attributed to Lévy processes can be generated by a simpler composite random walk process where the turning behavior is spatially dependent [Benhamou, 2007, Codling et al., 2008, Bénichou et al., 2006] [Benhamou, 2007] particularly argued that although the observed movement patterns may resemble those of Lévy flights (evidenced by a step length frequency distribution that fits well on a log-log plot), the actual underlying generating process is not necessarily a Lévy flight. The author demonstrates how Composite Brownian Walks (CBW) can generate search patterns that mimic those produced by Lévy flights and, under specific circumstances, can be more efficient than Lévy flights [Benhamou, 2007]. CBWs are obtained by combining two "classical" random walks, which are stochastic processes characterized by a step length distribution with a finite variance. These walks act together to imitate the intensive and extensive search behaviors of foraging animals. One of these processes is responsible for producing frequent steps with an exponentially distributed length and relatively small mean, while the other generates sporadic steps with an exponentially distributed length and relatively large mean. Benhamou [Benhamou, 2007] demonstrates how simulations from CBWs can be mistakenly identified as Lévy flights. This highlights the crucial distinction between the observed patterns and the inference of the model that generated the data, emphasizing the importance of the methods used for such deductions. Building on the ideas presented by Benhamou [Benhamou, 2007], one could argue for the plausibility of considering eye movements, specifically the alternation between fixations and saccades, as being generated from a mixture of simpler random walks rather than Lévy flights. A more rigorous explanation of this concept will be provided in the next chapter. For now, it is sufficient to consider the following pictorial representation [Benhamou, 2007]. One potential solution involves the utilization of an Ornstein-Uhlenbeck (OU) model with switching parameters as a mechanistic framework to explain eye movements. This model enables the switching between feeding (fixation) and relocation (saccades) based on a switching signal that represents the outcome of a decision-making process.

## 3.6 The Ornstein-Uhlenbeck model

The Ornstein-Uhlenbeck process, initially developed by Leonard Ornstein and George Eugene Uhlenbeck, is a stochastic process widely used in various fields such as financial mathematics and the physical sciences.

Originally, it was employed as a model to describe the velocity of a Brownian particle subjected to friction. This process is characterized as a stationary Gauss-Markov process, possessing the properties of being a Gaussian process, a Markov process, and temporally homogeneous. In fact, it is the only nontrivial process that satisfies these three conditions, with the allowance of linear transformations of the space and time variables [Gardiner, 2010]. Over time, the Ornstein-Uhlenbeck process tends to revert back to its mean function, leading to its classification as a mean-reverting process.

From a conceptual perspective, the process can be viewed as a modification of the continuous-time random walk, or Wiener process, where adjustments have been made to induce a tendency for the walk to return to a central location. The strength of this attraction increases as the process moves further away from the center. In the realm of discrete-time processes, the Ornstein-Uhlenbeck process corresponds to the continuous-time analogue of the AR(1) process.

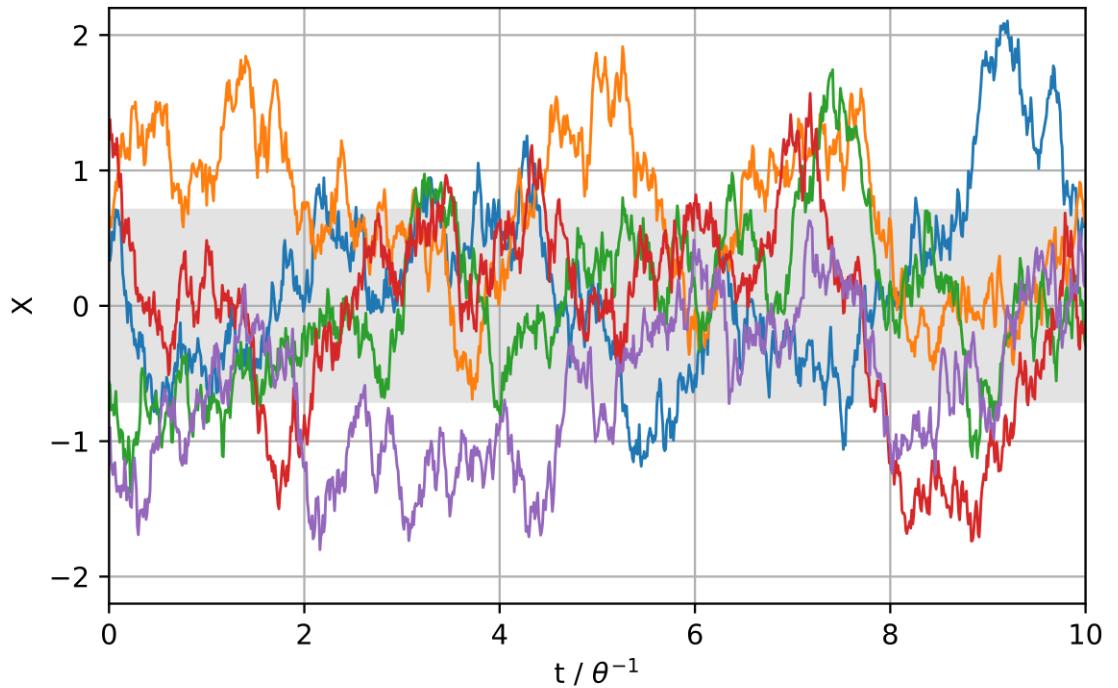


Figura 10: *Five simulations of the Ornstein-Uhlenbeck model with  $\Theta = 1$ ,  $\sigma = 1$ , and  $\nu = 0$ .*

The Ornstein–Uhlenbeck process  $x_t$  is defined by the following stochastic differential equation:

$$dx_t = -\theta x_t dt + \sigma dW_t \quad (16)$$

where  $\theta > 0$  and  $\sigma > 0$  are parameters and  $W_t$  denotes the Wiener Process. An additional drift term is sometimes added:

$$dx_t = \theta(u - x_t)dt + \sigma dW_t \quad (17)$$

where  $\mu$  is a constant. The Ornstein–Uhlenbeck process is sometimes also written as a Langevin equation of the form

$$\frac{dx_t}{dt} = -\theta x_t * \sigma n(t) \quad (18)$$

The term  $\eta(t)$ , also referred to as white noise, is used as a substitute for the hypothetical derivative  $dW_t/dt$  of the Wiener process. This is done in order to incorporate the Ornstein–Uhlenbeck process into stochastic differential equations.

However, it is important to note that the derivative  $dW_t/dt$  does not exist because the Wiener process is not differentiable anywhere, making the Langevin equation only heuristic in nature [Gardiner, 2010].

In physics and engineering fields, it is common practice to represent the Ornstein–Uhlenbeck process and similar stochastic differential equations by assuming that the noise term is a derivative of a differentiable interpolation (such as Fourier interpolation) of the Wiener process [Gardiner, 2010]. In an intuitive sense, focusing solely on the drift term, we can infer that the instantaneous change of  $x$ , denoted as  $dx(t)$ , relies on the difference between the current state  $x(t)$  and a fixed value  $\mu$ . If the difference  $(\mu - x(t))$  is positive,  $dx(t)$  will be positive as well, leading to an increase in  $x(t)$ . Conversely, if  $(\mu - x(t))$  is negative,  $dx(t)$  will be negative, causing  $x(t)$  to decrease. Consequently, the drift term acts as a force that drives  $x(t)$  towards the steady state or attractor value  $\mu$ . This property is why the Ornstein–Uhlenbeck process is regarded as a mean-reverting process.

It is worth noting that the parameter  $\theta$  governs the strength of the attraction towards the steady state  $\mu$ . With larger values of  $\theta$ , the difference  $(\mu - x(t))$  is amplified, resulting in a faster movement towards the steady state. Conversely, when  $\theta$  is close to zero, the attraction becomes weaker.

This characteristic leads to  $\theta$  being referred to as the dampening force or centralizing tendency, as it influences the rate at which the process converges towards the steady state value  $\mu$ .



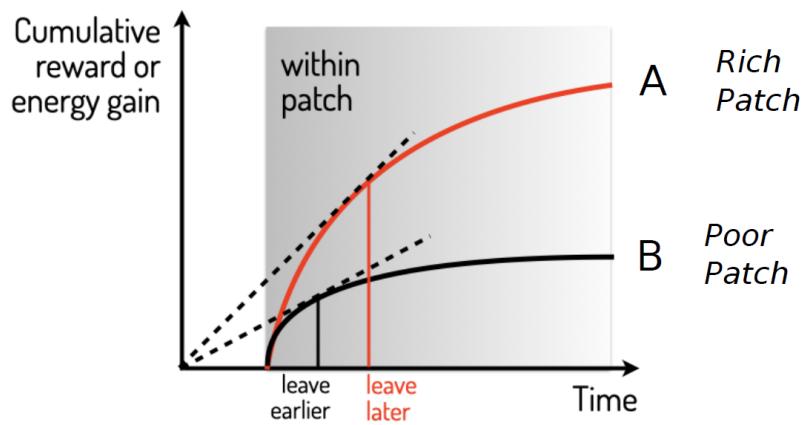
Figura 11: *Scanpath example generated by Ornstein-Uhlenbeck process*

### 3.7 Marginal Value Theorem

As mentioned earlier, Lévy Flights, also known as Correlated Random Walks (CB-Ws), are effective for modeling the spatial properties of foraging paths or scanpaths in patchy environments. However, they do not account for patch handling, which involves the decision-making process of when to move to the next patch based on the energy intake rate and travel costs. This exploration/exploitation dilemma has been addressed in Optimal Foraging Theories, such as Charnov's Marginal Value Theorem (MVT), which balances energy gain and consumption in patchy environments [Stephens and Krebs, 1986, Charnov, 1976].

Charnov's MVT states that a forager in a patchy environment should leave the current patch when the marginal rate of food intake drops to the long-term average rate of food gain across all patches in the environment. The MVT assumes the existence of multiple patches differentiated by their quality and characterized by a gain function, which represents the expected energy gain for a specific patch at time  $t$ . The gain function is assumed to be a well-defined, continuous, deterministic, and negatively accelerated function, reflecting the decrease in the rate of energy intake as

time increases.



Charnov [Charnov, 1976] demonstrated that to maximize the average rate of energy intake, the forager should choose the patch residence time such that the marginal rate of energy gain function at the time of leaving equals the long-term average rate of energy gain for the entire habitat. This optimal residence time, also known as the giving-up time, can be found by drawing a tangent to the energy gain function and projecting the point of tangency onto the x-axis.

The graph in the previous Figure illustrates this solution using two types of patches, A and B, classified as "rich" and "poor," respectively. The travel time to the patches is assumed to be constant, starting at the same point for both patches.

The principles of the Marginal Value Theorem have been applied to describe patch leaving times in human visual behavior, particularly in visual search experiments. [Wolfe, 2013] showed that the MVT successfully predicts patch leaving times when the patches are considered roughly identical. However, as experiments introduce more complex structures such as different patch qualities or limited access to information, human behavior deviates from Charnov's Theorem. [Cain et al., 2012]) extended the MVT to quantify human strategies in multiple-target search, introducing the potential value theorem [McNamara, 1982]. They demonstrated that individuals searched longer when they expected more targets to be present.

In the next section, we propose a comprehensive computational model of attentive eye guidance based on the principles of Optimal Foraging Theory (OFT). Unlike previous studies that focused on specific visual search tasks, we examine human visual attentive behavior along the free viewing of videos, particularly conversations between people. In this dynamic environment, patches are represented by small regions of the stimuli containing objects of interest.

### 3.8 Summary

This chapter discusses the theoretical approach adopted for the project, focusing on the development of a predictive gaze modeling system using machine learning, particularly reinforcement learning. The section begins by introducing machine learning as a branch of artificial intelligence, explaining its three main types: supervised learning, unsupervised learning, and reinforcement learning. The emphasis is placed on reinforcement learning, which trains agents to make decisions based on rewards and penalties in interaction with the environment. The chapter explores various reinforcement learning algorithms, evaluating their suitability for gaze modeling and comparing their principles, mechanics, strengths, and limitations. The goal is to identify the most effective algorithm for accurate gaze behavior prediction. Additionally, implementation details and considerations for training the models using these algorithms are provided to optimize predictive accuracy and efficiency. Furthermore, the section delves into contextual multi-armed bandits (CMABs), an extension of the traditional multi-armed bandits problem. CMABs incorporate contextual information into the decision-making process, enabling more personalized and adaptive choices. Several adaptive algorithms for CMABs, such as Logistic UCB, Epsilon Greedy, Adaptive Greedy, and Thompson Sampling, are introduced and compared in terms of their performance. The section also explores stochastic models for eye movement data, highlighting the limitations of classical approaches like the Central Limit Theorem. Alternative stochastic models, including Lévy flights, are discussed for capturing the complex dynamics of eye movements. The connection between eye movement modeling and foraging strategies observed in animals is emphasized, suggesting that eye movements can be seen as a form of foraging for valuable information in a scene. Finally, the section touches on the concepts of foraging theory and movement ecology, where optimal foraging theory (OFT) and the Lévy Flight Foraging (LFF) hypothesis are introduced. These concepts provide insights into the underlying mechanisms and strategies employed by animals and eye movements during visual exploration. Overall, this chapter presents a comprehensive theoretical foundation for the project, encompassing machine learning, reinforcement learning, contextual multi-armed bandits, stochastic modeling of eye movements, and the connection between eye movements and foraging behavior. These concepts lay the groundwork for developing a predictive gaze modeling system that takes into account the complexities and dynamics of human visual attention.

# Capitolo 4

## The OUTS-CMAB Gaze Model

Consider a dataset consisting of multiple videos displaying social interactions, specifically conversational clips that include both audio and video components. The main focus of this chapter is to model the deployment of attention through gaze by a human subject who is viewing and listening to these clips. When humans find themselves immersed in realistic and ecological situations involving other individuals, attention deployment is aimed at monitoring the behavior, intentions, and emotions of others, even in the absence of a specific external task [Foulsham et al., 2010].

In such circumstances, the internal goal of the observer is to control attention in order to maximize the implicit reward obtained from focusing on signals that hold social value [Anderson, 2013]. Despite the experimental evidence supporting these tendencies, their general modeling remains challenging. In order to effectively employ the mechanisms of selection, integration, and sampling that underlie the multifaceted phenomenon of attention, sensory systems must cope with the influx of multimodal events, including visual and audiovisual stimuli, captured from the external world.

This chapter thus aims to model gaze and predict where a single observer directs his/her attention while viewing social interactions. The objective is to understand how humans allocate their attention in such contexts. Recall that our ultimate goal is to develop a model that can predict which areas (patches) of the video will attract the observer's attention. This requires an understanding of the mechanisms involved in attentional selection and integration, as well as the ability to handle the large amount of multimodal information present in the videos.

This research is highly relevant as it can contribute to a better understanding of human attentional processes during the observation of social interactions. Furthermore, the findings obtained from this study can have practical applications, such as in the field of human behavior analysis or the design of intelligent user interfaces.

In our daily lives, we actively direct our gaze to observe and gather visual information, including social cues such as others' emotions and intentions [Shepherd and Platt, 2007, Guy et al., 2019]. This process also implies that the dynamic pattern resulting from our continuous sampling of the visual environment provides valuable insights into plans, goals, interests, and potential sources of rewards, as well as expectations about future events [Kowler, 2011, Henderson, 2017], personality, and social traits.

In the context of conversations, [?] demonstrated that observers predominantly focus their gaze on the people in videos, particularly their eyes and faces, and that gaze fixations are synchronized with the individuals who are speaking. This finding is not surprising, as visually-mediated social interactions are not unique to humans and have likely evolved in early primate ancestors due to selective pressures associated with group living [Shepherd and Platt, 2007]. Modeling attention in such scenarios necessitates considering the value of social cues. Consequently, the question arises as to whether it is possible to extract the implicit value of multimodal cues that drive an observer's motivation from behavioral data.

Even before addressing this pressing inquiry, the audio-visual nature of these stimuli presents a challenge in terms of how gaze should be guided within the context of multimodal perception (audio and visual).

## 4.1 Model input: an environment of multimodal patches

The input to the model at time  $t$  is the multimodal landscape, which we define as the time-varying ensemble of audio-visual patches  $\mathcal{W}(t) = \{\mathcal{P}_p(t)\}_{p=1}^{N_p}$ . These serve as regions of gaze attraction. Each patch is shaped as a 2-D Gaussian with localisation parameter (mean)  $\boldsymbol{\mu}_p$  and shape parameter (covariance matrix)  $\boldsymbol{\Sigma}_p$ . One example is provided in Figure 12 displaying the set of computed patches  $\mathcal{W}(t)$  as Gaussian blobs that overlay the original video frame; the patches correspond to the current speaker's face, the faces of the listeners, the speaker's hand gesture, and a center-bias patch. It is worth noting that the model needs not to rely upon any specific technique for deriving the pre-attentive representation  $\mathcal{W}(t)$ , as long as it captures relevant social multimodal information within the scene (persons, speakers, gestures, etc.). For what follows, we will practically exploit the pre-attentive processing steps computed as described in [Boccignone et al., 2020, D'Amelio and Boccignone, 2021].

The fundamental research questions in this context can be summarized as follows:

- What determines the value of a patch?
- How is gaze guided within and between patches?



Figura 12: The forager's patchy environment. At any time  $t$  the forager captures the multimodal landscape of social interactions as a set of audio-visual patches that convey different social value (speakers, faces, gestures, etc.); patches are shown as coloured Gaussian blobs that overlay the original video frame. See text for details

## 4.2 The spatial behaviour of the attentive forager: how to walk within and between patches

The temporal unfolding of gaze trajectory can be phenomenologically described as a biased random walk occurring at different scales. This process involves fine-scale movements within concentrated "information patches" (exploitation), alternating with coarse-scale relocations between patches (exploration). This perspective considers gaze trajectories as traced by a composite forager that seeks resources distributed in a patchy manner. A composite forager is capable of transitioning between exploitation and exploration, switching scales in its foraging walk within patches or during large-scale movements between patches.

In our case, the forager is represented as a stochastic entity, and both exploitation and exploration are accomplished through a biased Brownian walk, specifically an Ornstein-Uhlenbeck (OU) process, adjusted to the appropriate scale.

The scale adaptive or switching OU-process can be compactly formalised as follows.

The Ornstein-Uhlenbeck (OU) process models the movements or "walk" of a variable (in this case, the observer's point of focus) within a field (the observed scene). Its primary characteristic is that it is a mean-reverting process, implying that the variable tends to return to a mean or "normal" position after deviating from it. This "mean" can be considered the "attractive position" or the center of a patch of information in our case. If the rate is high, the point of focus quickly returns to the center, indicating a strong connection to the current patch. A lower rate allows for broader exploration within the patch, indicating a weaker connection. The dispersion or spread of the OU process represents the range or extent of the observer's attention within and between patches. This is influenced by the variance of the process. In the exploitation phase (within a patch), the dispersion is low, reflecting a detailed investigation in a small area around the center of the patch. In the exploration phase (between patches), the dispersion is higher, indicating broader and more sweeping movements as the observer searches for new patches. The forager's decision making process can be summarised via the evolution of state variables  $z_t$  and  $p_t^*$ : the first one is a binary random variable accounting for the switching from within-patch exploitation ( $z_t = 0$ ) to between-patch relocation ( $z_t = 1$ ); the second variable indexes the patch chosen to be handled at time  $t$ , thus  $p_t^* \in \{1, \dots, N_P\}$ . Based on the state variables, the evolution of the focus of attention (FOA) denoted by the vector of spatial coordinates  $\mathbf{r}_F(t) = (x_F(t), y_F(t))$  describes the spatial dynamics of attention deployment.

Given the state  $(z_t, p_t^*)$ , and following [Boccignone et al., 2020], the spatial dynamics of attention deployment is in turn obtained by evolving the FOA position  $\mathbf{r}_F(t)$  over time through the state-dependent stochastic differential equation (Ornstein-Uhlenbeck)

$$d\mathbf{r}_F(t) = \mathbf{B}_{p^*}^{(z_t)}[\boldsymbol{\mu}_{p^*}^{(z_t)} - \mathbf{r}_F(t)]dt + \mathbf{D}_{p^*}^{(z_t)}(\mathbf{r}_F(t))d\mathbf{W}^{(z_t)}(t) \quad (19)$$

that defines a mean reverting trajectory,  $\boldsymbol{\mu}_{p^*}^{(z_t)}$  being the attractor location (center of mass of the selected patch). Clearly, when  $z_t = 1$  the attractor serves as the target of a large scale gaze relocation (saccade); when  $z_t = 0$ , the attractor constrains local patch exploitation (fixational gaze movements, smooth pursuit).

The  $2 \times 2$  matrix  $\mathbf{B}_{p^*}^{(z_t)}$  controls the strength of the attraction (drift) of  $\mathbf{r}_F$  towards the location  $\boldsymbol{\mu}$ ;  $\mathbf{D}_{p^*}^{(z_t)}$  is a  $2 \times 2$  matrix representing the diffusion parameter of the 2-D Brownian motion  $\mathbf{W}(t)$ . Precisely, for the 2-D mean-reverting O-U process,  $\mathbf{B}_{p^*}^{(z_t)} = (b_{x,p^*}^{(z_t)}, b_{y,p^*}^{(z_t)})^T$ ,  $\mathbf{D}_{p^*}^{(z_t)} = (\sigma^{(z_t)})^2 \mathbb{I}$ , with  $\mathbf{W} = (W_x, W_y)^T$  denoting independent Brownian processes. Equation (19) can be integrated so that the explicit evolution of  $\mathbf{r}_F(t) = (x_F(t), y_F(t))$  in time between 0 and  $t$  can be obtained by numerically advancing the gaze position with the update equation from  $t$  to  $t' = t + \delta t$ , i.e.  $\delta t$

time units later, and initial condition  $x_0 = x_F(t)$ :

$$\begin{aligned} x_F(t') &= x_F(t)e^{-b_{x,p^*}^{(z_t)}\delta t} + \mu_x(1 - e^{-b_{x,p^*}^{(z_t)}\delta t}) + \sqrt{\gamma_x(1 - e^{-2b_{x,p^*}^{(z_t)}\delta t})}z(t) \\ y_F(t') &= y_F(t)e^{-b_{y,p^*}^{(z_t)}\delta t} + \mu_y(1 - e^{-b_{y,p^*}^{(z_t)}\delta t}) + \sqrt{\gamma_y(1 - e^{-2b_{y,p^*}^{(z_t)}\delta t})}z(t) \end{aligned} \quad (20)$$

with  $z \sim \mathcal{N}(0, 1)$ . As to the O-U parameters the drift terms  $b_{x,p}^{(z_t)}$  and  $b_{y,p}^{(z_t)}$  are set proportional to the width of the patch  $p$  if  $s_t = 0$ , or proportional to the distance to the arriving patch, otherwise. The diffusion terms are  $\gamma_x^{(z_t)} = \frac{\sigma^{(z_t)}}{b_{x,p^*}^{(z_t)}}$ ,  $\gamma_y^{(z_t)} = \frac{\sigma^{(z_t)}}{b_{y,p^*}^{(z_t)}}$  with  $\sigma^{(z_t)}$  set proportional to the average distance between patches if  $z_t = 1$ ; equal to 1, otherwise.

The bias in gaze trajectory is introduced by audio-visual patches that moment-by-moment emerge as relevant and rewarding within the multimodal landscape. Exploiting the foraging framework has gained popularity in the attention literature, extending its significance beyond being a metaphorical concept [Wolfe, 2013].

Whilst the above formalisation, is apt to describe the *spatial behavior* of forager exploration at two different scales, the crucial question that is still open concerns how to choose a patch of interest together with the decision of when to leave the currently exploited patch.

#### 4.2.1 Decision-making of the forager

The mechanism governing the oculomotor behavior, namely the switching between patch exploitation and exploration of the landscape, is primarily driven by a decision-making process. This process involves a critical comparison between the anticipated reward from the currently exploited patch and the average reward that could be reaped from moving and exploring other patches within the landscape.

The choice between these two states of exploitation and exploration, and the subsequent selection of the next patch to exploit, constitutes the range of behavioral decisions available to a hypothetical forager. This decision-making process is aptly modeled by Contextual bandits, with the chosen algorithm, Thompson Sampling, striking an optimal balance between exploration and exploitation amongst various patches.

A critical aspect of our model is the mechanism that governs patch exploitation. According to stochastic foraging theory, the duration spent within a patch is determined by the perceived potential value of the patch. The potential value, in turn, is a function of the forager's current expectations about the number of items present in the patch and the ease with which they can be located.

However, defining what constitutes an "item" can be challenging when internal goals are considered. For instance, when examining a patch containing a speaker's face, "items" could represent a multitude of details, including facial features, expressions, lip movements, or spoken words. Adding to this complexity, during the local patch exploration phase, it is a nontrivial task to determine how many items are being processed concurrently by the gaze.

The model, while in alignment with foraging literature and its perceptual applications, chooses to abstract away from the specific mechanisms driving gaze behavior within a region of interest for a given task. Instead, it emphasizes capturing the relevant phenomenological aspects statistically. This strategic approach allows us to address the macroscopic picture without getting mired in the minutiae of specific details.

Another key component of the model proposed here is the employment of a Poisson process for the encounter of patches and items within them. This stochastic process, coupled with the associated exponential waiting times, acts as a bridge connecting gaze points to the global and local scene characteristics.

Here, each patch is modeled as an independent generator of a Poisson process. The number of items within a patch is sampled from a Poisson distribution, providing an estimate of the perceiver's immediate information gain within the patch. This gain, measured instantaneously, is subsequently compared to the average gain achievable over the entire landscape. This comparison becomes instrumental in deciding when to relocate to another patch and how to select the next patch for exploitation, thereby mimicking the moment-by-moment actions of a forager.

The model robustness stems from its ability to make these comparisons and choices, thereby striking a balance between exploration and exploitation, which are fundamental to successful foraging behavior.

Furthermore, by drawing from established theories in foraging and perception, our model achieves a high level of abstraction while still capturing the critical phenomenological aspects of gaze behavior.

The choice of the Thompson Sampling algorithm is crucial. Recall from Chapter 3.2.3, that the key idea behind Thompson Sampling is to maintain a posterior distribution over the expected reward of each arm based on the rewards observed so far. At each step, an arm is chosen according to the current probability that it has the highest expected reward.

In Thompson Sampling, the balance between exploration and exploitation comes from the uncertainty in the posterior distributions. Arms that have not been tried much will have a wider distribution, leading to a higher chance of being selected and hence more exploration. On the other hand, arms with high expected rewards will also have a high chance of being selected, leading to exploitation.

More precisely here we use the following.

The environment provides a “ signal ” in terms of a context, namely  $\tilde{\mathbf{s}}_t \in \mathbb{R}^d$  [?]. Thus we can assume that the state  $\mathbf{s}_t \in \mathbb{R}^d$  is a function of the observed context and the action at time  $t$ ,

$$\mathbf{s}_t = \psi(\tilde{\mathbf{s}}_t, \mathbf{a}_t)$$

For example, in the most simple case,

$$\mathbf{s}_t^T = \left[ 1(\mathbf{a}_t = 1) \tilde{\mathbf{s}}_t^T, \dots, 1(\mathbf{a}_t = K) \tilde{\mathbf{s}}_t^T \right].$$

The introduction of the context lies at the heart of the definition of *contextual bandit*

$$\langle \mathcal{S}, \mathcal{A}, \mathcal{R} \rangle$$

and the MAB problem turns into that of a *contextual MAB* (C-MAB).

At the most general level we can assume a distribution on states  $\mathcal{S} = P(\mathbf{s})$  and a distribution on rewards  $\mathcal{R}_{\mathbf{s}}^{\mathbf{a}} = P(\mathbf{r} \mid \mathbf{s}, \mathbf{a})$ .

The interaction between the forager and the environment at any time  $t$ , unfolds as follows

1. the environment generates a state  $\mathbf{s}_t \sim \mathcal{S}$ ;
2. the forager selects action  $\mathbf{a}_t \in \mathcal{A}$
3. the environment generates a reward  $\mathbf{r}_t \sim \mathcal{R}_{\mathbf{s}}^{\mathbf{a}}$

The forager exploits observations and the reward related to each action selected in the past to make decisions in the present. The forager’s goal is to gather enough information related to the context vector and possible reward so to predict the highest expected reward action, given the state vector.

In this case the action-value function can be written as  $Q(\mathbf{s}, \mathbf{a}) = E[\mathbf{r} \mid \mathbf{s}, \mathbf{a}]$

We can thus formalise the TS algorithm for CMAB as follows. Consider the update of the posterior distribution over parameters  $\theta$ ,

$$P(\theta \mid \mathbf{s}_t, \mathbf{r}_t, \mathbf{a}_t) \approx P(\theta) P_R(\mathbf{r}_t \mid \mathbf{s}_t, \mathbf{a}_t, \theta) \quad (21)$$

We rewrite the TS-CMAB as in Algorithm ( 1 )

Generally speaking, in our model, each patch has a posterior distribution over its expected reward, which is updated based on the observed rewards. The decision of where to focus next is then made based on these distributions. By using the Thompson Sampling algorithm, the model can effectively balance the need to explore new patches (to gain more information) and the need to exploit known patches (to maximize the information gain). This makes the model highly effective in situations where the forager needs to adapt to changes in the environment and dynamically

---

**Algorithm 1** Thompson Sampling for contextual MAB

---

```

0: procedure TS-CMAB( $\mathcal{S}, P, P_R, \mathbf{r}$ )
0:   Initialise the posterior distribution  $P(\theta)$ 
0:   for  $t = 1, 2, \dots$  do
0:     for  $k = 1, 2, \dots, K$  do
0:       observe the context  $\tilde{\mathbf{s}}_t$ 
0:       compute  $\mathbf{s}_t = \psi(\tilde{\mathbf{s}}_t, \mathbf{a}_{t,k})$  {model inference}
0:        $\theta^* \sim P(\theta | \mathbf{s}_t, \mathbf{a}_{t,k})$  {sample  $\theta$ }
0:        $\hat{\theta} \leftarrow \theta^*$ 
0:     end for {select and apply action}
0:      $\mathbf{a}_t \leftarrow \arg \max_{\mathbf{a}} E_{P_R} [\mathbf{r}_t(\mathbf{s}_t) | \mathbf{s}_t = \mathbf{s}, \mathbf{a}_t]$ 
0:     Apply  $\mathbf{a}_t$  and observe  $\mathbf{r}_t = \mathbf{r}_t(\mathbf{s}_t, \mathbf{a}_t)$  {Update the posterior distribution P}
0:     Update  $P(\theta | \mathbf{s}_t, \mathbf{a}_t)$  via Bayes (Eq. 21)
0:   end for
0: end procedure=0

```

---

adjust its strategy. In order to exactly specify the needed context vector and the reward function, we further draw on optimal foraging and the Marginal Value Theorem

#### 4.2.2 Optimal Foraging Strategies and the Marginal Value Theorem

The decision-making process of foraging animals, particularly in relation to whether to stay or leave a current patch, presents a fundamental problem in ecology. Ecologists have long sought to understand how animals efficiently search for and exploit food patches within their environment. This problem encompasses the handling of patch cycles, which initiate when an animal departs from a patch to search for a new one. Once a patch is discovered, the animal's energy gain rate gradually declines as the food resources deplete, eventually leading to the departure from the patch and the start of a new cycle [MacArthur and Pianka, 1966]. The MVT suggests that foragers should exploit patches in a manner that maximizes the net rate of energy gain and predicts the optimal patch residence time. The MVT states that foragers should move to a new patch when the marginal rate of food intake drops to the long-term average rate of food gain across multiple patches in the environment. In this simplified model, the gain in energy serves as a proxy for the fitness of the foragers. It assumes that foragers possess knowledge about the environment, including the quality of different patches and the time required to travel between them. Consequently, the Marginal Value Theorem (MVT) predicts that the quality of a patch influences the decision to leave. Specifically, a poor patch that offers lower energy gain should be abandoned

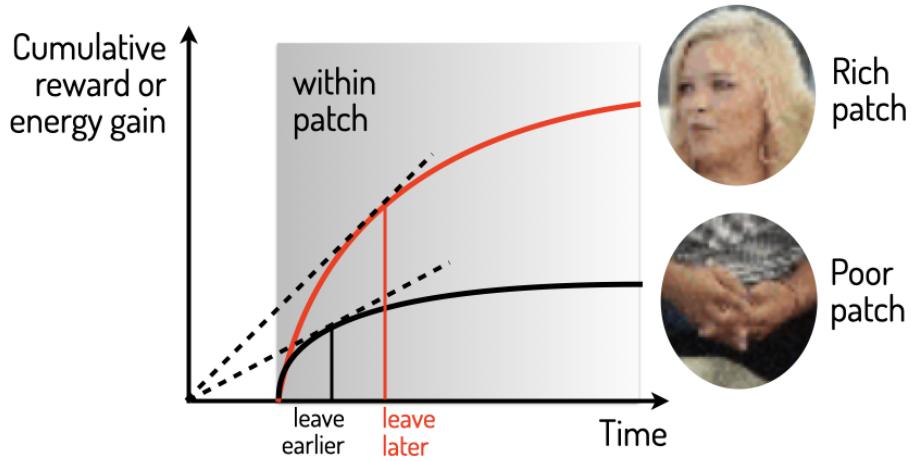


Figura 13: According to the Marginal Value Theorem (MVT), the prediction is that a patch with lower food resources, referred to as a "poor" patch, should be abandoned earlier compared to a patch with higher food resources, known as a "rich" patch. In terms of the time axis, the forager initially spends time traveling with no energy gain until it discovers a patch. The cumulative rewards obtained from exploiting the resources in the patch are represented by the shapes of the gain curves, with the red curve representing a rich patch and the black curve representing a poor patch. The optimal patch residence time is determined by the point of tangency where the tangent intersects the curve.

earlier. It is evident that foragers who remain in a patch for too long incur an opportunity cost, as they waste time exploiting a depleted patch while potentially missing out on fresher, unexploited patches.

In a stochastic environment where rewards are uncertain and do not follow a smooth pattern, an optimal forager should employ probabilistic reasoning when evaluating the value of a patch within the broader environment [McNamara, 1982]. The optimal time to leave a patch occurs when the expected rate of energy gain, rather than the observed rate, falls below the average rate for the environment. Stochastic foraging models typically consider the net energy gain  $G$  and the time taken to complete a foraging cycle  $T$  as random variables whose distribution depends on the foraging animal's behavioral strategy. The net energy gain,  $G$ , is influenced by the forager's time-dependent state, denoted as  $U(t)$ , which represents the forager's accumulated experience up to time  $t$  ( $G(U(t))$ ). For example,  $U(t) = u$  might signify the number of items or prey captured by the forager. The mean net rate of energetic gain, or the mean reward rate, achieved by the animal is expressed as the ratio of expectations  $E[G]/E[T]$ .

Considering a stochastic perspective in foraging studies proves useful in understanding the instantaneous reward rate and incorporating probabilistic reasoning into decision-making processes [McNamara, 1982].

$$g(u, t) = \lim_{\delta t \rightarrow 0} \frac{E[G(U(t + \delta t))|U(t) = u] - G(u)}{\delta t} \quad (22)$$

The expected reward rate over the next time interval, denoted as  $\delta t$ , serves as the stochastic counterpart to the continuous energy intake rate ( $\partial G/\partial T$ ) considered in the Marginal Value Theorem (MVT). This definition captures the probabilistic nature of foraging by estimating the average reward that the forager anticipates to gain within the upcoming time interval  $\delta t$ . It provides a stochastic representation of the continuous energy intake rate considered in the MVT.

The general rule followed by the forager while examining a patch is to depart from the patch when

$$g(u, t) < Q(t) \quad (23)$$

The decision to leave a patch occurs when the instantaneous reward rate drops below a predetermined "quality" threshold, denoted as  $Q$ . This threshold, which determines the perceived value of the patch, is influenced by factors such as the overall richness of the environment, the distance between patches, and other relevant considerations like predation risk. One method, utilized in Bayesian foraging approaches [Iwasa et al., 1981, Rodríguez-Gironés and Vázquez, 1997], involves calculating  $g(u; t)$  where  $u$  represents the experiential state of the forager. For instance, when considering a patch containing a discrete number of items, denoted as  $m$ , the experiential state  $U$  can be represented by the pair  $(n; t)$ , where  $n$  represents the number of items "consumed" within the time period  $t$ . In this context,  $G(U(t))$  is expressed as  $G(n; t)$ , and  $g(u; t)$  is denoted as  $g(n; t)$ . If the forager spends a duration  $t_{WP}$  within the patch, the remaining items can be represented by  $k = m - n$ . By considering these variables and their relationships, concrete methods can be developed to assess the optimal time for leaving a patch based on the available items and the elapsed time. A straightforward method involves considering a patch that contains a discrete number of items, denoted as  $m$ . Let  $n$  represent the number of items "consumed" within a given time period  $t$ . In this context, the experiential state  $U$  can be represented as the pair  $(n; t)$ , where  $G(U(t))$  is equivalent to  $G(n; t)$  and  $g(u; t)$  is equal to  $g(n; t)$ . If a forager spends time  $t_{WP}$  within the patch, the remaining items can be represented by  $k = m - n$ .

When foragers search for food items randomly, it is assumed that the time required

to find one item follows an exponential distribution.

$$P(T \in [t, t + \delta t]) = \lambda e^{-\lambda t} dt = A k e^{-A k t} dt, \quad (24)$$

where the rate  $\lambda = A k$  depends on  $A$ , the searching efficiency of the forager. The probability of capturing at least one item, conditionally on the  $k$  remaining, is  $P(\delta t | k) = 1 - e^{-A k t}$  it been calculated [Iwasa et al., 1981, Rodríguez-Gironés and Vázquez, 1997] that if the initial distribution of  $m_p$  items in patch  $p$  (prior, with  $k = 0$ ) follow a Poisson law,  $Pois(P_p) = \frac{e^{-P_p} P_p^{m_p}}{m_p!}$  then simply

$$g_p(t w_p) = \rho_p e^{A t w_p} \quad (25)$$

where  $\rho_p = \kappa v_p(t w_p)$  is the value of the patch  $p$  at time  $t w_p$  opportunely scaled by a positive constant  $\kappa$ .

The relationship between the foraging efficiency parameter  $A$  and the forager's behavior can be observed from Equations. The parameter  $A$  influences the rate at which the forager transitions from one item to another, ultimately affecting the instantaneous intake rate. However, studies have shown that individuals tend to concentrate their foraging efforts in areas with high rewards [de Knegt et al., 2007, Kazimierski et al., 2016]. This behavior leads to an increase in the handling time of each item, consequently prolonging the expected time until the next item is encountered within the patch. In our case, this effect is accounted for by setting  $A = \frac{\emptyset}{v_p(t)}$ , recalling that  $v_p(t) \in [0; 1]$  is the value associated to the patch  $p$  at time  $t$ , while  $\emptyset$  is a positive constant defining the baseline foraging efficiency.

The expected average gain from the environment for all patches  $q$  except the current one can be obtained by considering the potential intake rate at  $t_W = 0$ , i.e., via Eq. 25  $g_q(0) = \rho_q$ ,  $q \neq p$ :

$$Q(t) = \frac{1}{N_p - 1} \sum_{q \neq p} \rho_q(t). \quad (26)$$

Rather than straightforwardly use the deterministic rule given in Eq. ??, we allow the forager to perform a probabilistic decision; namely the behavioural state decision  $s(t) \in \{0, 1\}$  is sampled following a Bernoulli law,  $Bern(s(t) | \xi(t))$ . The parameter  $\xi$ , denoting the prior probability of staying within the patch is obtained using a logistic rule accounting for a stochastic comparison on the difference  $g_p(t_{W_p}) - Q$ , thus

$$\xi(t) = P(stay | g(t), Q(t)) = \frac{1}{1 + e^{-\beta(g_p(t_{W_p}) - Q(t))}}, \quad (27)$$

$$s(t) \sim Bern(\xi(t)). \quad (28)$$

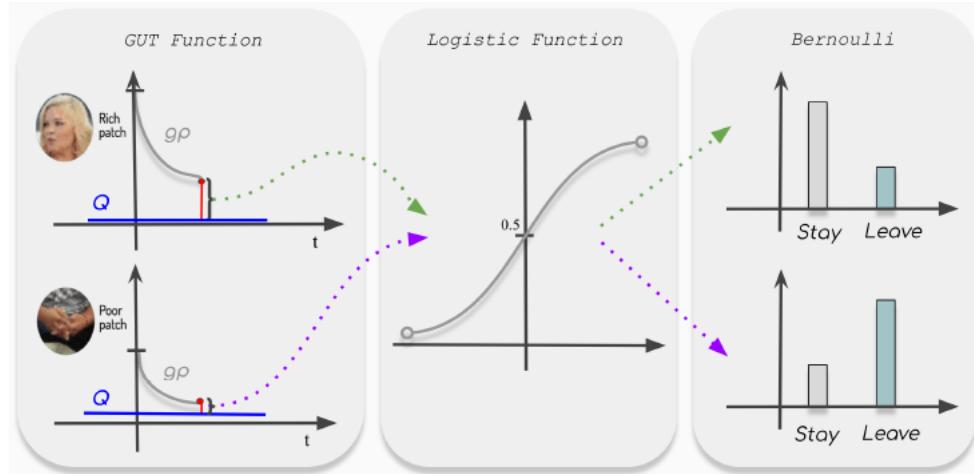


Figura 14: *The switching behavior can be described as follows. The first block illustrates the typical pattern of the instantaneous reward rate for two types of patches: rich and poor. These patterns can be seen as Giving Up Time (GUT) functions, which represent the time it takes for the reward rate to approach the quality threshold  $Q$ . For poorer patches, the GUT function reaches the threshold faster compared to richer patches. The decision to stay or go is made at each time step by sampling a Bernoulli random variable, as shown in the third block. The parameter of this random variable is determined by the distance between the GUT function and the quality threshold at that specific time, suitably scaled by a logistic function (as depicted in the second block). This decision-making process allows the forager to assess whether the reward rate is above or below the threshold, influencing the choice to remain in the patch or move on to a new one.*

### 4.3 Gaze model operationalisation: Context and Reward

The process of model execution commences with the identification and extraction of areas of interest, referred to as 'patches', from each source of audio-visual data. Each patch holds its unique significance, as emphasized by the principles of Foraging Theory. Herein, every patch is assigned an inherent value that portrays its appeal or "palatability" with respect to a specific resource under consideration.

Elaborating further on this theoretical framework, it becomes crucial to establish a comprehensive context for each individual patch. This is achieved by leveraging features intrinsic to the patch, as well as its spatial relationship to the current focal point of the observer. Thus, the model considers not only the inherent characteristics of the patch but also the dynamics of the observer's gaze at a given instant.

### 4.3.1 The Context Vector

Each patch  $p \in 1 \cdots N_p$  at time  $t$  is characterized as a vector  $\mathbf{c}_{p,t} \in \mathcal{R}^3$ :

$$\mathbf{c}_{p,t} = (\ell_p, d_{p,p^*}, \phi_{p,p^*}) \quad (29)$$

where:

- $p \in 1 \cdots N_p$ : the Patch ID, namely the identifier that is assigned to each distinct patch in the scene. Each patch carries its unique characteristics, and the Patch ID helps in tracing and analyzing these individual attributes;
- $p^*$ : the currently observed/exploited patch;
- $\ell_p \in (1, \dots, N_\ell)$  identifies the priority map from which the patch  $p$  is generated. This can be seen as the 'type' of the patch, defining its nature or category within the scope of our attention model.
- Euclidean Distance  $d_{p,p^*} = \|\boldsymbol{\mu}_p - \boldsymbol{\mu}_{p^*}\|$ : it measures the shortest 'as-the-crow-flies' distance between the observer's current gaze position and the proposed patch. It offers a quantifiable way to estimate the 'effort' or 'cost' that the observer might need to incur to shift their attention from their present focus to a new patch. This estimated cost plays a crucial role in the decision-making process, particularly when the observer is deciding whether to continue scrutinizing the current patch (exploitation) or move their gaze to explore a new one (exploration).
- angle  $\phi_{p,p^*}$ : the degree of deviation between the observer's current line of vision and the direction towards the proposed patch. This angular measurement is vital as it encapsulates the relative positioning of various patches from the observer's current viewpoint. It provides an added dimension to the spatial context of the scene, thus helping paint a more detailed, nuanced picture of the observer's visual environment and the potential decision-making process that guides their attention allocation.

At the beginning of each decision round the world generates a context vector  $\tilde{\mathbf{s}}_t$  of fixed dimension obtained by joining every feature vector  $\mathbf{c}_{p,t}$  associated to each patch:

$$\tilde{\mathbf{s}}_t = [\mathbf{c}_{1,t} | \cdots | \mathbf{c}_{p,t} | \cdots | \mathbf{c}_{N_p,t}] \quad (30)$$

It is hence assumed that there is a fixed number of choices (arms or patches), from which an agent must choose one as his action in each round  $t$ .

### 4.3.2 The reward function

The reward function quantifies the 'goodness' or 'value' of a state-action pair and is fundamentally related to the expected reward rate  $g(u, t)$ . It incentivizes the learning agent to make choices that maximize this function, reflecting the forager goal to maximize the average reward rate. The reward function provides an operational definition of the goal of an agent and plays a significant role in determining the policy that the agent learns. It is a motivational impulse that guides a general observer to choose one patch over another at any given moment, as previously mentioned, dependent on an implicit gratification tied to social value characteristics. In the proposed model, these are represented as patches containing objects of interest, specifically the faces of speakers/non-speakers and lower-level cues such as movement, colors, texture, etc. This approach aligns with the principles of the Optimal Foraging Theory (OFT), especially when observing dynamic environments like conversations among people in videos. In this context, the patches are dynamically interesting areas, and the observer exerts cognitive effort to shift attention from one area to another. The reward function, therefore, becomes a measure of the satisfaction or reward that the observer derives from focusing on a specific patch of the scene. The interest in a patch, and thus its reward function, changes based on its perceived appeal. Therefore, the reward function can be viewed as an analogous concept to the rate of energy intake. By adopting principles from the Marginal Value Theorem, the agent (or the observer) should redirect its attention to a new patch when the anticipated reward (the "marginal value") from the current patch falls below the average expected reward from the entire scene. The problem posed, therefore, is how to define a reward function able to mine the connection between such hidden and somewhat abstract quantities (e.g. interest in a patch according to an endogenous goal) to the visible characteristics of the viewed scene. Notably, the aim of this work is to eventually show that this can be learnt directly from eye tracking data.

Specifically, it is assumed that at the beginning of each round the world (the viewed scene) generates a set of covariates  $\tilde{\mathbf{s}}_t$  (the context). Each patch (action) is associated with an arm of a C-MAB model. The latter are operationalized via parametric functions of the context  $f_p(\tilde{\mathbf{s}}_t, \theta_p)$ , where  $\theta$  is a vector of parameters. Stochastic binary rewards are then generated for each arm ( $\mathbf{r}_{p,t} \in \{0, 1\}$ ) through a function of the context vector:

$$r_{p,t} \sim Bernoulli(f_p(\tilde{\mathbf{s}}_t))$$

which is different for each arm but is the same throughout all rounds. A common approach for dealing with scenarios exhibiting binary rewards is to use binary classification algorithms such as logistic regression as black-box oracles. Hence, the probability of receiving a reward after pulling the  $p$ -th arm is specified as:

$$p(r_{p,t} = 1 | \tilde{\mathbf{s}}_t; \boldsymbol{\beta}_p) = \frac{1}{1 + e^{-(\beta_{0,p} + \boldsymbol{\beta}_p \cdot \tilde{\mathbf{s}}_t)}} \quad (31)$$

Where  $\boldsymbol{\beta}_p$  is the vector of parameters of the Logistic Regression model associated to the  $p$ -th arm. The agent chooses an arm based on his previous knowledge, and the reward for the arm that was chosen is revealed, while the rewards for the other arms remain unknown.

The "reward reveal" is here intended as the actual action (patch) that was chosen by a human observer that is glimpsing at the same scene. Such quantity can be measured by resorting to eye-tracking data. If the action chosen by the agent and the human observer match, the agent earns a positive reward ( $r_t = 1$ ), no reward is obtained otherwise ( $r_t = 0$ ).

Subsequently, the agent is put in the context of the "correct" action (the current patch is set to the patch chosen by the human observer) and a new decision round starts.

Eventually, the expected reward  $E_{P_R} [\mathbf{R}_{p,t}(\tilde{\mathbf{s}}_t) | \tilde{\mathbf{s}}_t = \mathbf{s}, \mathbf{a}_t]$  associated with each arm at each time instant  $t$  delivers an approximation of the value of that patch:

$$v_p(t) \sim E_{P_R} [\mathbf{R}_{p,t}(\tilde{\mathbf{s}}_t) | \tilde{\mathbf{s}}_t = \mathbf{s}, \mathbf{a}_t] \quad (32)$$

The latter governs the expected reward rate of each patch as defined in Equation 25 through the  $A$  parameter.

As time progresses, the uncertainty about the true rewards decreases, leading the Thompson Sampling algorithm to better estimate the expected reward rate  $g(u, t)$  and select actions that yield the highest expected rewards. The actions represent decisions concerning the next patch to focus on, and the expected rewards equate to the values computed from the logistic functions based on the patches' contexts. By helping to make decisions that optimally balance exploration (investigating less 'rewarding' or less 'known' patches) and exploitation (concentrating on the current or known 'rewarding' patches), Thompson Sampling aligns seamlessly with the foraging behavior that this model aims to simulate. The Contextual Multi-Armed Bandit using the Thompson Sampling algorithm samples the expected values obtained at each moment in time. This expected value for each patch, denoted as  $v_p(t)$ , evolves over time based on the evolving context and choices made, ultimately reflecting the changing expected reward rate  $g(u, t)$

### 4.3.3 The OUTS-CMAB Gaze Model at the algorithmic level

The operationalization of the OUTS-CMAB model is outlined in Algorithm 2.

In a nutshell the audio-visual patches are computed and considered as the set of potential action to be chosen by the agent. For each action the Expected reward

(ER) is computed by the model according to Equation: 31. We consider the ER as the value of each patch (Equation: 32). The next step was to compute the ER for each patch based on its context (Equation: 31), which is a logistic regression-based estimation of the probability of receiving a reward from a specific patch given the context. The value of each patch (Equation 32) was set equal to this ER, indicating the reward the system anticipates to receive by choosing that particular patch. From there, an average of the expected rewards from all the patches, referred to as  $Q(t_n)$  (Equation 26), was calculated to provide a baseline against which the actual gain from each patch could be compared.

The gaze control algorithm then proceeded to the attentive stage where the system decided whether to exploit the current patch or explore other patches. If the current behavioral state was set to exploitation ( $s(t_n) = 1$ ), the system adjusted its parameters for Ornstein-Uhlenbeck (OU) sampling based on the current patch and behavioural state. As long as the system was within the patch, it continued the exploitation stage by shifting the gaze location using OU sampling. At each step of this local gaze shifting, the instantaneous expected gain  $g_p(t_{W_p})$  (Equation: 25) from the current patch was computed. This gain was then compared against the expected average gain  $Q$  from the environment (Equation: 28) to make a decision about whether to stay within the current patch or move to another one. A new behavioral state was sampled at each timestep  $t_n$  (Equation: 28), which helped decide between exploitation and exploration. If the system was in the exploration state, it would select the next patch with the highest expected value based on the Thompson Sampling (TS) rule. The parameters for the OU process were adjusted based on the state and the selected patch. The system then continued with the exploration stage, relocating the gaze location using OU sampling. The described procedure was repeated throughout the duration  $T$  of the simulated scenario, resulting in a prediction of the gaze location for each frame in the sequence. By integrating the Contextual Bandit model with the gaze control algorithm and utilizing Thompson Sampling, the system achieved effective visual exploration by striking a balance between exploitation and exploration in a diverse landscape. This approach enabled the system to dynamically adapt to the complexity and variability present in real-world visual environments, thereby enhancing its usefulness in a wide range of applications.

---

**Algorithm 2** Gaze control in a multimodal landscape

---

**Input:** Visual stream  $\{\mathbf{I}\}$ , audio stream  $\{\mathbf{A}\}$ , the duration  $T$  to be simulated, the video frame rate  $FR$ , the random walk sampling rate  $fs$ .

**Output:** Prediction of gaze

```

0: procedure OUTS-CMAB
1:  $\delta t = \frac{1}{FR}$ ,  $\delta u = \frac{1}{fs}$ 
2: Initialisation of first gaze location  $\mathbf{r}(t_1)$  on patch  $p(t_1) = j$ , with behavioural state  $s(t_1) = 0$  {Exploitation mode}
3: for  $n = 2$  to  $\frac{T}{\delta t}$  do
4:   {Landscape evaluation}
5:   Compute audio-visual patches  $\{\mathcal{P}_p(t_n)\}$  as potential actions (arms)
6:   Compute context vector  $\tilde{\mathbf{s}}_t$  (Equation 30)
7:   {Expected reward (ER) inference}
8:   Infer for each patch the ER from the context (Equation 31)
9:   Set each patch value to the predicted ER (Equation 32)
10:  Compute the expected average gain  $Q(t_n)$  from all patches in the landscape (Eq. 26)
11:  {Attentive stage}
12:  if  $s(t_n) = 1$  then
13:    {Exploitation: patch handling}
14:    Set the parameters  $\mu_p^{(z_t)}, \mathbf{B}_p^{(z_t)}, \Psi_p^{(z_t)}$  for OU sampling according to state  $s(t_n)$  and current patch indexed by  $p(t_n)$ 
15:    while within patch do
16:      {Exploitation: local gaze shifting }
17:      for  $j = 0$  to  $\frac{fs}{FR}$  do
18:        Sample the OU gaze relocation
19:         $\mathbf{r}_F(t_{n-1} + (j \times \delta u)) \rightarrow \mathbf{r}_F(t_{n-1} + (j + 1 \times \delta u))$ 
20:      end for
21:      {Behavioural state sampling}
22:      Compute the instantaneous expected gain  $g_p(t_{W_p})$  for current patch (Eq. 25)
23:      Compare current patch gain against the expected average gain  $Q$  from the environment (Eq. 28)
24:      Sample the behavioural state  $s(t_n)$  at time  $t_n = t_{n-1} + \delta t$  (Eq. 28)
25:    end while
26:  else
27:    {Exploration: patch-choice}
28:    Sample next most valuable attractor  $p(t_{n-1} + \delta t)$  according to TS rule
29:    Set the parameters  $\mu_p^{(z_t)}, \mathbf{B}_p^{(z_t)}, \Psi_p^{(z_t)}$  for OU sampling according to state  $s(t_n)$  and attractor  $p(t_n)$ 
30:    {Exploration: relocation gaze shifting }
31:    for  $j = 0$  to  $\frac{fs}{FR}$  do
32:      Sample the OU gaze relocation
33:       $\mathbf{r}_F(t_{n-1} + (j \times \delta u)) \rightarrow \mathbf{r}_F(t_{n-1} + (j + 1 \times \delta u))$ 
34:    end for
35:  end if
36: end for
37: end procedure=0

```

---

## 4.4 Summary

This chapter focuses on modeling the deployment of attention through gaze by a human subject during the viewing of social interactions. The goal is to understand how humans allocate their attention in such contexts and develop a model that can predict which areas of the video will attract the observer's attention. The chapter discusses the challenges of attentional selection and integration in the context of multimodal stimuli, including visual and audiovisual information. The chapter starts by highlighting the importance of attention deployment in monitoring the behavior, intentions, and emotions of others during social interactions. It emphasizes the role of attention in maximizing the implicit reward obtained from focusing on signals that hold social value. However, modeling attention in such scenarios is challenging due to the influx of multimodal events and the need to handle the large amount of information present in the videos. The chapter proposes modeling gaze behavior as a foraging process, where attention deployment is compared to a composite forager seeking resources distributed in a patchy manner. The foraging process involves fine-scale movements within concentrated "information patches" (exploitation) and coarse-scale relocations between patches (exploration). The chapter addresses the research questions of what determines the value of a patch and how gaze is guided within and between patches. To model gaze behavior as a foraging process, the chapter introduces the concept of the Marginal Value Theorem (MVT), which predicts optimal foraging strategies. The MVT states that foragers should move to a new patch when the marginal rate of reward drops to the average rate across multiple patches in the environment. The chapter discusses how the MVT can be applied to gaze behavior and decision-making in the context of social interactions. The chapter also discusses the challenges of modeling patch exploitation and the difficulty of defining what constitutes an "item" within a patch in the context of social interactions. It explains that the model focuses on capturing relevant phenomenological aspects in statistical terms rather than delving into specific mechanisms underlying gaze behavior. The chapter concludes by highlighting the significance of stochastic foraging models in understanding attention deployment and decision-making. It emphasizes the importance of probabilistic reasoning and estimating the expected reward rate within a patch to determine optimal foraging strategies. Overall, this chapter proposes a foraging-based model to understand and predict gaze behavior during the viewing of social interactions. By applying concepts from foraging theory and the Marginal Value Theorem, the model aims to provide insights into the mechanisms of attentional selection and integration in multimodal contexts.

# **Capitolo 5**

## **Results**

In this chapter, we present the results of our study on foraging behavior in patchy environments. We begin by providing an overview of the dataset used in the experiments and describing any preprocessing steps performed. Then, we discuss the results of the training phase, including the observers involved and the model's performance evaluation. Next, we present the results of the test phase, focusing on the test observers and the evaluation of the model's performance on the test data. We also analyze the model's ability to predict observer focus based on task-driven behavior experiments. Additionally, we explore the model's adaptability to dynamic patches and present the results of experiments involving the addition and dropping of patches. Finally, we discuss the implications of the results and address any limitations or challenges encountered during the experiments. First of all was been necessary to do the Quality evalutation. In this stage I need to make the mean of reward.

### **5.1 Dataset Description**

We begin by providing an overview of the dataset utilized in our study, which encompasses a collection of 78 videos capturing various foraging scenarios in patchy environments. These videos were meticulously chosen to encompass a wide range of ecological conditions and diverse foraging strategies, ensuring the representation of different scenarios in our analysis.

In addition to the aforementioned foraging dataset, we have incorporated another dataset consisting of 65 one-shot conversation scenes. Each scene in this dataset involves the presence of 1 to 27 distinct faces. To maintain consistency, the duration of the conversation videos was uniformly reduced to approximately 20 seconds. The resolution of these videos stands at 1280 by 720 pixels, with a frame rate of 25 frames per second.

To enhance the dataset's richness and relevance, eye-tracking recordings were obtained from 39 individuals who participated in the study. These participants, unaware of the experiment's objectives, had their eye fixations' positions and durations recorded using a Tobii X2-60 eye tracker operating at 60 Hz. The recorded eye-tracking data provides valuable insights into the participants' visual attention during the video scenes.

These parameters include the baseline foraging efficiency (represented as  $\Phi$ ), the logistic growth rate (denoted as  $\beta$ ), and the steepness of the exponential factor influencing patch visibility (represented as  $\kappa$ ). Through a grid search process, we maximized the metric scores using the procedure outlined in Section 5.2, which led us to identify the optimal parameter values:  $\Phi = 3$ ,  $\beta = 20$ , and  $K = 1$ . Was been used 10 subjects for training and the remaining 29 subjects from the eye-tracking dataset were exclusively reserved for the test evaluation phase of our study. This separation enables an unbiased assessment of the model's performance on unseen data and further strengthens the reliability and generalizability of our findings.

By incorporating these two datasets, we aim to provide a comprehensive and diverse collection of videos, along with corresponding eye-tracking data, to facilitate an in-depth analysis of foraging behavior and its connection to visual attention.

## 5.2 Qualitative Analysis

In order to conduct a comprehensive evaluation, a qualitative analysis was conducted. The analysis began by examining the cumulative rewards to identify areas where the best possible results could be achieved. The average reward distribution showed a clear emphasis on the "speakers," as expected, since they tended to generate a higher number of rewards compared to the non-speakers. This discrepancy in rewards can be attributed to the larger number of non-speaking subjects present in the examined videos.

An analysis of the distribution of average rewards per patch revealed a substantial emphasis on the 'speakers.' This outcome was not unexpected, as the 'speakers,' or the subjects in the audiovisual content that were actively speaking, are typically associated with higher information content and therefore tend to be assigned higher rewards.

The distribution of these rewards, however, wasn't equal. The 'speakers' received higher average rewards per patch compared to the 'non-speakers.' This discrepancy can be attributed to the presence of a larger number of non-speaking subjects within the examined videos, causing the lower average reward for the non-speakers.

Nonetheless, when considering cumulative rewards— which factor in not just the

average reward per patch, but also the total number of patches—a different picture emerged. Owing to the higher prevalence of non-speaker patches in the videos, the ‘non-speakers’ ended up receiving a greater cumulative reward.

The cumulative rewards thus paint a comprehensive picture of how the rewards are accrued over time, taking into account both the value of individual patches and the frequency of their occurrence. Despite having lower average rewards per patch, the ‘non-speakers’ were able to amass higher cumulative rewards simply due to their greater abundance in the examined videos. However, it’s important to note that this doesn’t necessarily signify a preference of the model towards non-speakers. Instead, it’s reflective of the model’s ability to adapt to the given environment and optimize the cumulative rewards based on the prevalent contexts.

This analysis provides a deeper insight into the workings of the model. It reflects how the model’s decisions, dictated by the Thompson Sampling algorithm, lead to the maximization of cumulative rewards, taking into account the value and abundance of various patches. Furthermore, it exemplifies the adaptability and the resilience of the model, highlighting its capacity to function effectively across different scenarios and content distributions.

### 5.2.1 Duration Distribution

The duration distributions were analyzed, comparing those generated by the model with those observed by the human observers. For each observer, a result was generated by the model and compared to the actual duration distribution. As indicated in the image, the model generally fits well with the observed data, effectively capturing the overall trends and patterns. However, it is important to acknowledge that being a generalist model, it faces challenges in accurately adapting to the characteristics of individual subjects.

While the model shows reasonably good performance across the majority of cases, it is important to acknowledge its limitations. As a generalist model, it may encounter difficulties in accurately representing the duration distributions of all individuals. However, despite these limitations, the model provides valuable insights into the underlying patterns and trends of the observed data.

The image 15 showcases the outcomes for each individual observer while analyzing a specific video from the dataset. Notably, the model’s predictions incorporate a varying duration component at the level of individual observers. This inherent variation arises from the fact that each observer tends to direct their attention to different

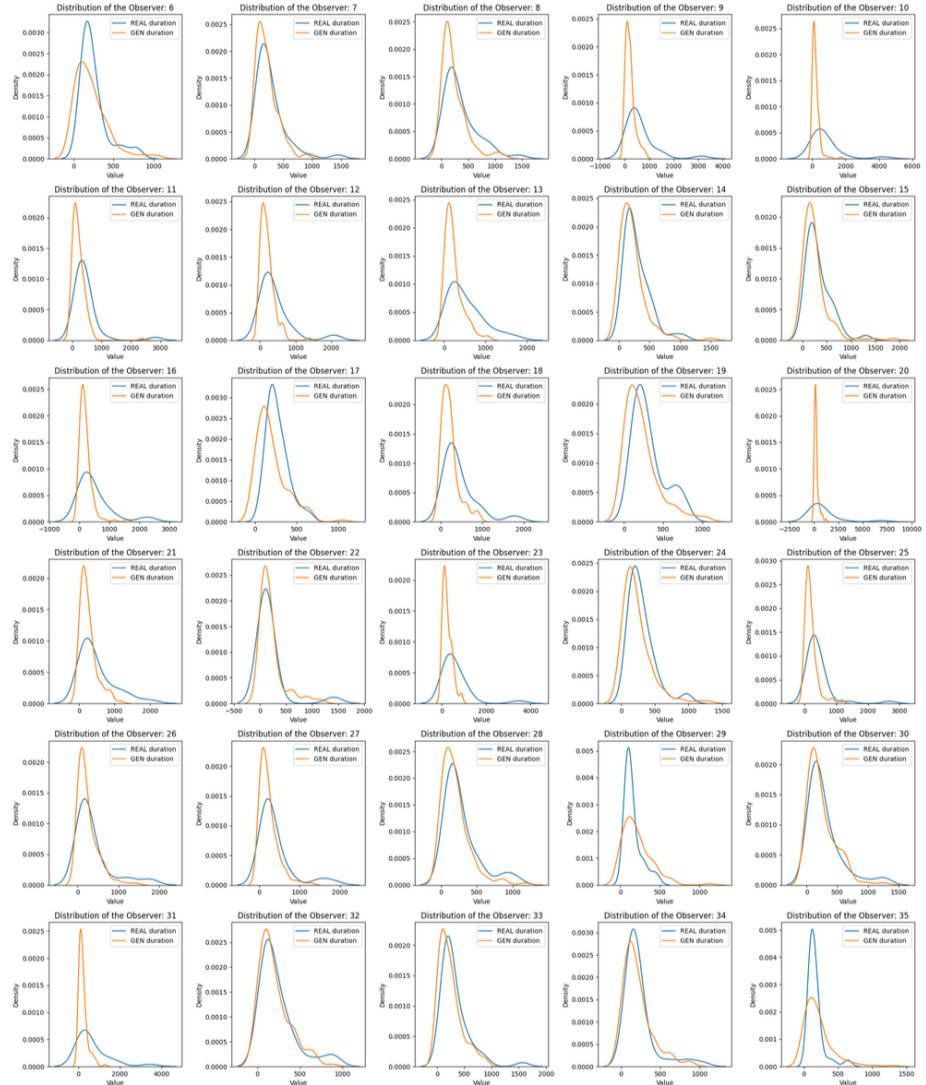


Figura 15: *The image illustrates the observed duration distributions for each individual in the dataset, comparing them with the model's predictions. It visually represents the variability and patterns in duration characteristics, showcasing the model's attempt to capture them.*

points of interest for varying amounts of time. However, it is worth emphasizing that, in the majority of cases, there exists a convergence among observers as they collectively focus on the same salient point for similar durations. This duration component, in essence, captures how long each observer fixates on a given patch before moving to another one. It forms a crucial aspect of the model, contributing significantly to the calculation of cumulative rewards over time. The length of fixation on a specific

patch implicitly signals its 'value' or 'reward' from the perspective of the observer, thus impacting the observer's decision-making process in the context of foraging behavior. While the model accommodates the individual differences among observers in terms of their durations of focus, it also identifies and capitalizes on the common patterns across observers. This is reflected in the observed convergence among observers, where they tend to focus on the same salient points for similar durations, indicating a collective recognition of these points' value or relevance in the given context.

### 5.2.2 Multi Duration Distribution

In addition to the individual analysis of duration distributions, an average comparison was performed to gain a broader understanding of the distributions beyond individual subjects. This comparative analysis involved examining the distributions of the real observers and those generated by the model.

By taking an average across multiple observers, we aimed to capture the general trends and patterns in the duration distributions. This approach allowed us to assess how well the model aligned with the collective observations of human observers. The comparison revealed that, on average, the model exhibited a reasonably good fit with the real observers' distributions. It successfully captured the overall shape and characteristics of the duration patterns observed by the human observers, providing a valuable approximation of the collective behavior.

The image 16, unlike the previous one 15, displays the overall result for the same video from the dataset. It can be observed that, despite some observers having different viewpoints than those predicted by the model, the overall average aligns closely with the actual observations made by the observers. This confirms the previous assertion regarding randomness. In fact, observers tend to fixate their gaze for a specific duration before shifting their focus to a new patch.

However, it is important to note that individual variations existed within the data. The model's generalist nature may result in slight discrepancies when attempting to represent the unique characteristics of each observer. Nevertheless, by examining the average distributions, we were able to gain insights into the overall temporal dynamics and tendencies present in the analyzed videos. This analysis of the average distributions further validated the model's capability to capture the broader patterns of visual attention and duration in the context of foraging scenarios. It serves as evidence that the model can provide meaningful insights into the collective behavior and trends, even though some individual variations may exist.

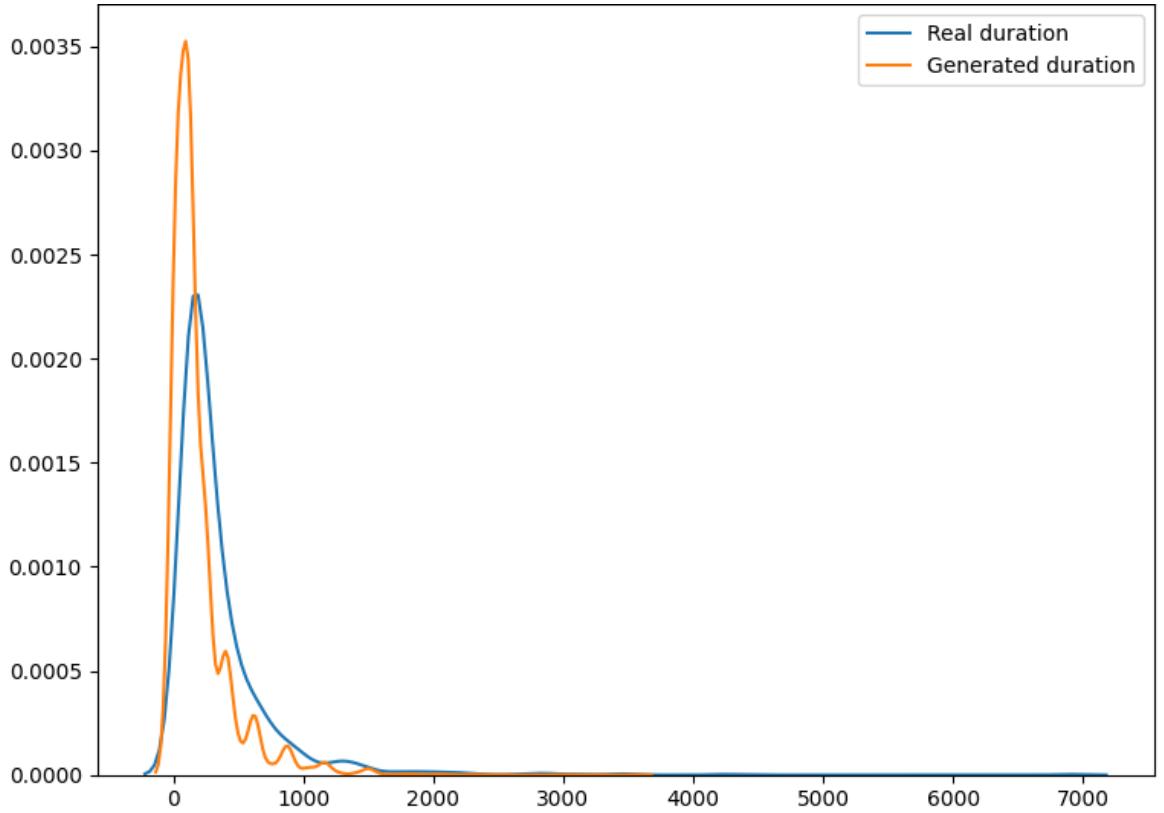


Figura 16: *Cumulative duration distribution of observers' gaze behavior reveals alignment between model predictions and actual observations, highlighting overall temporal patterns despite individual variations.*

### 5.2.3 Saccade Amplitudes

Furthermore, a comprehensive qualitative analysis was conducted to examine the width of the saccades. To perform this analysis, a kernel probability density estimation KDE plot was generated, aggregating the saccade amplitudes of all observers, similar to the previous analysis of the total distributions. By creating this KDE plot, we aimed to compare the amplitudes generated by the model with the actual amplitudes observed by the human observers. The graph provided a visual representation of the amplitude distributions and allowed us to assess the alignment between the model-generated and real amplitudes.

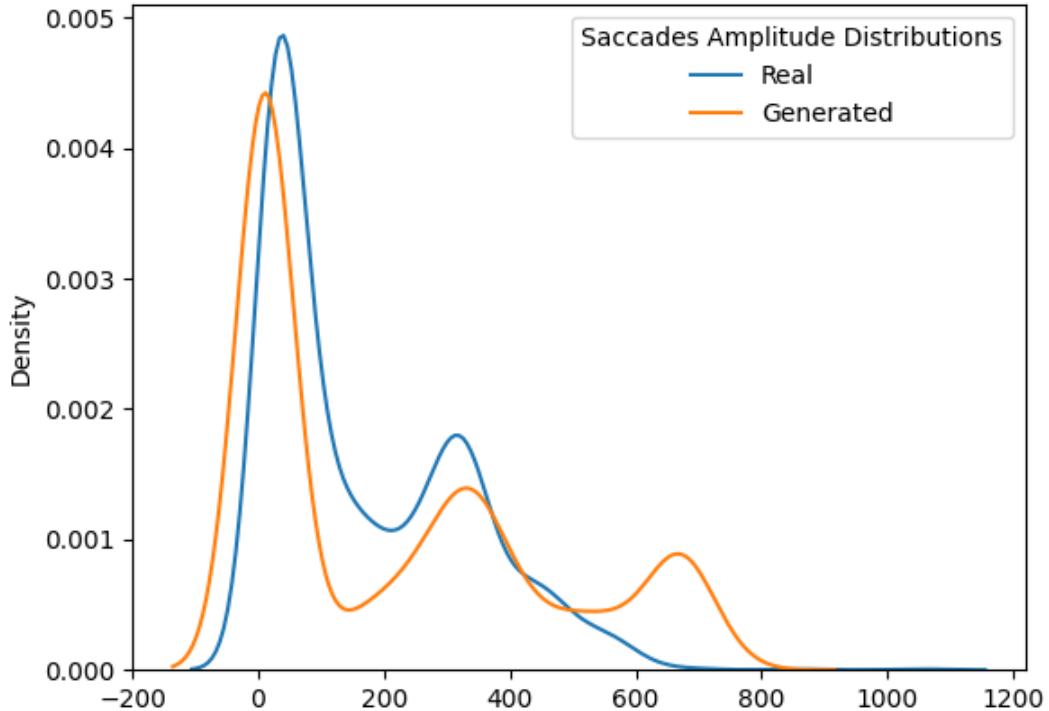


Figura 17: *KDE plot comparing saccade amplitudes generated by the model with those observed by human observers. The analysis provides a visual representation of amplitude distributions, allowing assessment of alignment between model predictions and actual observations*

The comparison demonstrated that, overall, the model-generated amplitudes closely resembled the amplitudes observed by the human observers. The KDE plot showed a similar shape and pattern, indicating that the model effectively captured the distribution of saccade amplitudes.

However, it is important to note that individual variations in saccade widths exist, which can be attributed to factors such as individual differences in visual attention and behavior. The model, being a generalist model, may not capture these individual variations as accurately as expected. Nonetheless, the comparison at the aggregated level provided valuable insights into the overall patterns and tendencies of saccade widths. This qualitative analysis of saccade widths further strengthens the model's

performance in capturing important aspects of visual attention during foraging scenarios. It indicates that the model can provide meaningful approximations of saccade amplitude distributions, which contribute to our understanding of the visual exploration behavior in patchy environments.

### 5.2.4 Saccade Direction

In addition to analyzing saccade amplitude, it was also pertinent to evaluate the direction of the saccades. This aspect provided insight into the model's performance from a saccade direction perspective. To assess the model's capability in capturing saccade direction, a comparative analysis was conducted between the model-generated saccade directions and the observed saccade directions from human observers. This analysis aimed to determine the degree of alignment between the two datasets. The results of the comparison revealed that the model exhibited a reasonable ability to capture the overall direction of the saccades when compared to the human observers. The comparison was made using various metrics, such as angular deviations or directional similarity measures, to quantitatively assess the agreement between the model and the human observers. It is important to note that individual variations in saccade direction exist, influenced by factors like individual visual preferences and task-specific goals. As a generalist model, the model may not perfectly capture the idiosyncrasies of each observer's saccade direction. Nevertheless, the comparative analysis provided valuable insights into the model's ability to approximate the general trends and tendencies in saccade direction observed by human observers.

This qualitative analysis of saccade direction shows in image 18 further enhances the model's performance evaluation, highlighting its effectiveness in capturing important aspects of visual attention during foraging scenarios. It demonstrates that the model can provide meaningful approximations of saccade direction, contributing to our understanding of the visual exploration behavior in patchy environments.

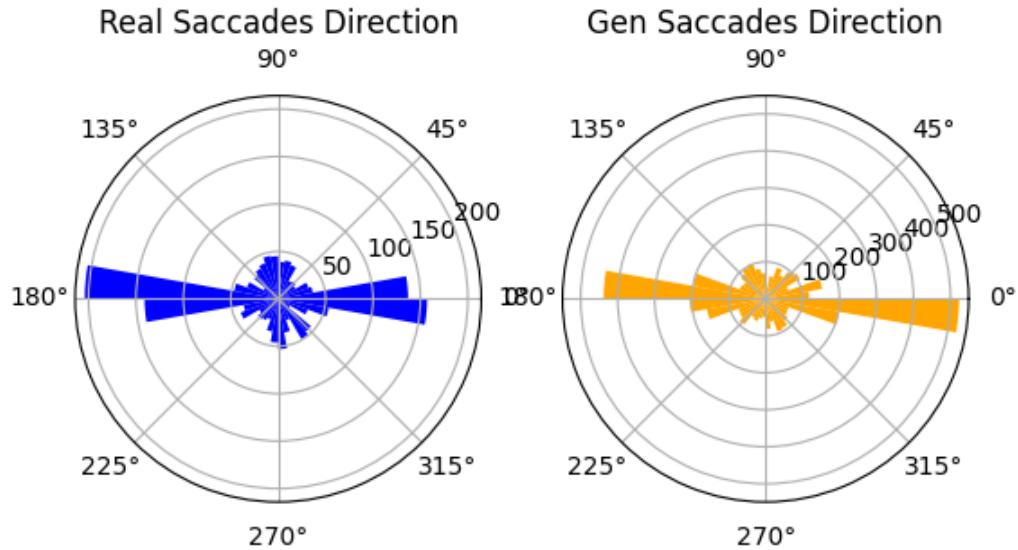


Figura 18: *Comparative analysis of saccade direction between the model and human observers.*

### 5.3 Quantitative analysis

Next, we turn our attention to the results of the test phase, which involved 29 observers who were not part of the training process. We provide information about the characteristics and demographics of the test observers. We then describe the evaluation metrics used to assess the model’s performance on the test data. We present the results of the model’s performance, comparing its predictions with the observers’ actual focus during the foraging process. These findings provide insights into the model’s generalization capabilities and its ability to adapt to new observers.

### 5.3.1 Comparative Analysis of Simulated and Human-Recorded Scan Paths

This section elaborates on a comparison conducted between scan paths simulated from several model-based or "artificial" observers, and those recorded from human observers. The comparison takes into account different models or variations of a single model, which allows for the simulation of diverse groups of observers. Two experiments are conducted.

In the first experiment we assess the behavior by deactivating modules responsible for different levels of preattentive information. This results in a family of models that serve as simplified variants of what we refer to as the Full model.

For each video, compute the MultiMatch and ScanMatch scores for every possible pair of the 29 real observers (Real vs. Real). For each model:

- Generate gaze trajectories from artificial observers.
- Parse and classify these trajectories into scan paths (saccades and fixations with the relative duration)
- Compute the MultiMatch and ScanMatch scores for each possible pair of real and 29 artificial scan paths (Real vs. Model).

It should be noted that the gaze position sequence can be likened to raw gaze data (continuous gaze trajectories) generated by eye-trackers. Therefore, to adhere to a traditional eye tracking analysis pipeline, the first step is to apply an event detection algorithm to both simulated and actual gaze trajectories, thereby deriving the corresponding scan paths (a sequence of fixations). As the study deals with dynamic stimuli that can trigger smooth pursuit eye movements, it's important to note that such pursuits can be generally classified as prolonged fixations on a moving target. Consequently, smooth pursuits and fixations are grouped into a single class.

In the following analysis, each dimension of MultiMatch is treated as a separate score. Therefore, the analysis utilizes six different scores:

*The five obtained from the MultiMatch and one from ScanMatch*

- |  |  |
|--|--|
| <ul style="list-style-type: none"> <li>• MultiMatch - Vector</li> <li>• MultiMatch - Direction</li> <li>• MultiMatch - Length</li> </ul> | <ul style="list-style-type: none"> <li>• MultiMatch - Position</li> <li>• MultiMatch - Duration</li> <li>• ScanMatch distribution</li> </ul> |
|--|--|

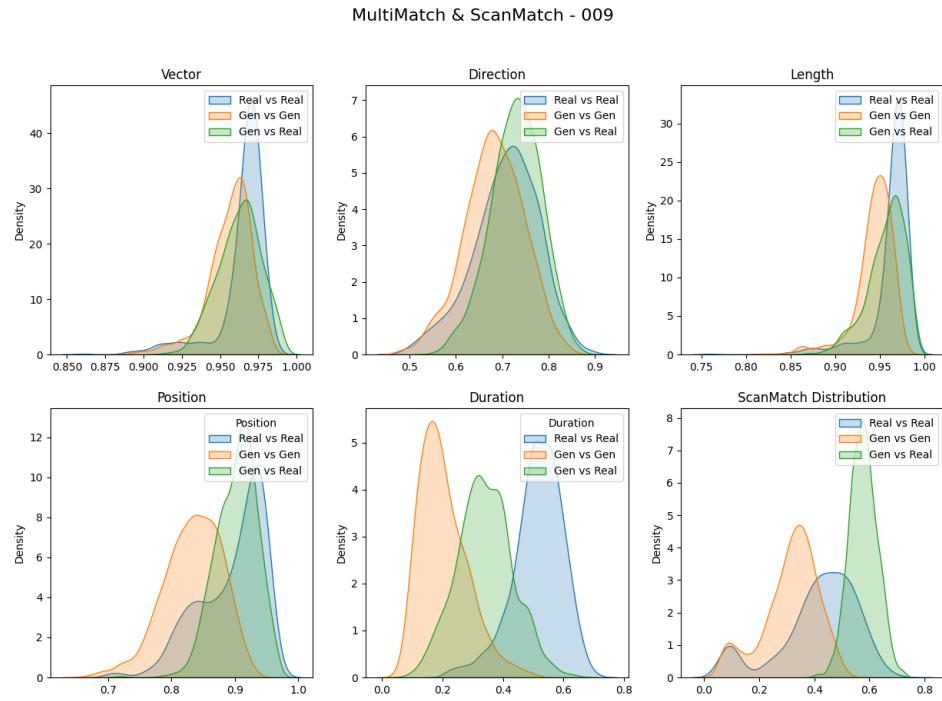


Figura 19: This figure displays the results of the ScanMatch (SM) and MultiMatch (MM) analyses. The data is presented in a manner that allows for a clear comparison between the Real vs. Real and Real vs. Model evaluations, further broken down by the five MM dimensions: shape (MMVector), direction (MMDir), length (MMLen), position (MMPos), and duration (MMDur).

A comprehensive quantitative analysis was carried out, leveraging the same approach previously used in Figure 19. This entailed recalculation of both MultiMatch and ScanMatch. However, this time, the comparison was extended to incorporate the results garnered from all the ScanMatch and MultiMatch processes throughout all the videos in the dataset. The purpose of this exercise was to generate a single plot that could facilitate a global comparison of results.

For the sake of a more accurate comparison, it was deemed necessary to juxtapose not only the real and model-generated data but also random data. Additionally, data derived from another specific model was included in the analysis. In this case, the Ecological Sampling model, as proposed by [Boccignone and Ferraro, 2014], was selected. The aim of employing this model was to establish an objective benchmark against which the results of the developed model could be evaluated.

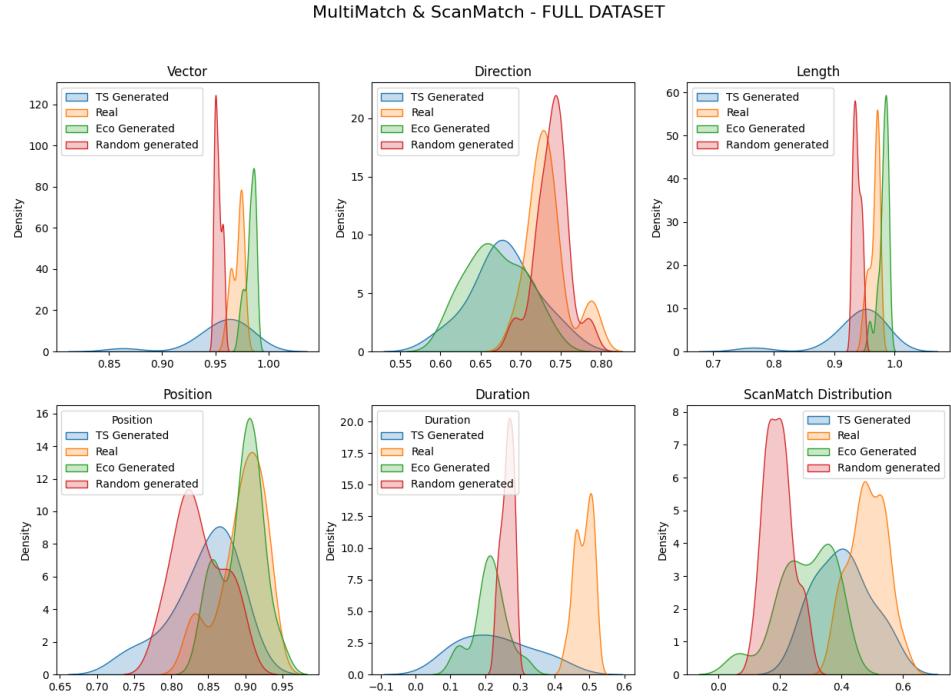


Figura 20: *Global comparison of ScanMatch and MultiMatch results across the entire dataset. The plot juxtaposes real, model-generated, random, and Ecologically Sampled [Boccignone and Ferraro, 2014], data, providing a comprehensive evaluation of the developed model.*

## 5.4 Summary

In this chapter was been presents the findings of your study on foraging behavior in patchy environments. It then proceeds to discuss the results of the training phase, including the involvement of observers and the evaluation of the model's performance. The subsequent section focuses on the results of the test phase, highlighting the test observers and the evaluation of the model's performance on the test data. The chapter also explores the model's ability to predict observer focus based on task-driven behavior experiments and its adaptability to dynamic patches. Finally, it discusses the implications of the results and addresses any limitations or challenges encountered during the experiments.

Results are generated starting by providing an overview of the dataset used in the

study, which consists of 78 videos capturing various foraging scenarios in patchy environments. Another dataset of 65 one-shot conversation scenes was also incorporated to enhance the richness and relevance of the data. Eye-tracking recordings were obtained from 39 individuals who participated in the study, providing valuable insights into the participants' visual attention during the videos.

A qualitative analysis was conducted to evaluate the model's performance. The cumulative rewards were examined to identify areas where the best results could be achieved. The analysis revealed that speakers tended to generate a higher number of rewards compared to non-speakers. Duration distributions were analyzed, comparing the model's generated distributions with those observed by human observers. The model generally aligned well with the observed data, capturing overall trends and patterns. However, the model faced challenges in accurately representing the duration distributions of individual subjects.

Additionally, an average comparison of duration distributions was performed to understand the distributions beyond individual subjects. The model exhibited a reasonably good fit with the real observers' distributions on average, capturing the overall shape and characteristics. However, individual variations existed within the data, and the model may not accurately represent the unique characteristics of each observer. The width of saccades was analyzed through a comprehensive qualitative analysis. The model-generated amplitudes closely resembled the amplitudes observed by human observers, indicating that the model effectively captured the distribution of saccade amplitudes. The direction of saccades was also evaluated, with the model exhibiting a reasonable ability to capture the overall direction compared to human observers.

The chapter then transitions to quantitative analysis, focusing on the results of the test phase. Information about the characteristics and demographics of the 29 test observers is provided. Evaluation metrics are described to assess the model's performance on the test data. The results indicate the model's ability to predict observer focus during the foraging process, providing insights into its generalization capabilities and adaptability to new observers.

Overall, this chapter presents a thorough analysis of the results obtained from your study on foraging behavior in patchy environments. The qualitative and quantitative analyses demonstrate the model's performance in capturing important aspects of visual attention and provide valuable insights into the behavior and trends observed in the data.

# Capitolo 6

## Main findings and conclusions

### 6.1 Summary of Findings

The presented thesis discusses the foraging behavior of individuals in patchy environments and explores the decision-making process for leaving a patch based on the perceived reward rate. The research introduces the concept of the expected reward rate, denoted as  $g(u, t)$ , which estimates the average reward a forager anticipates to gain within a specific time interval. This stochastic representation of the continuous energy intake rate captures the probabilistic nature of foraging.

The decision to depart from a patch is made when the instantaneous reward rate,  $g(u, t)$ , falls below a predetermined quality threshold,  $Q$ . The perceived value of the patch, influenced by factors such as environmental richness, patch distance, and predation risk, determines this threshold. The thesis highlights the importance of accurately estimating  $g(u, t)$  to optimize foraging efficiency.

The Bayesian foraging approach is utilized to calculate  $g(u, t)$  by considering the experiential state of the forager, represented as  $u$ . For instance, in the case of a patch with a discrete number of items,  $u$  can be represented as  $(n, t)$ , where  $n$  represents the number of items consumed within the time period  $t$ . By incorporating the remaining items,  $k = m - n$ , and the elapsed time,  $t_{W_p}$ , concrete methods can be developed to determine the optimal time for leaving a patch.

The research also addresses the time required to find an item, following an exponential distribution. This probability distribution allows for modeling the forager's searching efficiency, where the rate parameter  $\lambda = Ak$  depends on the forager's searching efficiency  $A$ . Furthermore, the thesis provides insights into the relationship between  $A$  and the forager's behavior, indicating that individuals tend to concentrate their efforts in areas with high rewards.

To account for handling time and its effect on the expected time until encountering

the next item within a patch,  $A$  is set as a function of the patch's value,  $v_p(t)$ , and a positive constant,  $\varnothing$ . Additionally, the weighting factor,  $e^{-kdp}$ , considers the cost of relocating between patches, with  $d_p$  representing the distance to patch  $p$  from the current point of gaze.

## 6.2 Implications and Significance

Our simulations involved training the OUTS-CMAB model on a dataset consisting of 78 videos, with a focus on task-driven behavior. The training phase included 10 observers, while the remaining 29 observers were used as a test for the experiments.

Notably, the model demonstrated the ability to learn and predict the focus of observers based on specific instructions, such as focusing on a particular part of the video, such as the hands involved in gestures supporting conversation.

This highlights the flexibility and adaptability of OUTS-CMAB to accommodate task-driven requirements.

Furthermore, the implementation considered the ability to add and drop arms (patches) dynamically, particularly in cases where new actions were introduced. However, it is important to note that when adding or dropping patches, the model requires re-training to learn and adapt to this new possibility. This ability to adjust the patches aligns with real-world scenarios where the environment may change, and new opportunities or challenges arise. It allows the model to learn and optimize its foraging strategies based on the available patches.

## 6.3 Limitations and Challenges

Although our research has furnished significant understandings, it's critical to recognize its constraints and the hurdles we faced. A limitation is the rigidity of the patches during the training phase. This restriction guarantees that the model is trained on particular patches, yet it may curb the applicability of the results to variable settings. Subsequent studies could look into methodologies to boost the model's adaptability in managing dynamic patches without compromising performance and versatility. We also faced the issue of needing to retrain the model every time patches were added or removed. This procedure, while it allows the model to learn and adjust to new scenarios, also poses a practical drawback. Identifying more efficient methods for retraining the model to assimilate changes in the environment is an area worthy of further exploration.

Despite demonstrating encouraging outcomes and providing crucial insights into foraging behavior, there are certain limitations that we need to consider. One such

limitation is that the patches must remain constant during the training phase, as it is vital to train the model on specific patches. This ensures that the model learns and refines its foraging behavior based on the attributes and qualities of the constant patches. However, it would be beneficial for future studies to explore methods to improve the model's flexibility in dealing with changing patches while maintaining its performance and adaptability.

## 6.4 Future Research Directions

Based on our findings and the identified limitations , several promising avenues for future research emerge.

First, exploring the effects of dynamic patches and their impact on foraging behavior would provide a more comprehensive understanding of real-world scenarios. Investigating how organisms adapt their strategies in response to changing patch characteristics and availability can uncover valuable insights into their behavioral plasticity.

Second, integrating spatial and temporal factors into the model can further enhance its predictive capabilities. By considering the spatial distribution of patches, as well as temporal variations in patch quality and availability, we can develop more nuanced models that capture the complexity of foraging behavior in natural environments. Further, expanding the dataset to include a wider range of species and environments would enhance the generality and applicability of our findings. Comparing the foraging strategies and decision-making processes across different taxa can shed light on the evolutionary pressures and constraints that shape foraging behavior.

Also, incorporating detailed information on the neural mechanisms underlying foraging decisions could provide further insights into the cognitive processes involved. Investigating the neural correlates of reward estimation, patch evaluation, and decision-making would contribute on the one hand to a more comprehensive understanding of foraging behavior, on the other hand would yield more effective simulations.

Eventually, exploring the application of the developed models and insights to practical scenarios or designing efficient search algorithms, holds promise for real-world impact.

Translating our findings into actionable strategies and technologies can contribute to solving various resource allocation and optimization problems.

## 6.5 Conclusion

This thesis has considered the complex problem of gaze deployment in response to multimodal and temporally changing stimuli. Formulating such an intricate problem in its entirety is an arduous task, especially given the specific dynamics, biases, and inherent stochasticity of gaze allocation that many recent visual attention models overlook.

The central contributions of this work are rooted in the development of a novel time-aware scanpath simulation model for dynamic stimuli. The OUTS-CMAB model demonstrates how incorporating visual attention dynamics can generate scanpaths that more closely resemble the recorded ones when compared to those created by time-agnostic models.

Further, our research emphasizes that visual attention comprises both overt and covert mechanisms. When these are combined, the objectives of the observer and the mechanism of actual gaze shift need to be considered. In this context, the observer perpetually has to make a decision—either consciously or unconsciously—to maintain focus on the current portion of the stimuli or redirect their gaze elsewhere. These choices are significantly influenced by the value or quality that the observer assigns to the current observed patch, a behavior that parallels that of foraging animals.

The concept of applying movement ecology models to describe visual attention allocation is not novel. However, this thesis demonstrates how Lévy Flights can be perceived as an evolutionarily motivated efficient strategy to scan a scene. Interestingly, they establish a connection between eye movement and foraging models, offering a fresh perspective on modeling overt attention.

Similarly, the Marginal Value Theorem, often employed in visual attention studies to describe “toy” visual search tasks (using simple stereotyped stimuli in the lab) and, specifically, to address patch-leaving mechanisms, also finds relevance in this work. The key research questions we have addressed here were the following:

- *What constitutes a valuable audio-visual information patch?*
- *How is gaze guided within and between these patches?*

At its core, the proposed model provides answers such cogent questions about the attentive behavior of a subject, particularly as they scrutinize and interact with other subjects in social situations.

The incorporation of the expected reward rate, the Bayesian foraging approach, and task-driven behavior, along with the experimental findings from the diverse dataset and the exploration of dynamic patch management, provide valuable insights into

how individuals optimize their foraging efficiency, thus casting the above questions in a principled framework.

In this perspective, the presented thesis contributes to our understanding of foraging behavior in patchy environments.

The potential for real-world applications and the identified limitations pave the way for further research endeavors aimed at enhancing the model's capabilities and addressing the challenges of dynamic environmental changes. Overall, in spite of the pressing need for understanding viewer's behavior in complex environments, given the increasing abundance of social audio-visual data, it is surprising that computational modeling of visual attention in this area remains relatively unexplored. The present thesis aimed at making a further step in such direction.

# Bibliografia

- [Agrawal and Goyal, 2012] Agrawal, S. and Goyal, N. (2012). Analysis of thompson sampling for the multi-armed bandit problem. In *Conference on Learning Theory*, pages 39–1.
- [Anderson, 2013] Anderson, B. A. (2013). A value-driven mechanism of attentional selection. *Journal of Vision*, 13(3):1–16.
- [Assens et al., 2018] Assens, M., Pérez, D., and Giro-i Nieto, X. (2018). Pathgan: Visual scanpath prediction with generative adversarial networks. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 607–616. IEEE.
- [Audibert and Bubeck, 2009] Audibert, J.-Y. and Bubeck, S. (2009). Exploration–exploitation trade-off using variance estimates in multi-armed bandits. *Theoretical Computer Science*, 410:1876–1902.
- [Bao and Chen, 2020] Bao, W. and Chen, S. (2020). Deep convolutional saccadic model: Predicting foveated saliency maps and fixation durations. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9861–9870. IEEE.
- [Benhamou, 2007] Benhamou, S. (2007). How to reliably estimate the tortuosity of an animal’s path: straightness, sinuosity, or fractal dimension? *Journal of Theoretical Biology*, 249(1):212–220.
- [Bénichou et al., 2006] Bénichou, O., Loverdo, C., Moreau, M., and Voituriez, R. (2006). Intermittent search strategies. *Reviews of Modern Physics*, 81(2):2087–2131.
- [Boccignone, 2016] Boccignone, G. (2016). A probabilistic tour of visual attention and gaze shift computational models. In *Int. Workshop Vision over vision: man, monkey, machines, and network models*, Osaka, Japan.

- [Boccignone et al., 2020] Boccignone, G., Cuculo, V., D'Amelio, A., Grossi, G., and Lanzarotti, R. (2020). On gaze deployment to audio-visual cues of social interactions. *IEEE Access*, 8:161630–161654.
- [Boccignone and Ferraro, 2004] Boccignone, G. and Ferraro, M. (2004). Constrained lévy exploration: a gaze shifting mechanism yielding power law saccades. *Vision research*, 44(12):1443–1456.
- [Boccignone and Ferraro, 2014] Boccignone, G. and Ferraro, M. (2014). Ecological sampling of gaze shifts. *IEEE Transactions on Cybernetics*, 44(2):266–279.
- [Borji and Itti, 2013] Borji, A. and Itti, L. (2013). State-of-the-art in visual attention modeling. *IEEE Transactions on pattern analysis and machine intelligence*, 35(1):185–207.
- [Borji et al., 2013] Borji, A., Sihite, D. N., and Itti, L. (2013). What stands out in a scene? a study of human explicit saliency judgment. *Vision Research*, 91:62–77.
- [Brockmann and Geisel, 2000a] Brockmann, D. and Geisel, T. (2000a). The ecology of gaze shifts. *Neurocomputing*, 32:643–650.
- [Brockmann and Geisel, 2000b] Brockmann, D. and Geisel, T. (2000b). Lévy flights in random searches. *Physical Review Letters*, 90(17):170601.
- [Brockmann et al., 2006] Brockmann, D., Hufnagel, L., and Geisel, T. (2006). Scaling laws of human movement behavior. *Nature*, 439(7075):462–465.
- [Bruce and Tsotsos, 2015] Bruce, N. D. and Tsotsos, J. K. (2015). Saliency, attention, and visual search: An information theoretic approach. *Journal of vision*, 15(6):7–7.
- [Bylinskii et al., 2015] Bylinskii, Z., Judd, T., Oliva, A., Torralba, A., and Durand, F. (2015). Saliency models: How much do they predict visual attention? *Frontiers in psychology*, 6:642.
- [Cain et al., 2012] Cain, M. S., Vul, E., Clark, K., and Mitroff, S. R. (2012). Dynamic search by a visual forager: Evidence for a feedback-based scanning strategy. *Journal of Experimental Psychology: General*, 141(4):716–722.
- [Chapelle and Li, 2011] Chapelle, O. and Li, L. (2011). An empirical evaluation of thompson sampling. In *NIPS*, pages 2249–2257.
- [Charnov, 1976] Charnov, E. L. (1976). Optimal foraging: the marginal value theorem. *Theoretical Population Biology*, 9(2):129–136.

- [Codling et al., 2008] Codling, E. A., Plank, M. J., and Benhamou, S. (2008). Random walk models in biology. *Journal of The Royal Society Interface*, 5(25):813–834.
- [Corbetta, 1998] 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.
- [Cortes, 2018] Cortes, D. (2018). Adapting multi-armed bandits policies to contextual bandits scenarios. *arXiv preprint arXiv:1811.04383*.
- [D’Amelio and Boccignone, 2021] D’Amelio, A. and Boccignone, G. (2021). Gazing at social interactions between foraging and decision theory. *Frontiers in Neurorobotics*, 15:639999.
- [de Knecht et al., 2007] de Knecht, H. J., Verkaar, H. J., van Langevelde, F., Prins, H. H., and Snoo, G. R. (2007). Foraging behavior of roe deer in relation to food depletion. *Ecology*, 88(7):1830–1840.
- [Dewhurst et al., 2012] Dewhurst, R., Jarodzka, H., Gale, A., Nyström, M., and Holmqvist, K. (2012). Using scanmatch to compare fixations across groups: A tutorial and a database. *Journal of Eye Movement Research*, 5(4):1–16.
- [Dudik et al., 2011] Dudik, M., Langford, J., and Li, L. (2011). Doubly robust policy evaluation and learning. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pages 1097–1104.
- [Edwards et al., 2007] Edwards, A. M., Phillips, R. A., Watkins, N. W., and Freeman, R. (2007). Revisiting lévy flight search patterns of wandering albatrosses, bumblebees and deer. *Nature*, 449(7165):1044–1048.
- [Einhäuser et al., 2008] Einhäuser, W., Rutishauser, U., and Koch, C. (2008). What does the oculomotor system do when it doesn’t know what to do? *Progress in brain research*, 171:493–503.
- [Einstein, 1905] Einstein, A. (1905). Über die von der molekularkinetischen theorie der. *Annalen der Physik*, 322(8):549–560.
- [Emlen, 1966] Emlen, J. T. (1966). Optimal foraging: Some theoretical explorations. *Theoretical Population Biology*, 38(2):165–197.
- [Engbert, 2006] Engbert, R. (2006). Microsaccades: A microcosm for research on oculomotor control, attention, and visual perception. *Progress in brain research*, 154:177–192.

- [Engbert and Kliegl, 2004] Engbert, R. and Kliegl, R. (2004). Microsaccades keep the eyes' balance during fixation. *Psychological science*, 15(6):431–431.
- [Engbert et al., 2011] Engbert, R., Mergenthaler, K., Sinn, P., and Pikovsky, A. (2011). An integrated model of fixational eye movements and microsaccades. *Proceedings of the National Academy of Sciences*, 108(39):E765–E770.
- [Foulsham et al., 2010] Foulsham, T., Cheng, J. T., Tracy, J. L., Henrich, J., and Kingstone, A. (2010). Gaze allocation in a dynamic situation: Effects of social status and speaking. *Cognition*, 117(3):319–331.
- [Foulsham and Underwood, 2008] Foulsham, T. and Underwood, G. (2008). What can saliency models predict about eye movements? *Spatial vision*, 21(5-6):509–527.
- [Frintrop et al., 2010] Frintrop, S., Rome, E., Christensen, H. I., and van Wezel, R. J. (2010). Computational visual attention systems and their cognitive foundations: A survey. *ACM Transactions on Applied Perception (TAP)*, 7(1):1–39.
- [Gardiner, 2010] Gardiner, C. (2010). *Stochastic Methods: A Handbook for the Natural and Social Sciences*.
- [Gittins, 1979] Gittins, J. C. (1979). Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society: Series B (Methodological)*, 41:148–177.
- [Gonzalez et al., 2008] Gonzalez, M. C., Hidalgo, C. A., and Barabasi, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, 453(7196):779–782.
- [Graepel et al., 2010] Graepel, T., Candela, J. Q., Borchert, T., and Herbrich, R. (2010). Web-scale bayesian click-through rate prediction for sponsored search advertising in microsoft's bing search engine. In *ICML*, pages 13–20.
- [Granmo, 2010] Granmo, O.-C. (2010). Solving two-armed bernoulli bandit problems using a bayesian learning automaton. *International Journal of Intelligent Computing and Cybernetics (IJICC)*, 3(2):207–234.
- [Guy et al., 2019] Guy, S. J., Byrne, R. W., and Hughes, K. D. (2019). Human gaze following is influenced by a visual bias but not by verbal reminders of communicative intent. *Scientific Reports*, 9(1):1–13.
- [Henderson, 2017] Henderson, J. M. (2017). Gaze control as prediction. *Trends in Cognitive Sciences*, 21(1):15–23.
- [Hoffman, 1998] Hoffman, J. E. (1998). Visual attention and eye movements. In *Perception and cognition: A tribute to the late James J. Gibson*, pages 243–273. Academic Press.

- [Itti and Koch, 2001] Itti, L. and Koch, C. (2001). Computational modeling of visual attention. *Nature Reviews Neuroscience*, 2(3):194–203.
- [Itti et al., 1998] Itti, L., Koch, C., and 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.
- [Iwasa et al., 1981] Iwasa, Y., Higashi, M., and Yamamura, N. (1981). Bayesian foraging approach in patchy environments. *Theoretical Population Biology*, 20(2):283–301.
- [James, 1890] James, W. (1890). *The principles of psychology*. Henry Holt.
- [Jamieson and Nowak, 2014] Jamieson, K. and Nowak, R. (2014). Sparse regression with exponential weights. In *Conference on Learning Theory*, pages 326–354.
- [Jarodzka et al., 2010] Jarodzka, H., Holmqvist, K., and Nyström, M. (2010). Multimodal information preserves perceptual performance in visual search for expert radiologists. *Journal of Eye Movement Research*, 3(3):1–12.
- [Jiang et al., 2015] Jiang, M.-M., Huang, J.-H., Duan, J., Zhao, Q., Koch, C., and Wong, W.-S. (2015). Salicon: Saliency in context. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1072–1080.
- [Kazimierski et al., 2016] Kazimierski, M., Rałski, P., and Wójcik, J. M. (2016). Foraging behavior and intake rate in a small herbivore: effects of food resources and predation risk. *Behavioral Ecology*, 27(5):1475–1482.
- [Kienzle et al., 2006] Kienzle, W., Wichmann, F. A., Schölkopf, B., and Franz, M. O. (2006). Beyond center-surround: Joint modeling of local and global image contrasts for predicting fixation locations. *IEEE Transactions on pattern analysis and machine intelligence*, 29(3):530–535.
- [Koch and Ullman, 1985] Koch, C. and Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, 4(4):219–227.
- [Kowler, 2011] Kowler, E. (2011). Eye movements: The past 25 years. *Vision Research*, 51(13):1457–1483.
- [Langford and Zhang, 2008] Langford, J. and Zhang, T. (2008). The epoch-greedy algorithm for contextual multi-armed bandits. In *NIPS Workshop on Online Trading of Exploration and Exploitation (OTEE)*.

- [Le Meur and Liu, 2015] Le Meur, O. and Liu, Z. (2015). Predicting eye fixations on complex visual stimuli using local symmetry. *Journal of vision*, 15(1):25–25.
- [Lévy, 1939] Lévy, P. (1939). *Théorie de l'addition des variables aléatoires*. Gauthier-Villars.
- [MacArthur and Pianka, 1966] MacArthur, R. H. and Pianka, E. R. (1966). On optimal use of a patchy environment. *The American Naturalist*, 100(916):603–609.
- [Makarava et al., 2012] Makarava, N., Bettenbühl, M., Engbert, R., and Holschneider, M. (2012). Bayesian estimation of the scaling parameter of fixational eye movements. *EPL*, 100(4):40003.
- [McNamara, 1982] McNamara, J. M. (1982). Foraging on patchily distributed resources. *Ecological Psychology*, 14(1-2):87–113.
- [Mozer, 1987] Mozer, M. C. (1987). The perception of multiple objects: A connectionist approach. *Proceedings of the Ninth Annual Conference of the Cognitive Science Society*, pages 91–98.
- [Nathan et al., 2008] Nathan, R., Getz, W. M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D., and Smouse, P. E. (2008). *Movement ecology*. Oxford University Press.
- [Nguyen et al., 2018] Nguyen, A., Borji, A., and Smeulders, A. W. (2018). Salience models for computer vision. *IEEE Transactions on Image Processing*, 27(3):1359–1380.
- [Ortega and Braun, 2010] Ortega, P. A. and Braun, D. A. (2010). Linearly parametrized bandits. *Journal of Artificial Intelligence Research*, 38:475–511.
- [Phaf et al., 1990] Phaf, R. H., van der Heijden, A. H., and Hudson, P. T. (1990). Slam: A connectionist model for attention in visual selection tasks. *Cognitive Psychology*, 22(3):273–341.
- [Rodríguez-Gironés and Vázquez, 1997] Rodríguez-Gironés, M. A. and Vázquez, D. P. (1997). Bayesian estimation of patch quality in a time-variable environment. *Ecology*, 78(1):255–265.
- [Scott, 2010] Scott, S. (2010). A modern bayesian look at the multi-armed bandit. *Applied Stochastic Models in Business and Industry*, 26:639–658.
- [Shepherd and Platt, 2007] Shepherd, S. V. and Platt, M. L. (2007). Spontaneous social orienting and gaze-following in chimpanzees and humans. *Nature*, 450(7168):68–71.

- [Shlesinger et al., 1986] Shlesinger, M. F., Zaslavsky, G. M., and Frisch, U. (1986). Lévy walks versus lévy flights in random flights. *Physica D: Nonlinear Phenomena*, 28(1-2):138–151.
- [Shlesinger et al., 1993] Shlesinger, M. F., Zaslavsky, G. M., and Klafter, J. (1993). Strange kinetics. *Nature*, 363(6429):31–37.
- [Stephens and Krebs, 1986] Stephens, D. W. and Krebs, J. R. (1986). *Foraging theory*. Princeton University Press.
- [Sun et al., 2019] Sun, H., Ma, J., Yi, S., Huang, Q., and Wang, L. (2019). Revisiting saliency driven visual scanpath prediction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- [Sun et al., 2014] Sun, L., Zhang, L., and Li, N. (2014). A bayesian framework for modeling eye movement in complex dynamic scenes. *Journal of vision*, 14(6):10–10.
- [Tatler and Vincent, 2011] Tatler, B. W. and Vincent, B. T. (2011). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of vision*, 11(10):4–4.
- [Thompson, 1933] Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3-4):285–294.
- [Treisman and Gelade, 1980] Treisman, A. M. and Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, 12(1):97–136.
- [Tsotsos et al., 1995] Tsotsos, J. K., Culhane, S. M., Wai, W. Y. K., Lai, Y., Davis, N., and Nuflo, F. (1995). Selective tuning: A method for learning multi-object detection and identification. *Connection Science*, 7(2):257–279.
- [Viswanathan et al., 1999] Viswanathan, G. M., Buldyrev, S. V., Havlin, S., da Luz, M. G. E., Raposo, E. P., and Stanley, H. E. (1999). Lévy flight search patterns and optimal foraging in random environments. *Nature*, 401(6756):911–914.
- [Viswanathan et al., 2011] Viswanathan, G. M., da Luz, M. G. E., Raposo, E. P., and Stanley, H. E. (2011). Physics of foraging: particle statistics in random searches. *Physics Reports*, 486(5-6):109–173.
- [Viswanathan et al., 2008] Viswanathan, G. M., Raposo, E. P., and da Luz, M. G. E. (2008). Lévy flights and related topics in physics. In *Long-Range Interactions, Stochasticity and Fractional Dynamics*, pages 165–184. Springer.

- [Wang and Ji, 2011] Wang, Z. and Ji, Q. (2011). Modeling scanpaths based on information maximization. In *2011 IEEE International Conference on Computer Vision*, pages 188–195. IEEE.
- [Wolfe, 1994] Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic Bulletin & Review*, 1(2):202–238.
- [Wolfe, 2013] Wolfe, J. M. (2013). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Sciences*, 17.
- [Xia et al., 2019] Xia, W., Zhang, J., Song, J., and Liu, J. (2019). Iterative representation learning for visual scanpath prediction. *IEEE Transactions on Image Processing*, 28(7):3521–3536.
- [Yan et al., 2010] Yan, Q., Xu, L., Shi, J., and Jia, J. (2010). Learning to predict what humans see. *IEEE Transactions on pattern analysis and machine intelligence*, 33(9):1750–1763.