Département de Psychologie
Université de Fribourg (Suisse)

Using eye movements to isolate information use for faces and ecologically valid information sampling in the wild

Thèse de Doctorat
(cumulative)

Présentée à la Faculté des lettres et des sciences humaines de l'Université de
Fribourg (Suisse)

en vue de l'obtention du titre académique de
Docteur en Psychologie

par

Michael Papinutto
(De Sumiswald, Suisse)

Approuvée par la Faculté des lettres et des sciences humaines sur proposition
des professeurs Roberto Caldara (premier rapporteur), Denis Lalanne (deuxième
rapporteur) et Filipe Cristino (troisième rapporteur)

Fribourg, le 3 décembre 2021
Le Doyen, Prof. Dominik Schöbi

This thesis was directed by Prof. Dr. Roberto CALDARA and co-directed by Prof. Dr. Denis LALANNE.

The University of Fribourg and the Human-IST institute supported this thesis.

Thesis examination committee:

Prof. Dr. Roberto **CALDARA**, University of Fribourg, thesis director, examiner

Prof. Dr. Denis **LALANNE**, University of Fribourg, thesis co-director, examiner

Prof. Dr. Filipe **CRISTINO**, Nottingham Trent University, examiner

Prof. Dr. Chantal **MARTIN-SOLCH**, University of Fribourg, Chair of the examination committee

Prof. Dr. Nicolas **RUFFIEUX**, University of Fribourg, assessor

Fribourg, 8th september 2021

Abstract

The eyes sample the visual world at a high speed. The purpose of such fast sampling is to counteract the low-resolution information extracted by the extremities of the retina. As such, eye movements constantly position the targets at the retina centre with high resolution. Determining where the eyes land gives access to the information sampled and perhaps to the information processed by the brain, opening doors to cognition. However, it has been shown that the position of the fixation does not necessarily reflect information use, as the eyes may extract information from the periphery. To overcome this problem, gaze-contingent paradigms allow one to precisely control the extent of information available around the fixation. In addition to this limitation of eye movement research, most eye-tracking studies are performed in confined laboratory settings that do not guarantee the generalisability of their findings to the real world. This thesis aims first to evaluate information use in facial recognition utilising gaze-contingent paradigms. Second, it aims to ensure the ecological validity of eye movement studies in the laboratory.

Despite the assessment, almost 50 years ago, of the amount of information that can be read in a single fixation and the growing number of studies evaluating facial information sampling, the perceptual span for faces—the facespan—has not been determined yet. The first contribution of this thesis is therefore an evaluation of the facespan with a gaze-contingent moving window paradigm. Several window sizes showing the target facial information were used in a face recognition task. Our results indicate that an aperture of at least 17° of visual angle allows for normal performance and oculomotor behaviour. Within this aperture, 7° were preserved from the Gaussian aperture, hence, evaluating the facespan at 7°. However, as in reading, this quantity might be modulated by different parameters such as expertise, disorders, or age.

For this reason, the second contribution is an evaluation of the modulation of the quantity of facial information extracted in the case of non-hearing individuals. We presented faces upside down or upright to a group of hearing-impaired persons and a control group in a face identification task. This time, a window extending according to the duration of the fixation was used. Based on our results, the hearing-impaired people were less affected by the inversion effect than the control group. In addition, the hearing impaired extracted more information from a face at each fixation than the control group.

Those studies performed in a confined laboratory raise the more general question of ecological validity for which previous studies did not find a consensus. Thus, the last contribution is an answer to this question. To this end, we compared the eye movements of participants in the real world and in the laboratory. Conditions were kept as similar as possible between the two environments. Our results revealed that, unlike in the laboratory, participants in the real world were regularly watching their next steps. Regarding similarities, participants in both groups looked at actionable objects, signs and faces in the same way.

In conclusion, these contributions clarify the minimal quantity of facial information needed at each fixation and its modulation by deafness. In addition, the results of studies on eye movements obtained in the laboratory seem to be valid, but only under specific conditions.

Résumé

C'est à grande vitesse que les yeux échantillonnent le monde visuel. Le but d'un tel échantillonnage est de contrebalancer l'extraction d'informations à faible résolution aux extrémités de la rétine. Ainsi, les mouvements oculaires positionnent constamment les objets au centre de la rétine afin de bénéficier d'une haute résolution. C'est pourquoi, déterminer où les yeux se posent donne accès à l'information échantillonnée et probablement à l'information traitée par le cerveau, ouvrant les portes de la cognition. Cependant, la position d'une fixation ne reflète pas nécessairement l'information utilisée, pouvant être extraite dans la périphérie. Pour surmonter ce problème, les paradigmes de contingence au regard permettent de contrôler l'étendue de l'information disponible autour de la fixation. En plus de cette limitation des études en mouvements oculaires, la plupart des études sont réalisées dans des environnements de laboratoire ne garantissant pas la généralisation des résultats au monde réel. Dès lors, cette thèse de Doctorat vise premièrement à évaluer l'utilisation de l'information dans la reconnaissance des visages avec le paradigme contingent au regard et aussi, à s'assurer de la validité écologique des études de mouvements oculaires en laboratoire.

Malgré l'évaluation, il y a plus de 50 ans, de la quantité d'information pouvant être lue en une fixation et le nombre croissant d'études évaluant l'échantillonnage de l'information faciale, l'empan perceptif des visages—le facespan—n'a pas été déterminé. Ainsi, la première contribution vise à évaluer le facespan avec un paradigme de fenêtre mobile contingente au regard. Plusieurs tailles de fenêtres montrant l'information faciale ont été utilisées dans une tâche de reconnaissance de visages. Nos résultats ont montré qu'une ouverture d'au moins 17° permettait une performance et un comportement oculaire dans la norme. Dans cette ouverture, 7° étaient préservés de l'ouverture gaussienne, représentant le facespan. Cependant, comme en lecture, cette quantité peut être modulée par différents paramètres.

Pour cette raison, la deuxième contribution a pour but d'évaluer la modulation de la quantité d'informations faciales extraites dans le cas de la surdité. En effet, nous avons présenté des visages à l'envers et à l'endroit à un groupe de personnes malentendantes et à un groupe contrôle dans une tâche d'identification de visages. Cette fois, une fenêtre grandissante avec la durée des fixations a été utilisée. Nos résultats ont montré que les personnes malentendantes étaient moins affectées par l'effet d'inversion que le groupe contrôle. De plus, elles ont extrait plus d'informations du visage à chaque fixation que le groupe de contrôle.

Ces études réalisées dans un laboratoire soulèvent la question plus générale de la validité écologique sur laquelle les études précédentes n'ont pas trouvé de consensus. Ainsi, la dernière contribution tente de clarifier cette question. Dans ce but, nous avons comparé les mouvements oculaires des participants dans le monde réel et en laboratoire. Les conditions ont été maintenues aussi proches que possible entre les deux environnements. Nos résultats ont révélé que, contrairement au laboratoire, les participants dans le monde réel regardaient régulièrement leurs pas. Concernant les similitudes, les participants des deux groupes regardaient de la même manière les objets, les panneaux et les visages susceptible à l'interaction.

En conclusion, ces contributions clarifient la quantité minimale d'informations faciales nécessaire à chaque fixation et sa modulation par la surdité. De plus, les résultats d'études en mouvements oculaires obtenus en laboratoire semblent valables, mais dans des conditions spécifiques.

THIS THESIS IS DEDICATED TO MY GRANDFATHER WHO ALWAYS SHARED HIS CURIOSITY AND INTEREST ABOUT ANY LITTLE THINGS IN LIFE

Acknowledgements

When I had the chance to enter the world of scientific research, I was surrounded by people who amazed me with their wealth of knowledge, despite giving the impression that everything is easy. I remember having expressed the feeling of being a baby surrounded by grown-ups. To this feeling, I was answered that everyone was a baby once, but one day I would probably end up being the grown-up in scientific research. However, at the end of this doctoral journey, I do not feel like a grown-up yet. Indeed, each field I explored (neurosciences, medicine, computer sciences, clinical and cognitive psychology) opened a wealth of knowledge and new concepts I was eager to understand, learn, play with and use. Fortunately, I was surrounded by amazing people sharing their knowledge, helping me, which made me enjoy this doctoral journey even more. I would like to thank them for all they have brought me with the following few lines.

First and foremost, I would like to thank my supervisor, Prof. Roberto Caldara, and my co-supervisor, Prof. Denis Lalanne, for their guidance, help, and advice through each step of my thesis. I would also like to thank them for trusting and supporting me when I asked for this crazy idea to conduct a B.Sc. in computer science aside from my thesis. I would like to address special thanks to Sébastien Miellet and Julien Nembrini for the insightful and non-lasting discussions we had together on personal life, science, and many other topics before, during, and after completing my thesis. Those discussions helped me empower my skills and critical thinking and become a better version of myself throughout the years.

I would like to thank the thesis examination committee for their time, insightful comments and questions during the defence: Prof. Chantal Martin-Sölch (Chair), Prof. Roberto Caldara (Examiner), Prof. Denis Lalanne (Examiner), Prof. Filipe Cristino (External examiner) and Prof. Nicolas Ruffieux (Assessor).

As mentioned above, I had the chance to explore many fields and especially be part of three different teams. First, I would like to thank present and past members of the visual and social neurosciences teams: Yingdi, Xiniy, Valentina, Helen, Nayla, Lisa, Meike, Anne-Raphaëlle, Justin; with special thanks to my office-mates: Camille, Peter, Géraldine, Nicolas and Charles. Then, I would like to thank the Human-IST institute present and past members, Prof. Edy Portmann, Himanshu, Hamed, Sailin, Moreno, Jhonny, Manuel and all the other members, with special thanks to my office-mates: Florian, Simon, and Agnès and to the other PhD students with who we spent hours chatting about technology and many other topics during our Monday lunch and coffee breaks: Pierre, Raphaël, Pascal and Arnaud. Finally, I would like to thank the members of the clinical and health psychology team as they provided me with a step back on my computational skills and allowed me to fight with the impostor syndrome: Prof. Martin-Sölch,

Mathias, Maya, Tanya, Saba, Dahlila, Amélie, Romina, Laurent and the other members that I might have forgotten.

This page intentionally left blank.

Table of Contents

List of Figures

List of Tables

List of Acronyms

2D Two-Dimension(al).

3D Three-Dimension(al).

AOI Area Of Interest.

AR Augmented Reality.

ASD Autism Spectrum Disorders.

EA Eastern Asian.

EEG ElectroEncephaloGram.

EFRP Eye Fixation Related Potential.

EMHMM Eye Movements analysis with HMM.

ERP Event Related Potential.

ETG Eye-Tracking Glasses.

FIE Face Inversion Effect.

fMRI functional Magnetic Resonance Imagery.

FOR Frame of Reference.

HMM Hidden Markov Model.

NSS Normalised Scanpath Saliency.

ORE Other Race Effect.

POF Point of Fixation.

POR Point of Regard.

ROI Region Of Interest.

SR Super-Recogniser.

VF Visual Field.

VR Virtual Reality.

WC Western Caucasian.

WEIRD Western, Educated, Industrialized, Rich, And Democratic.

This page intentionally left blank.

We keep moving forward, opening new doors and doing new things because we are curious, and curiosity keeps leading us down new paths.

Walt Disney

1

General Introduction

VISION is the most developed sense in human beings, with almost half of the cerebral cortex dedicated to it. Human beings mostly rely on visual information to perform everyday life activities, contrary to most animals which rely mainly on other senses. The brain receives visual inputs from the eyes that sample visual information by moving continuously at high speed. These movements are the most rapid and precise ones performed by the human body, thanks to six ocular muscles per eye. These muscles allow the eye to yaw, pitch, and roll about vertical and horizontal axes seamlessly and effortlessly. Humans perform no less than 10,000 eye movements each day at a speed up to 900° per second (approximately 280 km per hour) to sample their visual world (Bahill, Adler, et al., 1975; Bahill, Clark, et al., 1975).

Eye movements are a crucial feature to extract visual information from the world. Their recordings have allowed for the creation of a robust and objective index to identify the what, when, and where of visual information extraction. Eye tracking has been employed in many studies and has proven to be a valuable cue for identifying information samples by the eyes during various activities. While information sampling delivers substantial cues regarding the information processed by the brain, it remains an inaccurate estimation of information use due to attentional decoupling between the visual attention at the gaze location and the visual attention paid to the rest of the visual field (VF). Precisely identifying information used by the brain requires constraining the available visual information around the gaze location. Thanks to the advent of technology, stimuli presented to participants can be modified online according to gaze locations and movements with a technique named the gaze-contingent paradigm. Moreover, this technique has led to promising findings for a better understanding of the representations built by the brain from visual inputs.

While eye-tracking studies have provided many fruitful results identifying the information sampled by the eyes and the information used in a broad range of topics and with diverse tasks, several questions remain open or at least have not received conclusive explanations. For instance, the quantitative evaluation of information use and its modulation have been addressed in reading and visual search in scenes, but it has not been addressed in the field of face processing. Moreover, most studies bringing knowledge to the scientific community have been performed in a confined laboratory using impoverished representations of the real world. The extensive amount of research conducted under such conditions raises the broader question of the ecological validity of eye movement studies and the possibility to generalise results obtained in those studies to the real world. Despite the assessment of this question across several topics, it is not yet clear whether eye movements recorded in the laboratory match eye movements performed in the wild. Furthermore, clarification is still required with regard to whether the interplay of top-down (related to previous knowledge) and bottom-up (related to the information presented) processes are involved to the same extent in both laboratory and real-world settings. Answering these questions will allow for 1) an identification of the quantity of facial information used, 2) inferences on the possible constructions of representation, 3) an evaluation of how information intake and processing are impaired in other populations. Crucially, ecological validity would formally validate or partially invalidate the generalisability of results

produced in the laboratory to the real world.

The studies presented in this thesis focused on eye movements during face and scene processing. The subsequent sections hence mainly describe the eye-tracking studies related to these two fields. The first study explores ocular behaviour when recognising faces and aims to evaluate the minimal quantity of facial information required at each fixation to reach a typical performance. The second study investigates how information use is modulated in non-hearing people. These two studies raise questions about information use and its modulation in face processing. As the two studies were conducted in constrained laboratory settings, they raise the question of the ecological validity of our results. The last study in this work consequently investigates eye movements during scene processing while navigating; it aims at studying the differences and similarities in information sampling between the laboratory and the real-world settings. This last study, raising the broader question of the ecological validity of laboratory studies in eye movement research, has theoretical and methodological implications for eye-tracking research.

This thesis comprises six chapters. The first chapter presents a brief overview of the anatomy and physiology of the eye, the history of eye-tracking technologies, and a literature review of the research fields of interest for the reader to understand and elaborate on the studies presented in this thesis. These studies are briefly described in Chapter 2 and are included in Chapters 3, 4 and 5. The last chapter discusses the key findings of this work as well as the potential flaws and future research to elaborate on the present contributions.

1.1 From the Physiology of the Eye to the Eye in Movement

Before producing a conscious representation of the world, light reflected or emitted by objects in our surroundings passes through a complex diffraction system to be focused upside down onto a small area at the back of the eye. This refraction system, comparable to a camera, adapts to 1) the depth thanks to tissue that acts similarly to a camera lens and 2) the quantity of light in the environment thanks to a system similar to a camera diaphragm. The whole system guaran-

tees that the right amount of light hits the back of the eye and with the correct focus. A sophisticated computational system composed of several neuronal layers then transduces light energy into a neuronal signal that the brain will further process to provide a conscious visual representation of the world. We constantly move our eyes to change this representation and gather an updated and complete picture of our surroundings. Finally, thanks to the brain's computational power, our vision is upright, appears to be constantly sharp, and feels uninterrupted, despite eye movements.

In this section, an overview of the anatomy of the eye is first presented. This is followed by a description of the physiology of the eye allowing for the transformation of light into a neuronal signal interpretable by the brain. Finally, the muscles moving the eyes and the resulting eye movements are described.

1.1.1 The Anatomy of the Eye

The eye encompasses several types of specialised tissues with specific functions. More specifically, the eye is composed of three major layers: the outer layer, which is a fibrous layer that protects the eye; the middle layer which is a vascularised layer providing blood to the parts of the eye that need it; and the inner layer which is mainly composed of the *retina* encompassing the photoreceptors and allowing for light to be transduced into a neuronal signal. Before reaching this last layer, light travels through several tissues to be bent, focused and then transduced. The first entry point of the light into the eye is a translucent and curved tissue named the *cornea*, which orients and focuses the light into a beam. The cornea is one of the most essential diffraction systems of the eye. The light beam then passes through the *pupil*, an aperture that varies in size depending on the amount of light in the environment—similarly to a camera, the brighter the light, the smaller the aperture will be. The regulation of the amount of light entering the eye and thus the constriction or dilatation of the pupil are performed by two antagonistic muscle groups in the coloured part of the eye, the *iris*. A diffraction lens, named the *crystal lens*, then refracts the light to ensure an optimal focus onto the retina. The crystal lens is attached to the *ciliary muscles*, which allow it to change its curvature and vergence. Finally, the light continues

on its pathway to the back of the eye through the *vitreous humour*, a gel-like substance, to meet the retina (see Figure 1.1.1 for a depiction of a horizontal section of the eye and Felten et al., 2015, for further anatomical details). In addition, six muscles are tied to the eye bulb to move it to align the retina with the target of interest. Those muscles are described in more detail in Section 1.1.3.

Figure 1.1.1

Horizontal section of a right eye
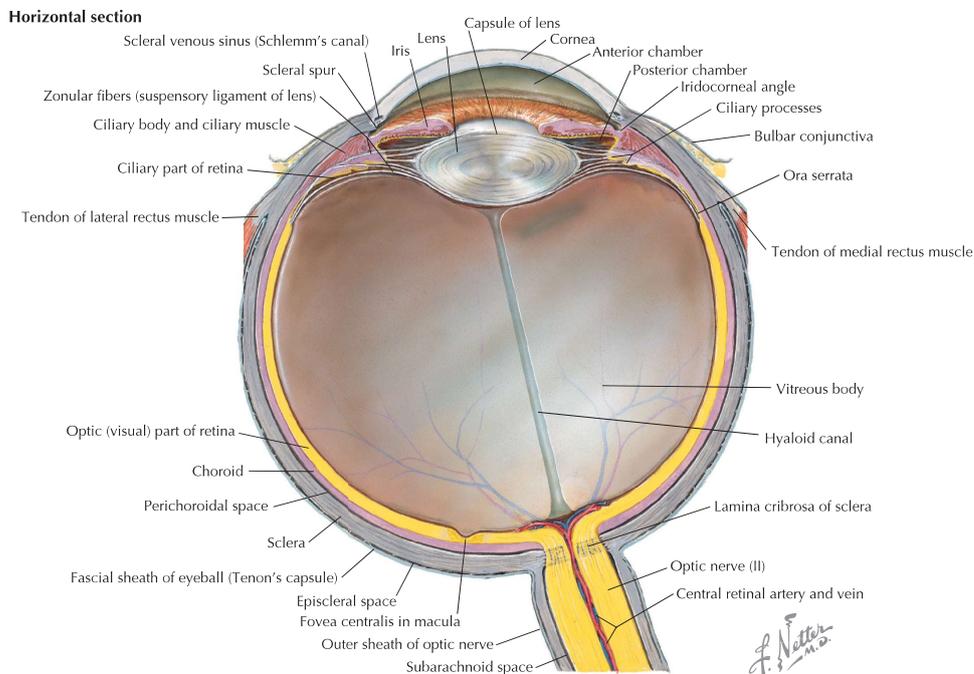
Note. Depiction of a horizontal section of the eye with the three major layers composing the eye. The outer layer encompasses the cornea and the sclera. The middle layer is a vascularised layer composed of the choroid, the ciliary body, and the iris. The inner layer consists of a part of the iris, the retina, and a part of the ciliary body. The light crosses the cornea, the pupil, and the lens, which focuses the light into a beam that continues across the humorous body to hit the retina. Reprinted from Felten et al. (2015).

1.1.2 FROM PHYSIOLOGY TO VISION

The retina is comprised of a sophisticated layering of cells to transduce the light's photonic energy into a neuronal signal that the brain will further process. The resulting VF is strongly related to the retina structure and has the same strengths and weaknesses. Nevertheless, the brain manages to correct for some issues attributed to the physiology of the retina.

Counterintuitively, light must cross all the retina's neuronal layers to hit the photoreceptors. More specifically, light travels across the nerve fibre, and the ganglion cells layer passes through the layers consisting of *bipolar cells* and other cells, as well as the body of the photoreceptors (inner plexiform, inner nuclear, outer plexiform, and outer nuclear layers), before hitting the photoreceptor layer that includes the photoreceptors embedded in the *pigment epithelium* (see Figure 1.1.2). The photoreceptors can be divided into two groups: *rods* and *cones*. The rods provide night—*scotopic*—vision and delivers a low-resolution, uncoloured vision. In contrast, the cones provide day—*photopic*—vision, yielding a colourful and vivid vision at a high resolution. In addition, the photoreceptors are distributed in a specific way in the retina. Indeed, the cones are mainly concentrated in an area called the *fovea* that is devoid of rods. More specifically, the number of cells in each group varies with the retina's eccentricity: the number of cones decreases, whereas the number of rods increases when moving away from the centre of the retina. Both groups of cells send chemico-electrical signals according to the qualitative and quantitative nature of the wavelength of light that hits them. For example, the rods will be more active in the shadow, whereas the cones are sub-divided into three sub-groups reacting to a specific wavelength range. At the molecular level, the light hits a molecule called the *retinal*, which changes its configuration. This configuration change activates the *opsin*—the larger structure encompassing the retinal—causing the photoreceptor to transmit the signal to the adjacent neuronal cells. Crucially, the differentiation in light reactions is made possible thanks to opsin (called *rhodopsin* in rods) that adjusts light absorption to a specific spectrum. Once processed, the transduced signal travels through the optic nerve to the brain (see Figure 1.1.2). The neuronal fibre bundle location in the retina is called the *blind spot* and is utterly devoid of photoreceptors. The transduction of light into the fovea, affluent in cones, occurs at an especially high resolution. Additionally, the neuronal signal of the fovea is exceptionally well represented in the brain, resulting in a high-quality signal. This phenomenon is known as *cortical magnifica-*

tion.

Figure 1.1.2

Drawing of a section through the retina
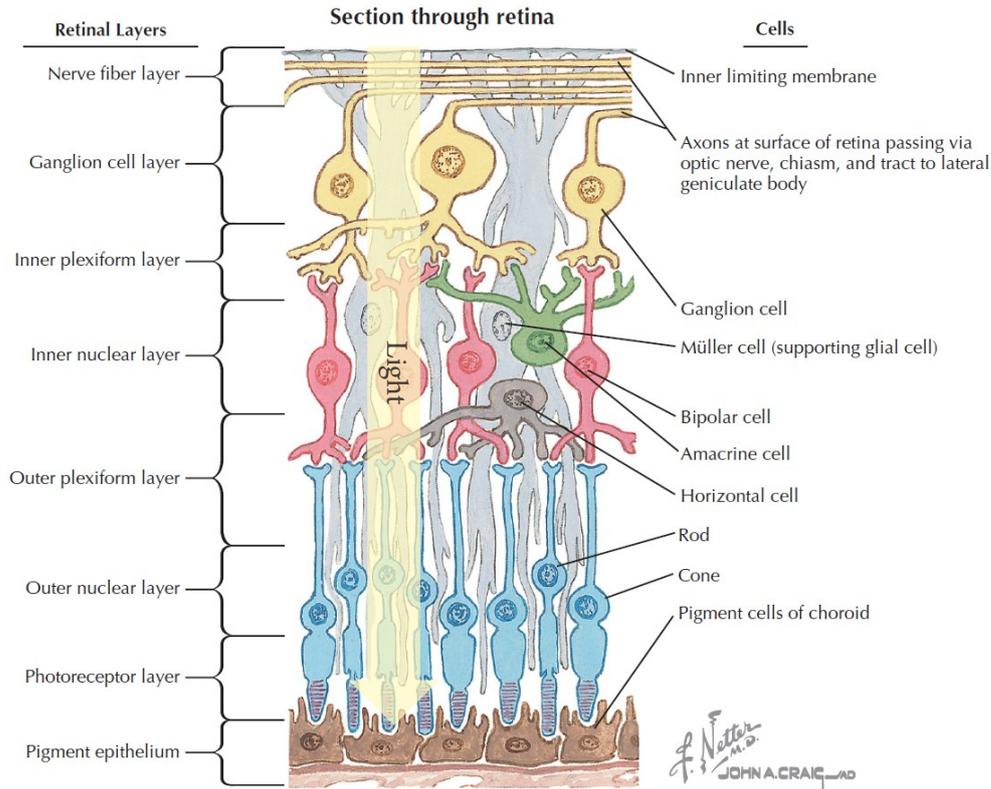
Note. Drawing of the cellular architecture of the retina. Retinal layers are decomposed into layers comprised of neuron bodies and layers made up of fibres as well as inter-layer connections. Each cell type is depicted in a different colour (*e.g.,* photoreceptors are depicted in blue). The transparent yellow arrow denotes the pathway of light through the retinal layers. Adapted from Felten et al. (2015).

Following the retina's architecture and the brain's processing, the resulting VF exhibits colourful vision at high visual acuity and temporal resolution at the centre. The central region, called the *foveal region,* is further processed by a vast number of neurons in the brain's visual areas. This high-quality region spans roughly a 2° visual angle and is also named the central

vision region. The vision outside the central region is called the *peripheral region* and results in a lower quality vision. Peripheral vision is divided into the *parafoveal region*—surrounding the foveal region—and the *peripheral region,* eliciting a lower resolution and less colourful vision that decreases with eccentricity. Crucially, the blind spot, entirely devoid of photoreceptors, does not provide any visual input to the brain. This region is filled in by the brain such that no lack of information is perceived. The visual perception is rendered in mere fractions of seconds and without any conscious effort (see Figure 1.1.3 below for a depiction of this architecture from the histology of an eye to the VF and the resulting human vision and, *e.g.,* Bear et al., 2007; Purves et al., 2011, for a description of the visual system).

1.1.3 The Eye in Movement

As stated above, the anatomy of the eye leads to maximal acuity at the centre of the VF that decays linearly with eccentricity. Clear vision is thus achieved when the visual input falls within the foveal region of the VF. To counteract such constraints, the efficiency of the human visual system resides in the movements of the eyes, allowing the high acuity centre of the VF to systematically lend to targets of interest. To achieve fast and accurate eye movements, six of the most rapid muscles of the human body make the eyeball pitch, yaw, and roll about both the vertical and horizontal axes.

The eye movements are produced with three pairs of antagonistic muscles, including four straight muscles and two obliques. The straight muscles are attached to the top, the bottom, and the sides of the eyeball. The two oblique muscles are connected to the eyeball's upper medial and lower lateral sides. The superior oblique muscle is unique because a trochlea (a structure similar to a pulley) bends it. The medial and lateral straight muscles are exclusively responsible for horizontal movements, while vertical movements result from the co-ordination of the two straight muscles (the inferior and superior straight muscles) and oblique muscles, depending on the horizontal position of the eye. When looking straight ahead, oblique muscles contribute to vertical movements; however, when looking away from the nose, vertical movements result solely from the action of the inferior and superior straight muscles. When looking

Figure 1.1.3

Depiction of human beings vision: from the eye to a representation of the vision
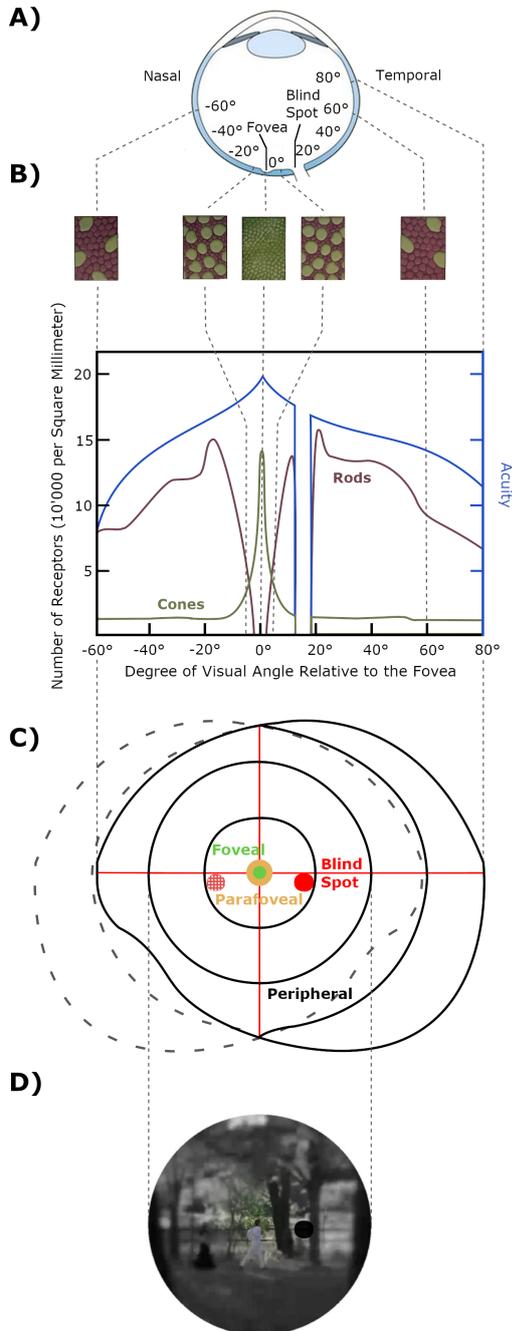
Note. Depiction of vision from the eye to the visual field (VF). A) Representation of a horizontal section of the eye. B) Illustration and distribution of the number of photo-receptors depending on eccentricity according to the fovea (at 0°). Cones are represented in green, and rods in purple. The blue curve represents the visual acuity achieved as a function of eccentricity from the fovea. C) Illustration of the right visual hemifield representing the foveal, the parafoveal, and the peripheral region as well as the blind spot due to the optical nerve. The dashed lines denote the left hemifield of view, and the partial fill indicates the left blind spot for a complete vergence. D) Representations of vision acuity drop-off with eccentricity using a retinal filter on an image. Adapted from figures in Bear et al. (2007), Purves et al. (2011), and Targino Da Costa and Do (2014).

towards the nose, only the oblique muscles are responsible for the vertical movements of the eye, with the superior oblique lowering the gaze and the inferior oblique raising the eye. Torsional movements are other exclusive movements imputed to oblique muscles (see Figure 1.1.4 for a depiction of the eye muscles, and see, *e.g.*, Purves et al., 2011, for a description of the muscles of the eye).

Figure 1.1.4

Depiction of the left eye muscles
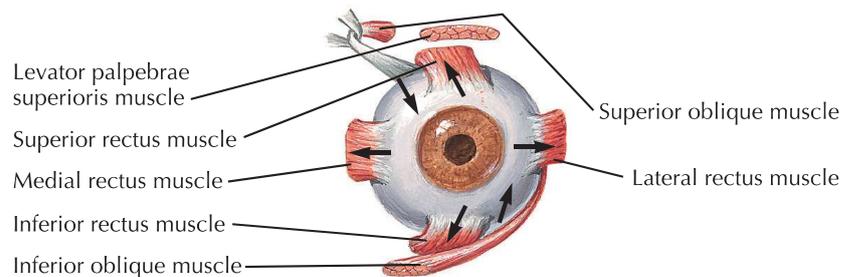
Note. Depiction of a left eye bulb with the six muscles attached to it. The four straight muscles are attached on the sides, top, and bottom. The trochlea of the superior oblique muscle, as well as the oblique inferior, are represented. The six muscles are responsible for all movements that the eye can perform. The arrows indicate the direction in which the ocular bulb moves when the muscle below the arrow contracts. Adapted from Felten et al. (2015).

The action of the muscles described above allows the eye to perform various movements to centre a target of interest and stabilise the gaze regarding head or scenery movements. The repertoire of eye movements entails six types of movements to either stabilise or move the eye. The former movements include *optokinetic* (for a review, see Bender & Shanzer, 1983) and

vestibulo-ocular movements (see Angelaki, 2004). These movements work together to stabilise the gaze and prevent a slide of the retinal image during movement either from the head or the visual world.

Furthermore, there are three types of movement to align the foveal region over a target: smooth pursuit, vergence, and saccades. First, as the name suggests, *smooth pursuit* occurs when one's gaze follows a moving object; it is characterised by a voluntary slow movement of tracking an object along its pathway (for a review, see Spering & Montagnini, 2011). Importantly, it shall be noted that the high visual acuity of the foveal region is preserved as long as the eyes move at less than a 3° visual angle per seconds (see, *e.g.*, Barnes & Smith, 1981; Westheimer & McKee, 1975). Second, *vergence movements* occur in conjunction with lens accommodation and constriction of the pupil's movement, allowing for a clear retina image. Crucially, both eyes do not always land on the exact same location; instead, they might be slightly shifted in location. Vergence movements thus correct for this shift and align both eyes' foveal regions, thereby specifically allowing for a focus on depth changes by increasing the convergence or divergence of both eyes (for a review, see, *e.g.*, Searle & Rowe, 2016; Sethi, 1986). Third, a *saccade* is a ballistic eye movement—the fastest movement that humans can perform. The main goal of saccades is to displace the fovea precisely and gather new information from the VF as fast as possible. When a target is found, and the eyes successfully lend on it with an optimal focus, the eye stops moving to gather visual information. This step, which occurs while the eye is still, is called *fixation*. However, even when the eyes fixate on a target, the eyes do not remain completely still but perform micro-movements. These micro-movements are called *micro-saccades* and presumably preserve neuronal activity from habituation (*e.g.*, Martinez-Conde et al., 2006; Rolfs, 2009).

Dedicated devices were created and used to capture eye movements; they are the subject of the next section. Two specific movements were of particular interest using those devices, namely, saccades and fixations. Both were deeply investigated, as they represent the course of information gathering and change in target. They are discussed in greater detail next.

1.2 Eye Movements: A Window to Cognition

Eye movements have proven to be valuable cues to gain insight into visual cognition. Over the years, technology has improved, thus allowing researchers to investigate when and where eye movements land precisely. Moreover, the advent of technology has allowed for the presented stimuli to be controlled, according to the eye movement's location, to study what feeds the visual system more finely. In addition, eye-tracking technology has become easy to use and affordable. The ease of access to such technology has yielded fruitful results in various domains, such as marketing, clinical psychology, or cognitive neuroscience. Altogether, eye tracking offers a window into cognition, allowing one to draw inferences about the brain and cognitive processes.

This section describes the evolution of the recording of eye movements and the research that led to their classifications according to their nature (steady versus moving). Then, those movements are analysed, and the general findings of the sampling of visual information using eye movement recordings are described. These first themes subsequently lead to a mandatory discussion on the coupling of visual attention with eye movements. Thereafter, gaze-contingent paradigms constraining the visual information (and thus the visual attention) are presented, and the general results obtained with such paradigms are briefly discussed. Finally, the contribution of eye movements to the investigation of the underlying cognitive processes is examined.

1.2.1 History of Eye-Tracking Research

The study of eye movements has been of great interest since antiquity. Already then, the restriction of clear vision to the fovea during fixations was alleged. The formal study of eye movements began with descriptive observations of the visual phenomena (Liversedge et al., 2011). One of the first descriptions was offered by Aristotle, who noted that the eyes moved together in different directions (Forster, 1963). Despite the earliness of the phenomenological description, it took several years, along with technological breakthroughs, before a clear depiction of eye

movements was established. Indeed, the complex nature and multitude of movements elicited by the eyes make them challenging to describe solely through observation. According to Galien (1970), the initial intuition that the eyes move vertically and horizontally was contradicted by the discovery of the two oblique muscles that allow the eyes to roll.

Conjointly with this discovery came the difficulties in interpreting the direction of the eye following a rotation. For instance, Porterfield (1759) denied the possibility of the eye rotating, as he did not feel the rotation of his own eyes. Despite these difficulties, eye movements have long been observed, and the eyes have been thought to glide continuously and uninterruptedly over the visual world. This assumption was strengthened by the perception of the visual experience being uninterrupted (Liversedge et al., 2011). However, Porterfield (1737) deduced that achieving clear vision while reading or looking at a picture requires one to move a small area of high acuity all over words or pictures. Based on this fact, he logically deduced that eye movements were probably sequential. Porterfield's intuition of a small area of high acuity was correct: the acuity of the human VF is maximal at the centre of the VF and decays linearly with eccentricity. More precisely, clear vision is achieved when visual input falls within the foveal region of the VF. This milestone is a key point in understanding how eyes move to sample and extract information from the visual world.

In the 19th century, the understanding of eye movements was deepened thanks to the development of dedicated devices, namely, *eye trackers*. Hering (1879) and Lamare (1892) developed one of the first apparatuses to record eye movements. They both developed a way to listen to the sound produced by oculomotor activity (see Figure 1.2.1). In this device, a tube was directly attached to the eyelid to listen to the crackling sounds produced by the contraction of ocular muscles. Listening to oculomotor activity allowed the researchers to empirically disclose that eye movements were not continuous but rather alternated between jerky and stable phases, confirming Porterfield's initial intuition. Javal (1879) coined the term saccades to describe these jerky eye movements. Later, a new and not only more precise but also more invasive technique was developed (Delabarre, 1898; Huey, 1898, 1900) that used a cup directly attached to the eyeball. The cup was connected to a lever drawing eye movements on a writable medium (see Figure 1.2.1 for a depiction of the eye tracker using a cup on the eyeball). The

Figure 1.2.1

Depiction of the cup and the photographic eye-trackers and resulting recordings
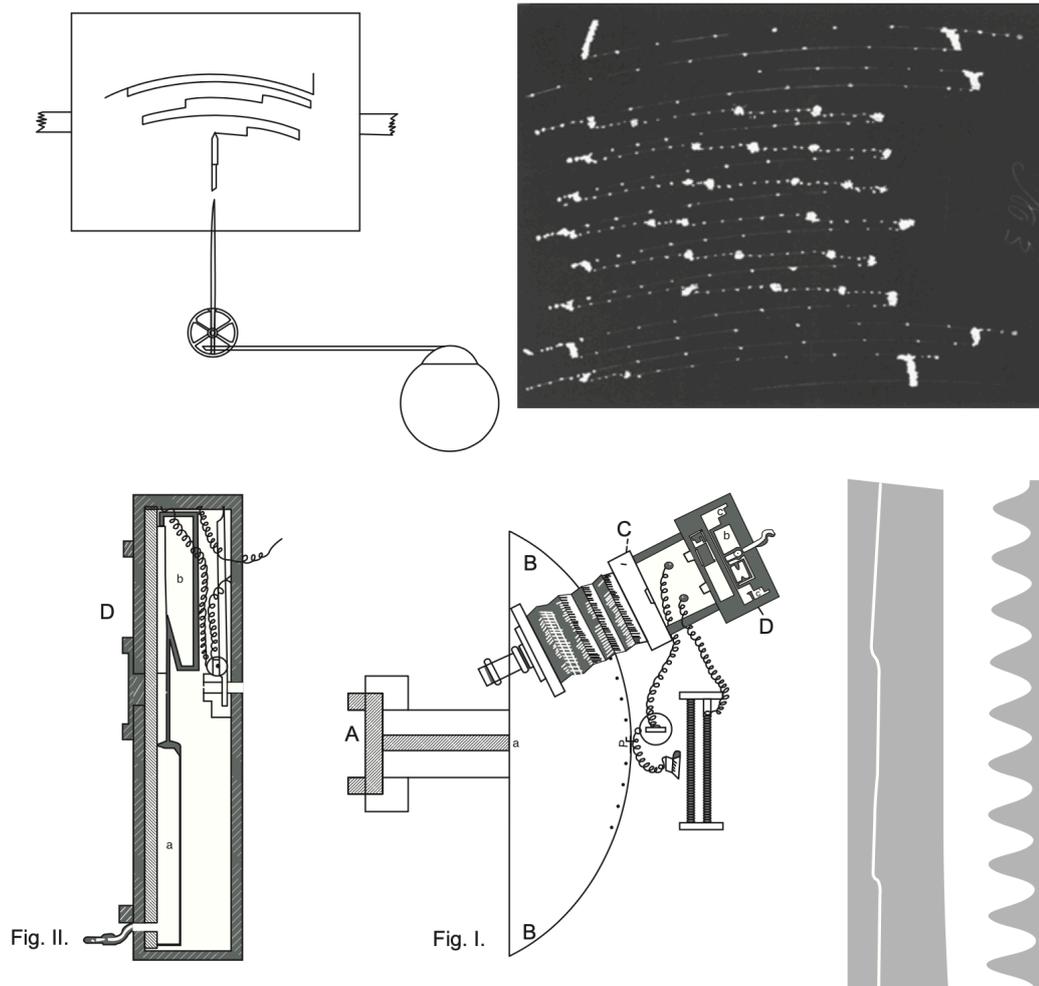
Note. Top left: Depiction of the cup eye tracker developed by Delabarre (1898) and Huey (1898). Top right: Recorded eye movements on a writable medium. Bottom left: Depiction of the photographic eye tracker developed by Dodge and Cline (1901). Bottom right: Recorded trace of eye movements using the photographic eye tracker. Reprinted from Liversedge et al. (2011).

sequence and location of eye movements could subsequently be studied on the medium.

Using these dedicated devices made it possible for researchers to study eye movements and draw empirical conclusions. In particular, the critical conclusion was drawn that the eyes move according to a repeated sequence alternating between steady and rapid eye movements. However, the cup used in those eye-tracking devices added constraints on the eyeballs. To relieve these constraints, Orschansky (1899) created a device without a direct physical connection between the eye cup and the data acquisition device—the connection was dropped thanks to the recording of light reflection from a mirror attached to the cup (Van Gompel, 2007). Later, a significant improvement to eye-tracking technology was achieved by Dodge and Cline (1901), who recorded eye movements through photography, completely relieving the constraints on the ocular bulbs (see Figure 1.2.1 for a depiction of the first photographic eye tracker). More precisely, they recorded light reflection directly from the eye without using a cup. This technique, called *video-based eye tracking*, is still utilised in modern eye-tracking devices (see Figure 1.2.1 for a depiction of this eye tracker).

The technique used by Dodge and Cline allowed the recording of eye movements to be completely free from any constraints on the participant's eyes. Nowadays, eye-tracking devices contain a camera that records the eyes and a range of diodes, emitting infrared light that reflects on the eye. To understand how light reflections can be used in modern eye tracking, it is important to note that the light entering the eye reflects on each compound of the eye (see Section 1.1.3 for a recapitulation of the different eye parts the light traverses throughout the eye bulb). Each stage of the reflection is known as a *Purkinje image*. There are four Purkinje images: the first (P1) is a reflection from the outer part of the cornea; the second reflection (P2) is emitted from the inner part of the cornea; the third (P3) is the reflection of light on the anterior parts of the lens, and the fourth reflection (P4) is emitted from the posterior part of the lens. Eye trackers typically use corneal reflection (P1) to map the eye positions on what an observer is viewing (see Holmqvist et al., 2011, for a detailed description of modern eye-tracking devices and Figure 1.2.2).

In the mid-20th century, the miniaturisation of photographic eye-tracking devices al-

Figure 1.2.2

Depiction of Purkinje images and first Purkinje image during eye tracking
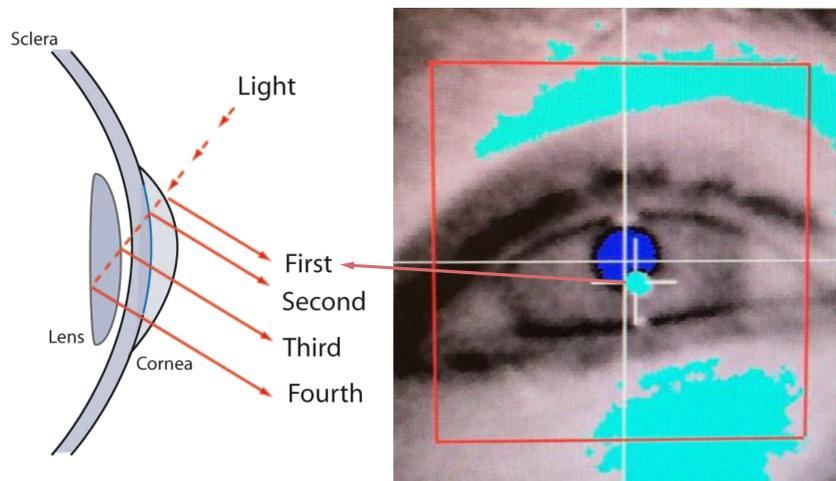
Note. Left: Reflections of light on the eye resulting in the four images of Purkinje. Reprinted from Hansen and Ji (2009). Right: A picture of the first Purkinje reflection used in eye tracking as observed in the user interface of the Eyelink eye tracker.

lowed for the development of wearable eye trackers. In the 1940s, Hartridge and Thomson (1948) developed the first portable (head-mounted) eye tracker allowing the recording of eye movements outside of a laboratory. A decade later, J. F. Mackworth and Mackworth (1958) improved portable eye tracking using techniques employed for television to superimpose one's gaze over a scene. This device, which was relatively bulky and heavy and covered a large amount of the field of view, recorded both the observer's field of view and eyes with a head-mounted ciné camera. More specifically, a series of prisms and lenses provided a magnified corneal reflection that was projected onto the recorded field of view. Later, N. H. Mackworth and Morandi (1967) and N. H. Mackworth and Thomas (1962) studied driving using a camera TV version of this device. However, these devices were still somewhat bulky and were not prominently used in eye movement research. Only in the 1990s, when Land (1992) built a lightweight video-based eye-tracking device, the recording of eye movements outside the laboratory became possible without significantly impacting on the observer's behaviour. This device was composed of a single camera recording the observer's field of view and the eyes through a mirror, providing an image of both the scene and the gaze position simultaneously (see Cognolato et al., 2018, for a review of head-mounted eye tracking).

In recent devices, the illumination and camera are mounted in the frame of glasses (slightly larger than usual glasses). A camera placed on the front of the glasses records an observer's field of view (see Figure 1.2.3). These capabilities have led to new perspectives on recording eye movements in real life using eye-tracking glasses (ETGs). Even more recently, the advent of virtual reality (VR) or augmented reality (AR) led to the possibility of integrating eye-tracking apparatuses in those devices using the same method as the one described above for ETGs. While eye tracking using VR and AR are not yet the norm when studying eye movements, they offer new opportunities to immerse observers in a fully controllable three-dimensional (3D) environment. On top of these improvements, recent advances in 3D modelling from two-dimensional (2D) images have led to the possibility of tracking eye movements using a simple webcam from either a computer or a phone (see, *e.g.,* Papoutsaki et al., 2016; Qu et al., 2017), whereas it previously required dedicated 3D mapping from infrared dot clouds. This technique, albeit far from video-based eye-tracking quality, provides an acceptable tracking quality for estimating coarse gaze location or evaluating the experimental impact on eye movements' characteristics.

17

Figure 1.2.3

Example of eye-tracking glasses (ETGs)

SMI ETG front view

SMI ETG rear view
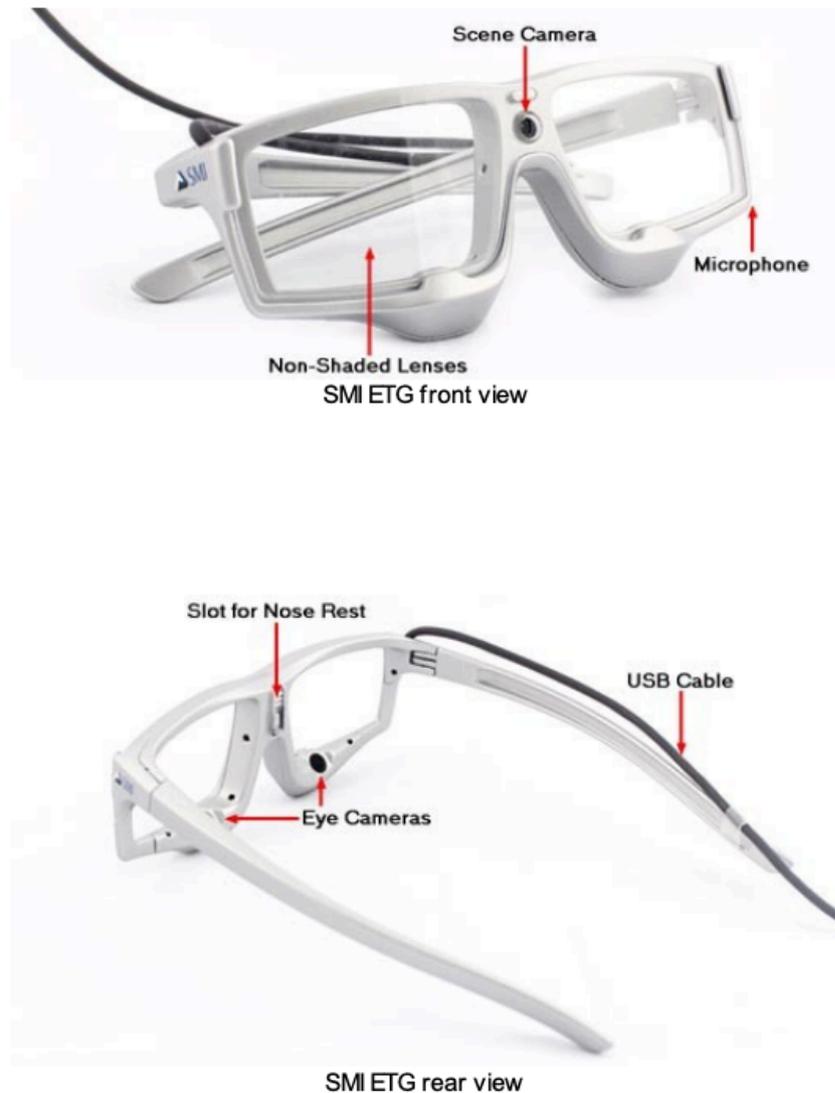
Note. Picture of modern ETGs by the SensoMotoric Instruments (SMI) compagny. A camera recording an observer's visual field (VF) can be seen on the front between the lenses. At the bottom, below each lens, two cameras are dedicated to the recording of the eyes. The black dots on the frame of the glasses are the infrared diodes. Reprinted from the SMI iViewETG User Manual.

18

Those advents in eye-tracking technologies went hand in hand with the rapidly growing amount of knowledge about the different types of eye movements. A cornerstone of our knowledge on ocular events elicited in the course of vision was the scientific definition of the term *saccade* and, crucially, the description of how steady eye movements, namely, *fixations*, land on relevant information to perform a specific task (Dodge, 1916). Nowadays, eye movements are more finely decomposed into events that more precisely describe the scanning of the visual environment and their functions (see Section 1.1.3 above for a description of the other types of eye movements). In contrast to other types of eye movements, the alternating sequences between saccades and fixations were found to occur consistently, even in other vertebrates and some invertebrates (Land & Nilsson, 2012). A surge in the amount of eye movement research followed the classification of eye movement patterns into these sequences of fixations and saccades. The former is thought to gather visual information, whereas the latter is believed to displace the location of information gathering. Crucially, though, as presented later, this sequence involves a more complicated process than initially thought. Moreover, questions regarding where fixations land in the visual environment took unrivalled prominence in the 20th century (Liversedge et al., 2011). Answering such questions requires accurate and rigorous eye movement analyses to assess the differences in eye movement patterns across tasks, conditions, and observers. The next section presents commonly used analyses.

1.2.2 ANALYSES OF EYE MOVEMENTS

Performing eye-tracking experiments produces a significant amount of spatio-temporal data, especially with a modern eye tracker that can record up to 2000 eye movement locations per second (see, *e.g.*, the EyeLink 1000 Plus from SR Research). The data collected translates into X and Y coordinates and timestamps. The collected raw data is then transformed into events characterising how the eyes moved, for example, into fixations and saccades. Afterwards, statistical analysis can be performed on the characteristics of the events to evaluate the impact of manipulation over oculomotor behaviour (see Figure 1.2.4).

In vision research, raw data is traditionally pre-processed and then classified into fixa-

Figure 1.2.4

Scheme of the workflow of eye movement analysis
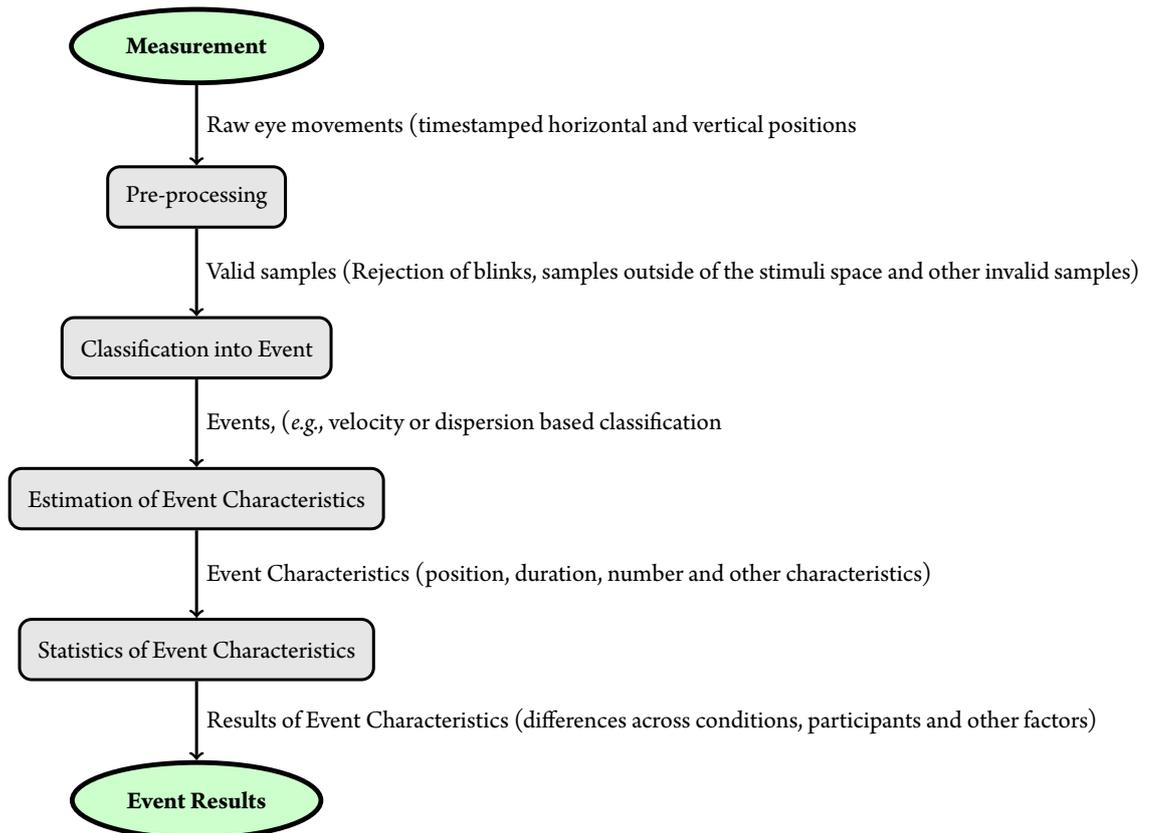
Note. Illustration of the workflow of eye movement analysis. Recorded eye movements are represented according to a time-stamped horizontal and vertical position. Then, pre-processing is applied to reject invalid samples, such as blinks or samples outside the stimuli space. The cleaned eye movement data is subsequently passed in an algorithm that will classify each position to an event according to velocity or dispersion. The resulting data is oculomotor events with their start and end position, their duration, and other characteristics. These characteristics are then analysed and compared across groups, conditions, or participants. Adapted from Lappi (2015).

tions and saccades. For this step, several methods exist to categorise raw eye movement data into those events. The two most common methods are dispersion-based and velocity-based thresholds, depending on the event type of interest. This pre-processing is most often based on a velocity threshold, allowing one to efficiently differentiate between stationary and ballistic eye movement events (see Figure 1.2.5). Nevertheless, for instance, micro-saccades are more accurately identified using clustering methods based on dispersion.

Once raw data has been categorised into events, several characteristics can be extracted from those events, such as the location, amplitude, duration, direction, and number of events occurring within a defined time frame. Vision researchers are traditionally mostly interested in the duration, number and location of fixations as well as the number, amplitude, and orientation of saccades. Still, micro-saccades and smooth pursuit characteristics can sometimes also be of interest (see Table 1.2.1 for examples of the typical range of values observed). Statistical analysis can then be applied to these characteristics to evaluate how a specific manipulation impacts those characteristics.

Table 1.2.1

Range of values observed for the characteristics of different eye movement events

	Duration (ms)	Amplitude	Velocity
Event Type			
Fixation	200–300	-	-
Saccade	30–80	4°–20°	30°–500°/s
Smooth pursuit	-	-	10°–30°/s
Micro-saccade	-	10'–40'	15°–30°/s

Note. Range of values observed for different characteristics of the most commonly studied types of eye movement events. Adapted from Holmqvist et al. (2011).

With a focus on where fixations land, location-based analyses determine regions in the stimuli that have been fixated, called regions of interest (ROIs) or areas of interest (AOIs). The

Figure 1.2.5

Pre-processing of eye movements to categorise fixations and saccades
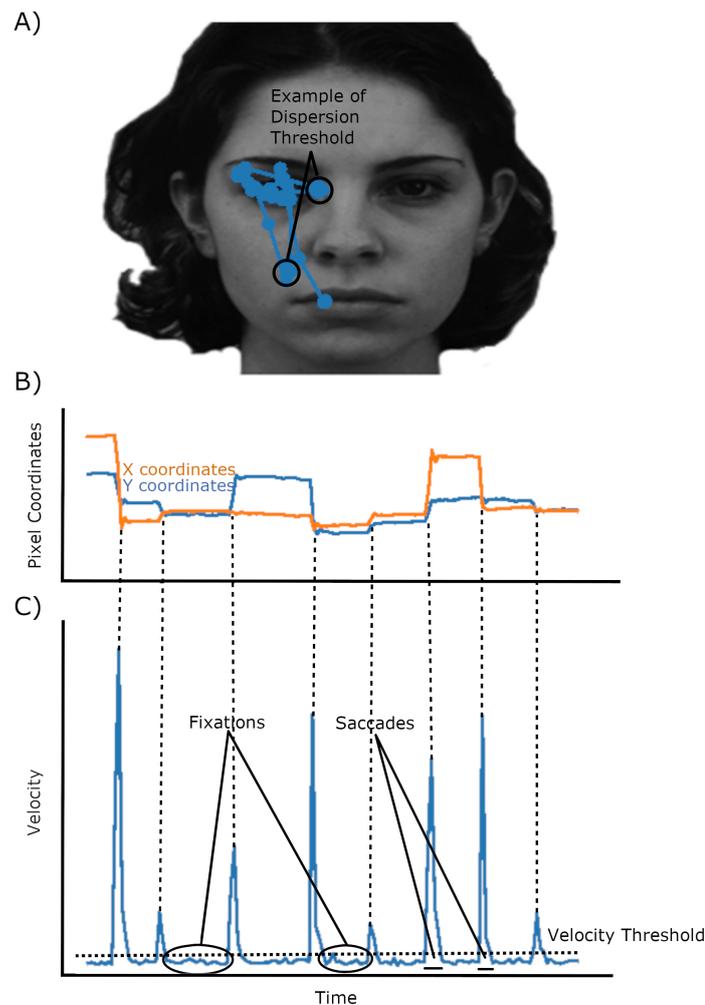
Note. Depiction of eye movement data analysis. A) Image of a face stimulus viewed by an observer. Dots represent points recorded by the eye tracker. B) Plot of the recorded X and Y co-ordinates as a function of time. C) Plot of the velocity computed from raw data. The dotted horizontal line represents the threshold velocity—when above the threshold, data will be categorised as saccades, whereas data below the threshold will be categorised as fixations.

number of fixation landings in those regions is subsequently compared between each ROI. For example, in face-processing research, ROIs are often defined over the eyes, nose, and mouth (see Figure 1.2.6). However, such analyses produce inconsistent results across studies, as researchers define the ROIs subjectively. Furthermore, those methods tend to discard many fixations occurring slightly outside the ROIs. Modern analyses have enabled the elimination of such constrained and subjective investigations using robust and data-driven techniques (for a comparison of both analyses, see Figure 1.2.6). These analyses use an approach similar to the one used in functional magnetic resonance imagery (fMRI). More precisely, this method smooths the landing position of fixations and then involves conducting statistical analysis for each pixel across conditions, participants, or groups (Caldara & Miellet, 2011; Lao et al., 2017).

Another conventional analysis performed on eye movements consists of the evaluation of the ordered sequential alternation between fixations and saccades, *i.e.*, scan paths. More specifically, this analysis aims at evaluating the distance between scan paths. Several methods exist to achieve this comparison. A prevalent method in the early stage of scan path comparisons is called the *string-edit distance* method. As a first step, this method involves dividing the stimulus space into ROIs and assigning letters to each region. Then, to compare scan paths, a metric is computed based on the substitution, deletion, and addition of ROI letter names in the sequence for the scan paths to match (see Figure 1.2.7 for a depiction of this method). The *ScanMatch* method (Cristino et al., 2010) improves and corrects many biases of the string-edit method by aligning scan paths and including gap penalty. ScanMatch crucially includes the temporal aspect of scan paths. Indeed, scan paths are binned according to time and location. More specifically, each letter can be repeated according to the duration of fixations. Moreover, semantic information about the fixated region can be included in the linear distance metric, such as Mannan's metric. Compared to, for example, Euclidean distance, this method computes the distance between the first scan path fixations and the nearest neighbour in the second scan path. This method critically allows one to avoid the use of ROIs.

Other vector-based methods have been developed, such as *MutliMatch* (Jarodzka et al., 2010). As stated above, linear distance methods eliminate ROIs but do not consider the temporality of ocular events. The MultiMatch method uses a vector created between two successive

Figure 1.2.6

Spatial analysis of eye movements (regions of interest [ROIs] vs iMap)
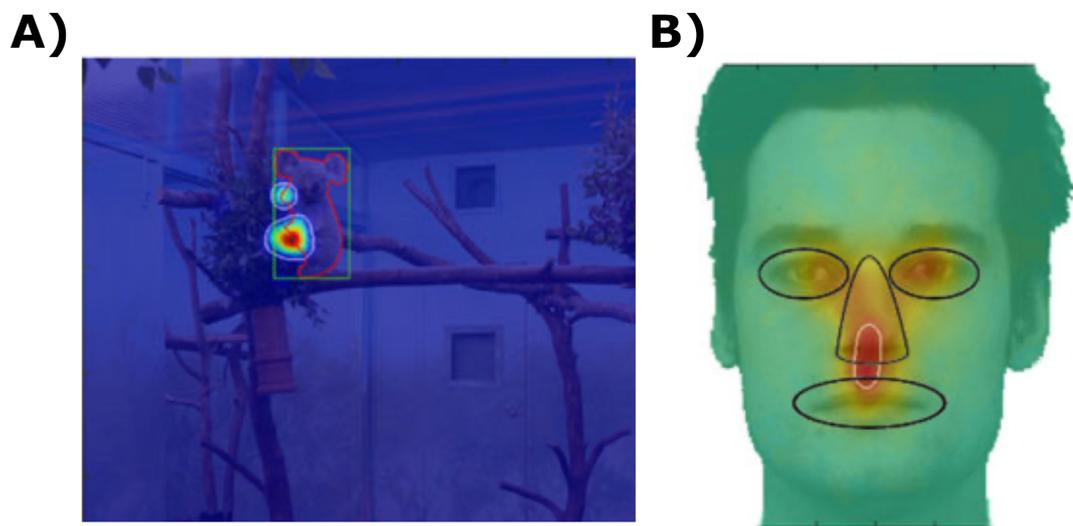
Note. A) Example of ROIs commonly used on trial in visual search (contour of the target in red and box in green). The coloured gradient inside the red contour represents the area fixated significantly above chance. The trial was extracted from the visual search task of Miellet et al. (2010). B) Example of the complexity of choosing adequate ROIs on face stimuli (the black contours indicate the commonly used ROIs on face stimuli). The coloured gradient shows the smoothed density of fixations with the area fixated significantly above chance, according to the *i*Map method circumscribed inside the white contour. This trial was extracted from Caldara et al. (2010). Adapted from Caldara and Miellet (2011).

Figure 1.2.7

String-edit method for scan path comparison
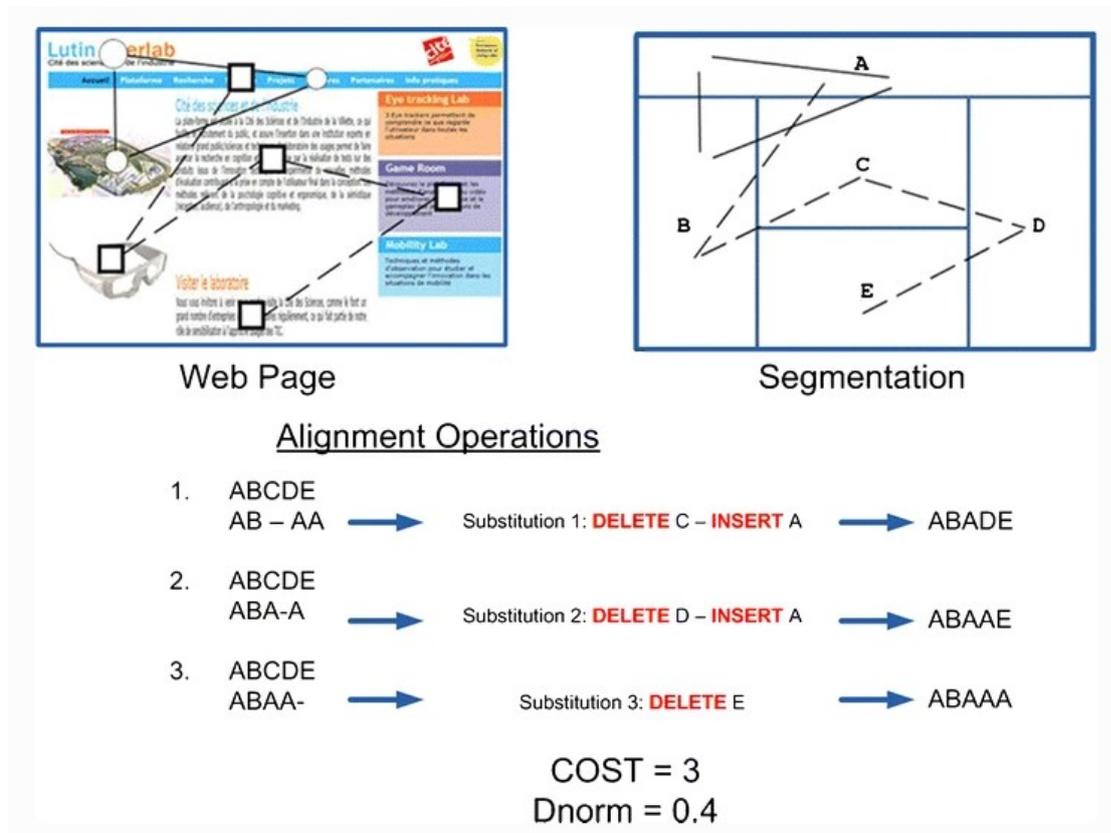
Note. Top left: Two scan paths can be seen reprinted on a web page. Top right: Segmentation of the web page into ROIs with scan paths superimposed. Bottom: Three alignment operations required to match both scan paths. Reprinted from Le Meur and Baccino (2013).

fixations to compare the two scan paths of interest. The strategy involves creating a combination of the first scan path vectors to match the second scan path (see Figure 1.2.8 for a depiction of this method). The advantage of this method is that it can be used to compare any saccade and fixation characteristics. However, it is restricted to the comparison of only two scan paths (see, *e.g.*, Anderson et al., 2015; Le Meur & Baccino, 2013, for a description of these methods).

Other methods indirectly measure the similarity between scan paths. Using saliency maps that represent the fixations' likelihood in a stimulus (see Section 1.2.3 for further information about saliency models), the method for comparing saliency maps computes a normalised score reflecting the extent of the saliency observed at the fixation location. This score can be obtained with various metrics (see, *e.g.*, Le Meur & Baccino,2013, for a description of these methods). However, the most frequently recommended and used method is *normalised scan-path saliency* (NSS); it seems to offer the ideal balance between false positives and false negatives (Borji, Tavakoli, et al., 2013; Bylinskii et al., 2016; Peters et al., 2005). After computing the NSS scores, statistical analyses can be computed to evaluate whether the distances of two scan paths are significantly different from each other or whether participants relied on saliency to the same extent (see Figure 1.2.9). Note that in place of a saliency map, any other map can be used, such as a motion map or another observer's fixation map, to evaluate the extent to which eye movements land on areas encompassing the characteristic depicted by the map.

A new method, known as the *eye movements with hidden Markov model* (EMHMM), was recently created to differentiate between the sampling strategies. An extension of this model includes a switching hidden Markov model (EMSHMM), capturing observers' cognitive state transitions (Chuk et al., 2019). These methods are based on the hidden Markov model (HMM), a extension of a finite state machine with hidden state that enables an evaluation of the likelihood of producing a transition between states when in a specific state. Relating such a computational method to eye movements, each hidden state represents a ROI to which fixations belong, and a transition is viewed as a saccade. The critical advantage of such a method is that it selects the optimal number of hidden states required to achieve an acceptable fit of eye movements—the unnecessary states are pruned. Note that a maximum number of hidden states is often specified for clarity. The clustering of an individual HMM allows for the differentiation between

Figure 1.2.8

Vector-based method for scan path comparison
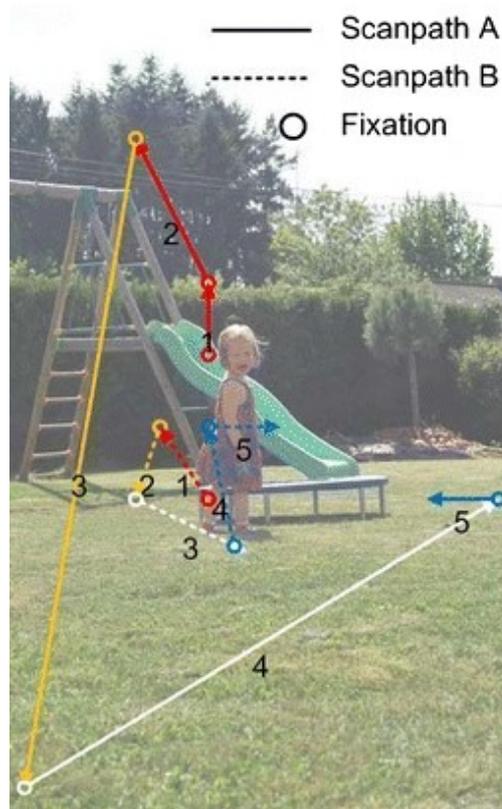
Note. Depiction of the vector-based method: MultiMatch. The aim is to align the saccadic vector between the two scan paths. The best match is represented with the same colours in the two scan paths (red, yellow, white, and blue). This leads to the following match first with Scan Path A and second with Scan Path B: [1,2]<->1, 3<->2, 4<->3 and 5<->[4,5]. Adapted from Le Meur and Baccino (2013).

Figure 1.2.9

Saliency method to compare scan paths
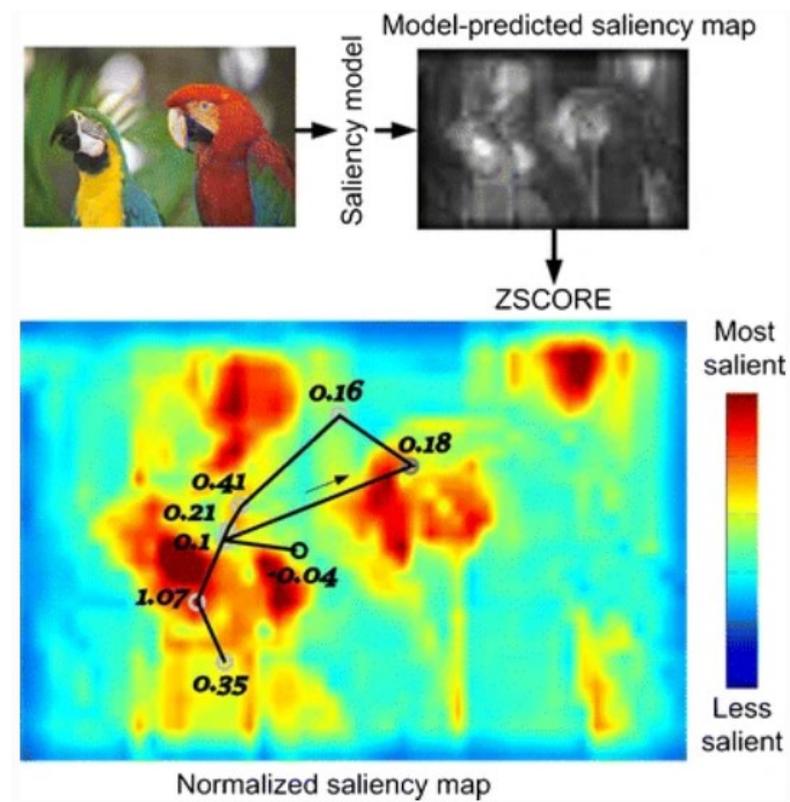
Note. Depiction of the saliency method to compare two scan paths. A scan path is superimposed on the normalised saliency map, and scores are extracted at each fixation location. To compare another scan path, the same procedure is used. Then, both scan paths' NSSs are compared. Reprinted from Le Meur and Baccino (2013).

sampling strategies (see Chuk et al., 2014; Hsiao et al., 2021; for further explanation of the EMSHMM methods, and for application of the method, see, *e.g.* Chuk et al., 2019; Tallon et al., 2021, as well as Figure 1.2.10 and Table 1.2.2).

Figure 1.2.10

Depiction of an analytical sampling strategy as modelled by the eye movements with hidden Markow model (EMHMM) with the corresponding hidden Markov model (HMM)
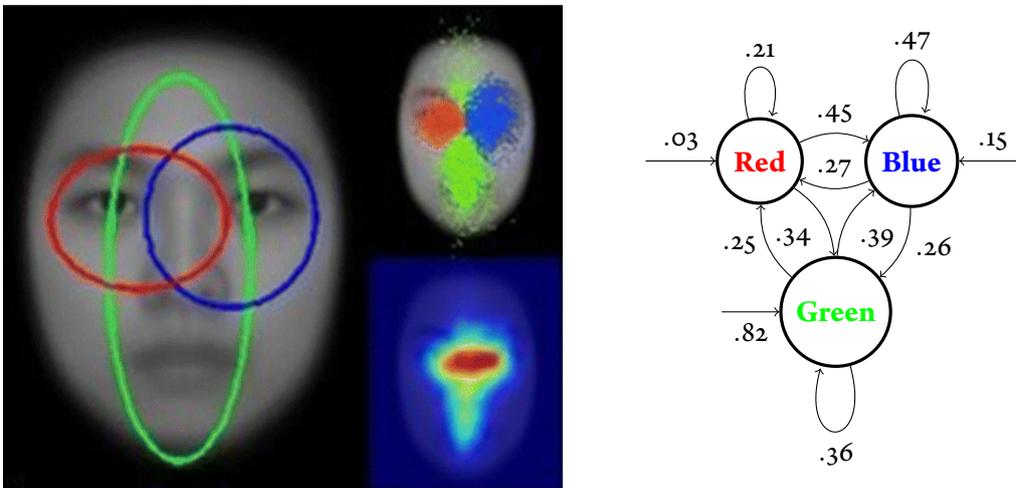
Note. Left: Depiction of eye fixations of individuals classified as using an analytical sampling strategy, the fixation heat map, and ROIs resulting from EMHMM. Right: The resulting HMM, with arrows from the void indicating an entry into the HMM (*i.e.*, the first fixation landing on one of the three ROIs). Arrows between states (red, blue and green) indicate possible transitions with their probability nearby. See also Table 1.2.2 for a tabular depiction of this HMM. Adapted from Chuk et al. (2019).

The methods described above allow for the unveiling of some of the impact of various paradigms on eye movements patterns. They are the main methods used not only in the papers presented in the literature review but also for the contribution of this thesis. However, numerous methods exist beyond those described in this non-exhaustive review of methods to analyse eye movements.

Table 1.2.2

Probabilities of a hidden Markov model (HMM) depicting an analytical sampling strategy, as modelled by the eye movements with hidden Markov model (EMHMM) in between state transitions

	To **Red**	To **Blue**	To **Green**
Prior	.03	.15	.82
From **Red**	.21	.45	.34
From **Blue**	.27	.47	.26
From **Green**	.25	.39	.36

Note. Between-state transitions' probabilities of the HMM depicted in Figure 1.2.10. Rows and columns represent the current ROI/state and the landing ROI/state of the HMM, respectively, with the probability of performing this transition showm at the intersection of the row and column. The *Prior* row indicates the probability of the first fixation landing inside one of the three ROIs. Adapted from Chuk et al. (2019).

1.2.3 EYE MOVEMENTS TO ISOLATE INFORMATION SAMPLING

Following the discovery of alternation between fixations and saccades and modern statistical analysis, studies converged towards the research question of where fixations land. Researchers used various types of stimuli containing either the gist or whole characteristics of the real world and studied eye movements landing positions to answer this question. These differences in stimuli were found to elicit different eye movement patterns. The preliminary results led to increased interest investigating information sampled by eye movements in the 20th century. This rise in research studies led to fruitful discoveries that have deepened our understanding of how eye movements sample the world to extract visual information in both the laboratory and the wild.

The first works answering the question of where fixations land were the seminal works of Stratton (1902) and C. H. Judd (1905b) that examined where eye movements land when participants were presented with geometric illusions. The novelty here rested in the combination of both the fixations analysis and the observers' judgements. As a result, the authors observed that fixation patterns differed according to the observers' judgements. Later, the rationale behind the landing position of eye movements was further described and investigated by Buswell (1935) and Yarbus (1967). They compiled their respective studies and others' research in two books that are milestones for eye tracking and visual perception research. Both observed that eye movements land in similar regions regardless of expertise or idiosyncratic differences. Additionally, Buswell and Yarbus already observed that eye movements changed according to the instruction given to observers (see Figure 1.2.11).

Interestingly, the exploration of pictures began with short fixations, followed by longer fixations, as if observers made an initial broad exploration before landing on specific ROIs containing diagnostic information. This coarse to fine visual processing was observed in most studies evaluating the path of eye movements on stimuli. Moreover, collapsing the location of fixations across observers revealed attractive areas in pictures, such as faces, as confirmed by contemporary authors (Gibson, 1966; N. H. Mackworth & Morandi, 1967). These two modes of information processing highlighted two well-known processing types: top-down and bottom-

Figure 1.2.11

Example of Yarbus' investigation of eye movements
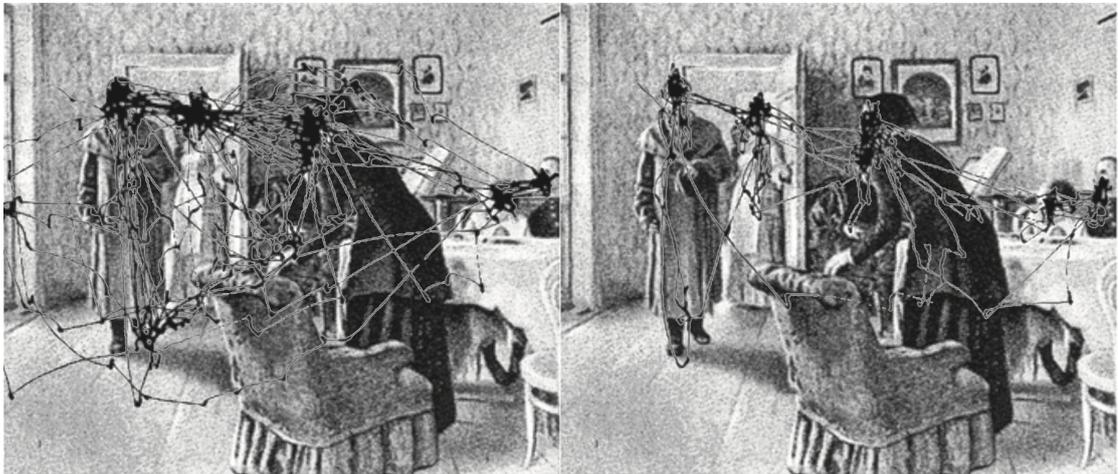
Note. Yarbus' investigation of a single observer's assessment of a picture (An Unexpected Visitor by Repin). Left: The observer was free to observe the picture. Right: The observer was asked to estimate the age of the persons in the picture. While both depictions highlight a focus on the face, this focus is even more present in evaluating age. Adapted from Liversedge et al. (2011).

up processing. Top-down processing refers to previous knowledge used to conduct a task. In contrast, bottom-up processing refers to the information emanating from the stimuli only, such as low-level features (*e.g.*, edges, brightness, etc.).

The visual informativeness emanating from stimuli, *per se* (such as edges, colour, luminance and depth; Henderson & Hollingworth, 1998) were modelled at the end of 20th century and in the early 21st century to take into account the low-level properties of stimuli that attract the gaze (see, *e.g.*, Itti & Koch, 2000, 2001; Itti et al., 1998). These models are called *saliency* models and provide fruitful information about where eye movements land (see Figure 1.2.12, an example of a saliency model computed on a face stimulus). However, the word saliency was used broadly in the literature and referred to the following (Eckstein, 2011):

1. Models considering only the low-level properties of a stimulus, regardless of the behavioural goal (bottom-up information);

2. Models integrating both bottom-up (stimuli characteristics) and top-down information (behavioural goals or knowledge);

3. The general visibility of features that can refer to either of the two definitions above to predict eye movements or even using eye movements to characterise salient areas in scenery.

In this thesis, we adopt the first definition of saliency: the models of low-level properties irrespective of the behavioural goal as defined in the initial model of saliency (Itti & Koch, 2000; Itti et al., 1998; Koch & Ullman, 1985). Although these models accurately evaluate where eye movements land during free viewing, they mostly fail when specific behavioural tasks are required (EinhÄ et al., 2008; Foulsham & Underwood, 2008; Pomplun, 2006; Torralba et al., 2006; for a review, see Tatler et al., 2011).

The possibility of tracking eye movements during real-world behaviour has raised substantial interest in the course of eye movements during specific natural tasks, especially those

Figure 1.2.12

Example of a recent saliency algorithm on a face stimulus
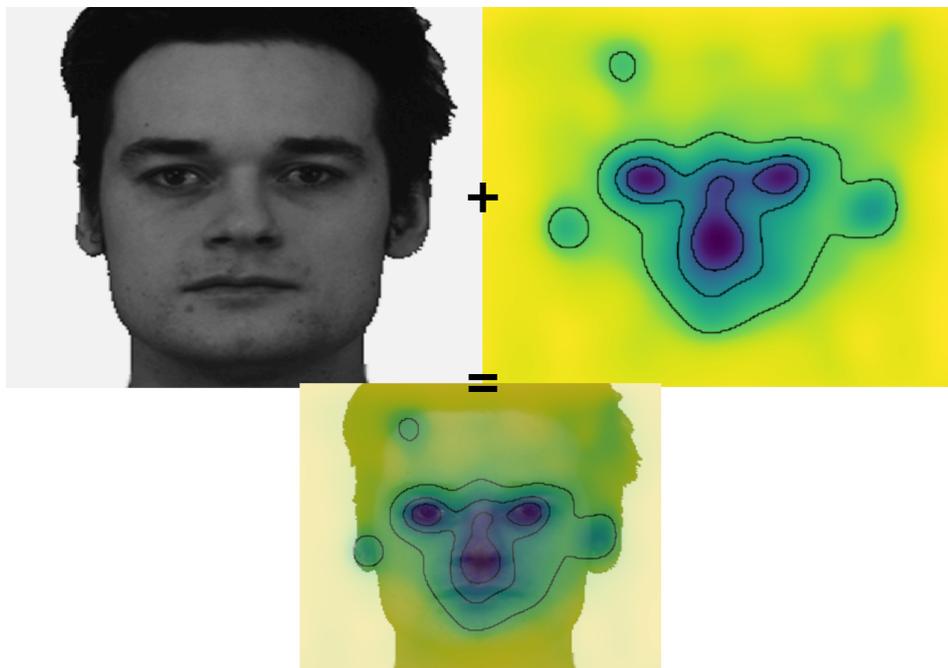
Note. Example of a saliency algorithm applied to a face stimulus. Left: The face stimulus. Right: The saliency computed by the algorithm (*DeepGazeII*; Kümmerer et al., 2017). This algorithm is, at the time of writing, the best-rated model in the MIT/Tuebingen saliency benchmark (Kümmerer et al., 2018).

involving hand–gaze co-ordination. For instance, Ballard et al. (1992) studied how eye movements relate to hand movements when participants were asked to build a tangible model composed of coloured bricks. They observed that the eye usually precedes hand movements. Later, the more general interplay between action and eye movements was investigated in the wild thanks to ETGs. Daily activities such as washing one's hands (Pelz & Canosa, 2001), making a cup of tea (Land et al., 1999) or making a sandwich (Land & Hayhoe, 2001), required an eye movement to precede an action; for example, before one fill a tank with water, eye movements towards the sink will be produced (see Figure 1.2.13). These studies collectively revealed that eye movements serve four purposes in action: locating, directing, guiding, and checking (Van Gompel, 2007). Such anticipatory strategies were also found in driving (see, *e.g.*, Land & Horwood, 1995; Land & Lee, 1994) and in ball sports, especially those with a ball moving at high speed such as a squash ball (see, *e.g.*, Chajka et al., 2006).

Figure 1.2.13

Eye movements during a tea-making task
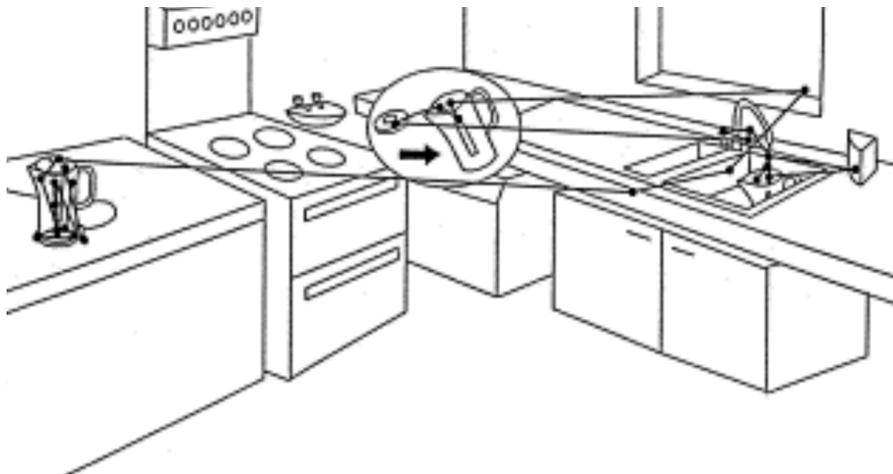
Note. Eye movements recorded when participants were asked to make tea. First, fixations are made on the kettle. The kettle is picked while looking at the sink. The action circled represents the removal of the lid of the kettle while moving towards the sink. At the sink, fixations land on the tap and then goes on the right handle that requires an action. Finally, fixations land on the water stream. Adapted from Land and Hayhoe (2001).

Eye movement sampling revealed that eye movements land according to the characteristics of stimuli engaging bottom-up processes; these processes were further modelled accurately. Additionally, the gaze is also directed according to the instructions provided to the observer and the observer's previous knowledge (*i.e.*, engaging top-down processes). When conducting everyday life activities, eye movements mostly anticipate actions. In addition, eye movement sampling was further dissociated in vision related to recognition and vision related to action (Milner & Goodale, 2006). Altogether, these findings are based on the strong assumption that what is fixated is processed and used to conduct everyday life activities in both recognition and action. Just and Carpenter (1980) posit this eye–mind hypothesis as follows:

"There is no appreciable lag between what is fixated and what is processed." (Just & Carpenter, 1980, p. 331)

This hypothesis would imply that when the eyes fixate on something, the object of fixation is immediately processed when it is fixated. Researchers conducting eye tracking excessively relied on and still rely on the eye–mind hypothesis and information sampling as a direct description of what the brain processes. However, as we will describe in the next section, this is not necessarily true partly because the study of attentional constraints challenges the eye–mind hypothesis. Indeed, the interplay between eye movements and attention is more complicated than this simplified approach of eye movements (see, *e.g.*, Findlay & Gilchrist, 2003).

1.2.4 ATTENTIONAL CONSTRAINT AND EYE MOVEMENTS

Experiments evaluating eye movement sampling have collectively shed light on where fixations land to gather visual information. These experiments were mainly driven by Just and Carpenter's (1980) eye–mind hypothesis. However, the location where the gaze lands was not found to match the attention location exactly.

One of the first signals of attention decoupling from the fixation location came from

Viviani (1990) who already discredited the eye–mind hypothesis by making claims such as, "What meets the fovea is only a part of what meets the (mind's) eye". Later, counterexamples to the eye–mind hypothesis were raised in reading. Indeed, a word that was not fixated could be processed extra-foveally. This processing was especially observed for the word to the right side[1] of the fixated word (for a review, see, *e.g.*, Rayner, 1998). Visual processing could hence occur beyond the foveal region. Allocating attention to the peripheral and parafoveal areas is known as *covert* attention and is not reflected by the fixation location. Fixational eye movements reflecting attention was termed *overt* attention (Eriksen & Colegate, 1971; Eriksen & Hoffman, 1972; Posner, 1982). Thus, the attention at the fixation location can drift towards another ROI in the peripheral vision to already pre-process information.

These shifts in attention during fixation are known as *pre-saccadic shifts of attention* (*e.g.,* Zhao et al., 2012) and allow the eyes to preview peripheral material (Henderson et al., 1989). Previewing the material outside of the foveal region impacts the precision of saccades (Melcher & Kowler, 1999; Vishwanath & Kowler, 2003, 2004) and facilitates the maintenance of perceptual stability and continuity across saccades (Melcher, 2005, 2007, 2009). According to Deubel et al. (2000), this preview constitutes the initial step for programming a saccade. Altogether, these pieces of evidence suggest a mandatory and robust coupling between the preview of regions beyond the foveal region and the programming of a saccade.

Those covert attention shifts in the peripheral vision were initially modelled conceptually as an *attentional spotlight* spatially constrained and moving in the peripheral field of view. Such a spotlight engages attention to pre-process material in the peripheral vision and then disengages to allocate attention to another location (Posner, 1980; Posner et al., 1978; Posner et al., 1980, see Figure 1.2.14). This pre-processing of material enhances the subsequent processing performed when foveation occurs. This model was further improved by introducing the possibility of varying the spatial extent of the spotlight and was called the *zoom lens model* (see Eriksen & James, 1986). However, the extent of attention allocated to the target of interest in the peripheral vision is related to the nature of the target rather than parts of the visual space (Duncan, 1984). Additionally, these models do not consider the fact that the eyes move

---

[1]In the case of a left-to-right reading direction. For other reading directions, this effect will occur according to the direction of reading.

and thus do not relate overt attention to covert attention. In this regard, Henderson's (1992) *sequential attentional model* integrates the course of eye movements.

Figure 1.2.14

Depiction of the pre-saccadic shift in attention on a face
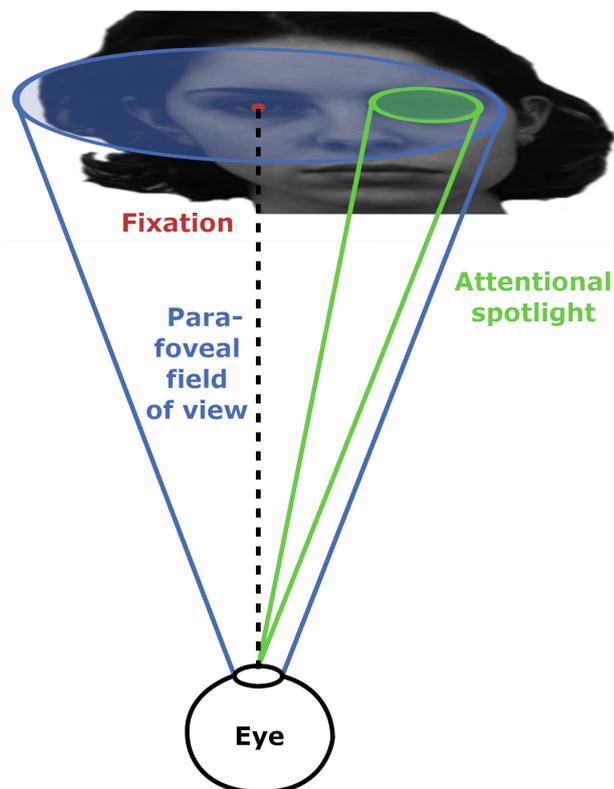
Note. While the fixation (the red dot) lands on the eye, the attentional spotlight (in green) explores the field of view (in blue). On faces, the attentional spotlight will most likely select the other eye to produce the next saccade to land the fixation on the other eye.

Later, in their *premotor theory of attention*, Rizzolatti and colleagues (Rizzolatti et al., 1987; Sheliga et al., 1997) proposed an even stronger link between saccades and shifts of covert

attention. Indeed, they postulated a unique control mechanism for both action and attention, but deriving from a weaker activation of the same neuronal circuits as opposed to the previous current of thought that postulated two independent mechanisms. This view was supported by behavioural (Rizzolatti et al., 1994), neuroimaging (Corbetta et al., 1998; Nobre et al., 2000) and neurophysiology evidence (Ekstrom et al., 2008; Moore & Armstrong, 2003; Moore & Fallah, 2001; Ruff et al., 2006).

In summary, in the course of eye movements, once a fixation has landed, attention is allocated to the centre of the fixation for information intake. When the fixated stimulus is sufficiently processed, attention shifts occur in the peripheral field of view to program the next saccade. Finally, after the execution of the saccades, attention is reallocated at the centre of the fixation location (Findlay & Gilchrist, 2003). Figure 1.2.15 presents a schematic representation of such a process.

Overall, overt and covert attention operate intricately. This entangling was found to be as deep as the neuronal level (Corbetta & Shulman, 1998). As such, the decoupling of overt attention and covert attention do not allow to identify the visual information extracted by an observer from the fixation location solely (Klein & Farrell, 1989; Murthy et al., 2001; Posner, 1982). Indeed, the attention might shift towards information in the peripheral region to preview the material of interest for the next fixation. Modelling attention allows one to conceptualise and make attentional demand tangible as a spotlight. Those modelling attempts seem to me to be similar to atomic models. Indeed, chemistry models depicted electrons as particles. However, electrons are more likely to be a continuous quantity distributed according to a certain probability within the space. In my opinion, attention likely follows such a pattern. It is thus even more important to control the information projected into the observer's VF in order to match, as closely as possible, the visual information sampled by the eye fixations and the visual information used for processing by the brain. Next, Section 1.2.5 presents a computational method to assess the issue of attentional decoupling.

Figure 1.2.15

Alternation between fixations and saccades from the active vision perspective
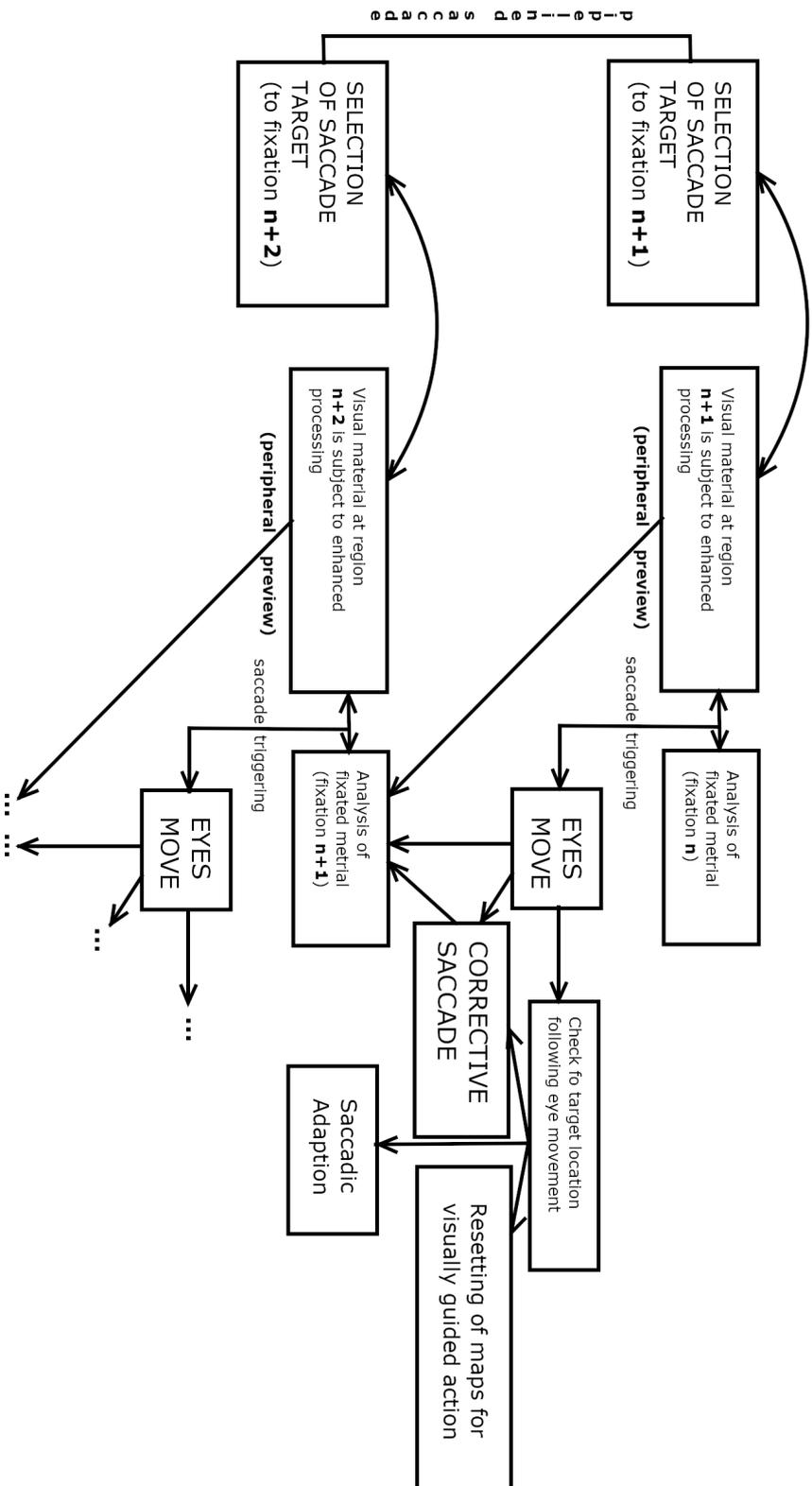
Note. Diagram modelling the alternation between fixations and saccades according to the active vision perspective. Once material has been selected for a saccade, it is previewed peripherally while the fixated material is still under processing (see Figure 1.2.14). Then, a saccade will be produced and corrected if necessary. Finally, the alternation between the analysis of the fixated material and the next target is repeated. Reprinted from Findlay and Gilchrist (2003).

1.2.5 GAZE-CONTINGENT PARADIGM TO ISOLATE INFORMATION USE

The gaze-contingent paradigm uses eye movements to make online changes to a stimulus according to the gaze position or shift. Areas around the gaze location can be changed to reveal, hide or alter the target visual information. These changes occur rapidly enough not to be noticed by the observer. This paradigm allowed for the recreation of an artificial tight coupling between overt and covert attention and yields fruitful results in all fields of vision sciences.

The improvement of eye tracking and computing technologies allowed gaze-contingent paradigms to be applied in laboratory experiments in the 1970s. McConkie and Rayner (McConkie & Rayner, 1975) were the first to employ the paradigm. Their experiment used a gaze-contingent moving-window paradigm, which reveals the target information in the foveal region while hindering the target information in peripheral vision. This paradigm allowed them to precisely control the information feeding the visual system (see Figure 1.2.16A) and to identify the minimal number of letters required to achieve normal reading. Their findings confirmed the rejection of the eye–mind hypothesis experimentally. Indeed, the minimal number of letters required to read effectively—*the perceptual span for reading*—was three to four letters to the left of the fixation (McConkie & Rayner, 1976a; Rayner et al., 1980; N. R. Underwood & McConkie, 1985) and 14–15 letters to the right of the fixation (Den Buurman et al., 1981; McConkie & Rayner, 1975; Rayner & Bertera, 1979; for a review of the perceptual span for reading, see Rayner, 1998). This number of letters comprises at least one word and the word directly to its right. McConkie and Rayner (1975) crucially noticed that completely blocking the shape or context of peripheral information led to a decrease in reading performance. These findings suggest that low-level stimuli characteristics outside of the gaze-contingent moving window remain necessary to process visual stimuli effectively. More precisely, peripheral information is required to program the next saccades efficiently.

The gaze-contingent paradigms provide multiple ways to manipulate information peripherally, parafoveally, or foveally by either revealing, blocking, or changing this information according to the gaze location, the duration of the fixation, the onset of a saccade, or other events' characteristics. These modulations of the gaze-contingent paradigm were used in var-

ious fields of vision research, such as reading or face processing. The following are among the most used gaze-contingent paradigms (see Figure 1.2.16):

1. **The moving window** (also called the *spotlight* paradigm in our lab), which reveals foveal information only and allows for the unfolding of overt and covert attention (McConkie & Rayner, 1975).

2. **The moving mask** (also called the *blindspot* paradigm in our lab), which hides foveal information but reveals parafoveal and peripheral information that allows for a direct assessment of parafoveal and peripheral information use (Rayner & Bertera, 1979).

3. **The boundary paradigm**, which changes peripheral information once an invisible boundary is reached (Rayner, 1975).

Several adaptions have been made to these paradigms. For instance, in reading, the *parafoveal magnification* (Miellet et al., 2009) was used to assess whether the size of the perceptual span in reading was due to the visual acuity drop off with eccentricity. Parafoveal magnification increases the size of letters with eccentricity (*i.e.,* the further away it is from the fixation, the larger the letter is; see Figure 1.2.16B). In face processing, a gaze-contingent, expanding, moving window has been used (see Miellet et al., 2013). This window expands according to the duration of the fixation, (*i.e.,* the longer the fixation, the larger the window). In the initial study, the expanding rate was 1° every 25 ms. This paradigm allows one to map both qualitative and quantitative information intake. The former information is accessed by evaluating the type of content extracted in the window, and the latter by evaluating the window's size. Finally, the last paradigm that is worth mentioning is the *iHybrid* technique (Miellet et al., 2011). This technique mixes high and low spatial frequencies from two stimuli, such as in the famous hybrid image illusion, named *Marilyn Einstein* and created by Schyns and Oliva (see Oliva & Schyns, 2017; Oliva et al., 2006). However, in the gaze-contingent moving-window version, the high spatial frequencies of one stimulus are drawn around the fixation location and the low spatial frequency of another stimulus in the parafoveal and peripheral regions (see Figure 1.2.16E)).

Additionally, these paradigms adapt to both discrete (*e.g.,* letters in reading experiments)

or continuous stimuli (*e.g.*, faces; see Figure 1.2.16 for the use of the gaze-contingent paradigm on text or face stimulus, and see Figure 1.4.2 in Section 1.4.2 for an example of its use in scene processing). Moreover, experimenters can adapt the way in which hidden information is masked. For instance, in face processing, a face can be covered with an average face or a uniform background. However, according to the previous results obtained in reading (McConkie & Rayner, 1975), completely masking parafoveal and peripheral information should be avoided, since such masking disturbs oculomotor behaviour by making it more challenging to program saccades. Additionally, the transition between revealed and hidden information in both the mask and window paradigms can be chosen to be either blurred progressively or hardly defined. A progressive blurred transition between revealed and hidden information should ideally be favoured to avoid attentional disturbances in the peripheral vision resulting from a moving hard border (see Figure 1.2.16C and D for a smooth border spotlight and blindspot, and Figure 1.3.4 for a comparison of hard and smooth borders for a gaze-contingent moving window).

Therefore, the aforementioned gaze-contingent paradigms offer a wide variety of techniques that precisely control the visual information that feeds one's visual system and hence isolate information used, uncovering the information that is processed by the brain. Moreover, they allow researchers to block diagnostic information and evaluate how the visual system reacts and adapts to such artificial visual impairments. Altogether, gaze-contingent paradigms are a powerful tool to artificially couple overt and covert attention and to study how the brain merges visual information resulting from the succession of fixations and saccades to deliver a continuous and coherent visual experience.

1.2.6 A WINDOW TO COGNITION

The previous sections explained that eye movements sample regions following bottom-up processes, (*i.e.*, sampling information related to low-level properties of the stimuli) and top-down processes (*i.e.*, sampling according to higher-order properties of the stimuli such as previous knowledge or instructions). Both of these processes are controlled in the laboratory but they are intertwined in real-world activities. Eye movements are guided through attention that was

Figure 1.2.16

Overview of gaze-contingent paradigms in reading and face processing
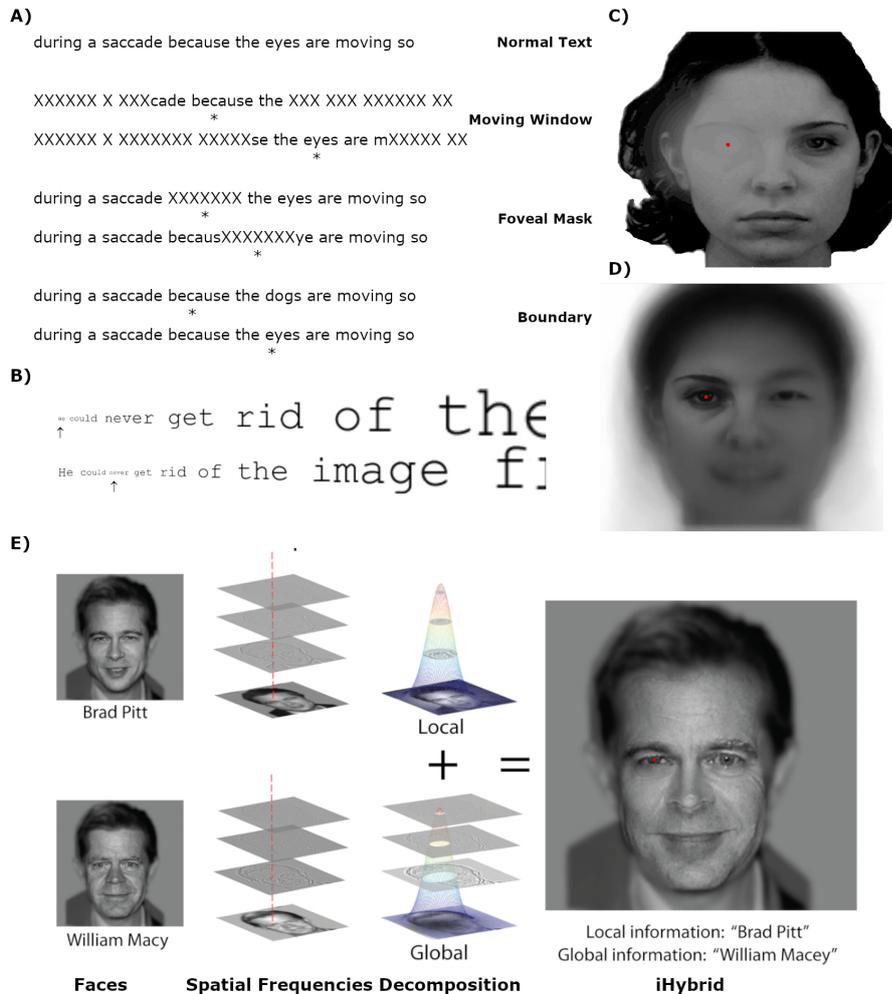
Note. A) Example of gaze-contingent paradigms used in reading. The unaltered text is depicted at the top, with examples of the moving window, the moving mask and the boundary gaze-contingent paradigms below it. The stars represent the current fixation location on the text. Adapted from Rayner (1998). B) Depiction of the parafoveal magnification paradigm. The arrow represents the fixation location. Adapted from Miellet et al. (2009). C) Example of the moving-window paradigm on a face stimulus. D) Depiction of the moving mask paradigm on a face stimulus. E) The *iHybrid* gaze-contingent technique. Adapted from Miellet et al. (2011).

found to be decoupled in overt attention (uncovered by eye movements) and covert attention (not reflected by eye movements). This implies that the information sample does not necessarily match information use. The gaze-contingent paradigms constitute a powerful tool to evaluate information use rigorously. However, while information use discloses information processed by the brain, it remains unclear how visual information is reconstructed to produce a continuous visual perception.

Perceptual continuity was initially imputed to continuous visual sampling. The eyes were thought to glide over the visual world continuously and uninterruptedly, thereby constantly gathering information. Even after the discovery of the alternating sequence between fixations and saccades, sampling was believed to be uninterrupted throughout this alternating pattern (Cattell, 1900; Javal, 1878), which means that the integration of visual information was thought to occur not only during fixation but also before, after, and, crucially during saccades. However, our perception is low or suspended while our eyes move. Indeed, when individuals attempt to see their own eyes moving in front of a mirror, they will notice that the moving eyes cannot be perceived (Erdmann & Dodge, 1898). This suppression of perception during saccades is known as *saccadic suppression* (Tatler & Land, 2011). Using the gaze-contingent paradigm, Grimes (1996) experimentally proved the lack of perception during saccades. In his experiment, he made changes to stimuli when participants produced saccades—these changes remain mostly unnoticed.

While most information seems to be integrated during fixation, the role of saccades in the integration of visual information should not be neglected. A complex system matches the representation before and after a saccade. This integration of information throughout the execution of a saccade is known as *transsaccadic integration* and aims at merging information from previous and subsequent fixations. However, the oculomotor system is far from perfect despite the shift of attention to peripherally preview the next location to be fixated. It presents mismatches between the targeted location and the landing position, and this mismatch remains unconscious and leads to the impression of a continuous and stable visual experience. In short, saccadic suppression and the lack of precision in fixation landing reflect the feat of the visual system to produce a stable and continuous visual perception (Findlay & Gilchrist, 2003).

As a consequence, these results partly invalidate the hypothesis of transsaccadic integration and thus bring forward the hypothesis of sparse representation (Rensink, 2000) or even that no representations were reconstructed from visual sampling (O'Regan & Noë, 2001). However, relating transsacccadic integration to fixation in memory tasks suggests that saccadic suppression should not be overrated, as memory accumulates across successive fixations and recalls of previous contents could occur during saccades (Hollingworth, 2006; Hollingworth & Henderson, 2002; Tatler et al., 2005). The continuity in visual perception should thus be imputed to the brain itself. Indeed, some changes and visual target displacements occurring when eyes move can be noticed under certain conditions but remain unavailable to consciousness.

The alternating sequence of fixations and saccades, as well as the lack of information integration during saccades, has led to the consideration of eye movements, as tools of a passive visual system delivering only samples of visual information extracted from fixations as snapshots. Following this thought process, the successive sampling of retinal images from fixations was believed to be merged to build a representation of the VF. McConkie and Rayner (1976b) even proposed that an integrative visual buffer was used to this end. More precisely, this visual buffer would store several retinal images for the brain to merge. In contrast to this trend, the *active vision* philosophy (see Findlay & Gilchrist, 2003) considers human vision to be a dynamic process during which the eyes continually sample the visual environment. Crucially, following this view, the visual system actively selects and processes the environment during eye movements and thus highlights the importance of understanding the saccadic system.

Beyond the representation constructed from the information sampled by the eyes, another fact that makes eye movements a window to the mind is the tight coupling between brain activity and eye movements. One straightforward example is that when recalling a memory, we move our eyes accordingly (see Ehrlichman & Micic, 2012). Regarding clinical psychology, brain activity was also found to be modulated by eye movements, which were used as an efficient treatment for the distress associated with traumatic memories consisting of a regular motion of the eyes from left to right (see, *e.g.*, Shapiro, 1999; and for a review, see Valiente-Gómez et al., 2017). Additionally, in face processing, a recent study revealed a tight coupling between brain activity and eye movement sampling strategies (Stacchi et al., 2019). More generally, it has been shown that eye movements relate to brain activity during perceptual and cognitive

processing (for a collection of papers discussing this topic, see Nikolaev et al., 2014). Finally, recent evidence has even demonstrated that brain areas implicated in eye movements are also involved in perception in mice (Ahmadlou et al., 2018).

Altogether, these facts suggest a tight coupling between eye movements on the one hand and, on the other hand brain activity, specifically and cognition more broadly. Therefore, eye tracking is a powerful and non-invasive approach for studying information sampled by the eye, information used by the brain, and the underlying cognitive processes. However, one should be aware of this methodology's limitations and avoid making a direct link between information processed and information sampled by fixations. Indeed, it is likely that while integration seems low during saccades, trans-saccadic integration occurs with a certain probability according to the number of stimuli changes, the material previously fixated, and the landing location of the fixation. To this end, active vision offers valuable guidelines. With those perspectives presented, the next sections describe the eye movements patterns elicited when people recognise faces and when they search scenes. More particularly, regular eye movements and the recording of eye movements using gaze-contingent techniques are explained. Next Sections present eye movements sampling and information use during face recognition.

1.3 Eye Movements in Face Processing

Faces are among the first and preferred stimuli processed by newborns (McKone, Robbins, et al., 2011; Morton & J.son, 1991; however, see Arcaro et al., 2017). This early preference highlights the expertise gathered in the processing of faces and the importance of faces for human beings. Indeed, when one looks at a face, a large variety of information is conveyed (Jeffery & Rhodes, 2011), such as age, identity, gender, or trustworthiness, which crucially guides our social interactions. Moreover, faces are among the stimuli to which humans are the most exposed. Interestingly, they cannot be ignored, even when it could be advantageous to do so (*e.g.*, Suzuki & Cavanagh, 1995). Additionally, patients with visual neglect or extinction have been found to distinguish faces, especially expressive faces, more often than other shapes (Fox, 2002; Vuilleumier, 2000; Vuilleumier & Schwartz, 2001), hence explaining the particular interest in

47

the investigation of face processing. As described above, eye movements were proven to be valuable cues to understand the underlying cognitive process. As such, they were envisioned to investigate face processing as early as Yarbus (1967) seminal work that highlighted observers' particular attraction towards faces.

This section focuses on facial identity processing mainly through two paradigms: a face matching task and a face recognition task. In the first task, participants must indicate whether a face is the same as the previous one presented. The second task is more complex and requires participants to learn a set of faces. After the learning phase, a set of faces, including both new and previously learnt faces, is presented. Participants must then indicate whether they have learned the presented face. While both of these tasks imply facial identity processing, they require different memory involvement and cognitive resources.

1.3.1 Facial Information Sampling

Seminal work on facial information sampling, conducted by Yarbus (1967), has already shed light on the landing location of eye movements. His participants were exposed to face photographs, which they freely explored visually for several minutes. From of his experiments, Yarbus discovered a triangular pattern alternating between the left eye, the right eye, and the mouth, with longer fixations on the eyes. This triangular scan path was confirmed in subsequent studies of face processing using free-viewing, matching or recognition tasks (see, *e.g.*, Althoff & Cohen, 1999; Groner et al., 1984; Henderson et al., 2005), and the pattern was thus thought to be universal (see Figure 1.3.1 for an example of a typical triangular scan path).

However, this systematic observation of a triangular pattern was exclusively observed in Western, educated, industrialised, rich and democratic (WEIRD) participants. Although this category of participants is implicitly assumed to be representative of the population, it is a specific sub-set (see, *e.g.*, Henrich et al., 2010). Moreover, this triangular scan path was observed from aggregated eye movements across stimuli and observers (Arizpe et al., 2017; Mehoudar

Figure 1.3.1

Triangular scan path elicited by observers when looking at a face
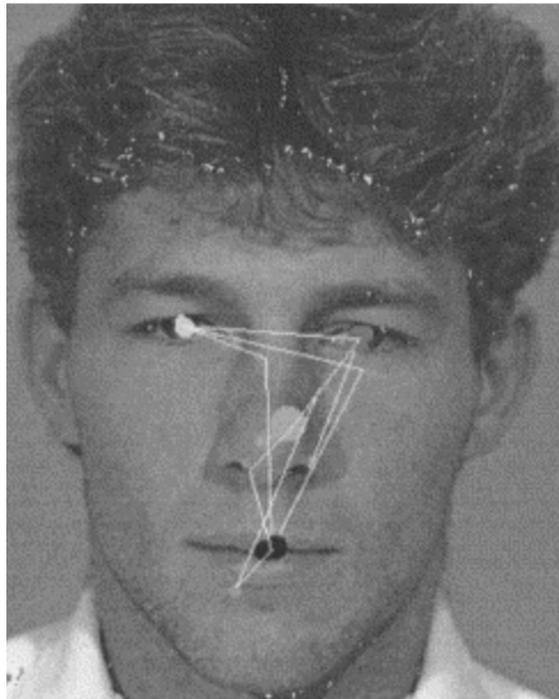
Note. Eye movements elicited by a face recognition task in an observer that exhibits a typical triangular scan path pattern. Adapted from L. Williams et al. (1999).

et al., 2014). Additionally, stimuli manipulations can affect the sampling pattern and cause it to deviate from the triangular pattern. For those reasons, the universality of facial information sampling following a triangular pattern has been challenged. Investigating a broader population, such as clinical patients or participants from other cultures, was required and revealed differences in eye movement patterns. The next sub-sections describe the particularity of eye movement patterns observed in different participants or types of stimuli.

Clinical population

First, clinical patients have been found to elicit a pattern of fixation that deviates from the triangular pattern. For instance, schizophrenic patients have been found to elicits fewer eye movements that are restricted to the upper left region of the face (see, *e.g.* Green et al., 2003; Manor et al., 1999; Phillips & David, 1997, 1998; Streit et al., 1997; L. Williams et al., 1999; for a review see Bortolon et al., 2015, and see Figure 1.3.2 for an example of a typical scan path elicited by a schizophrenic patient). Patients with bipolar disorder also exhibit a pattern deviating from the triangular pattern with very long fixations and a narrow scan path (Bestelmeyer et al., 2006; Streit et al., 1997). Regarding autism spectrum disorder (ASD), a vast amount of research has confirmed both lower face recognition performance in patients with ASD than healthy controls and the eye avoidance hypothesis present in such syndrome (Tanaka & Sung, 2016) supporting that ASD eye movements land mostly on the mouth (see, *e.g.*, Dalton et al., 2005; Davis et al., 2017; Pelphrey et al., 2002; Ristic et al., 2005; Spezio et al., 2007; but see Hedley et al., 2015; Schauder et al., 2019; Vettori et al., 2020). Interestingly, it was found that when the face was decomposed into puzzle pieces, ASD fixation locations were more likely to be on the eyes and mouths (Albrecht et al., 2014), characterising the difference between the whole face and stimuli manipulation in terms of information sampling. Apart from those manifest disorders, it is noteworthy that induced sadness can also elicit an eye avoidance mechanism: sadness-induced participants produced more fixations whilst avoiding the eyes than participants without induced sadness (Hills et al., 2017).

Prosopagnosia

Disorders directly related to face perception impact the fixation pattern elicited in facial identity processing. Prosopagnosia is one of the most studied face processing disorders.

Figure 1.3.2

Schizophrenic patients eye movements over a face
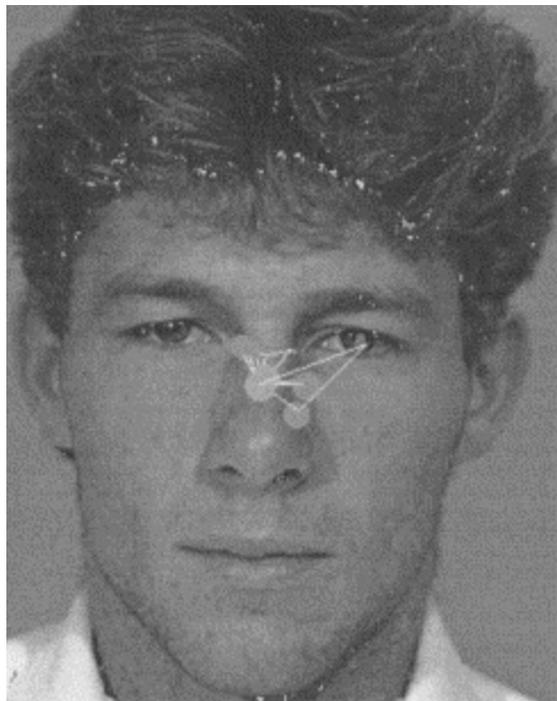
Note. Eye movements elicited by a face recognition task in a schizophrenic patient. This patient exhibits a scan path pattern deviating from the typical triangular scan path observed in control participants. Adapted from L. Williams et al. (1999).

This disorder results in an incapacity to recognise faces. The study of eye movement patterns in prosopagnosia has revealed a deviation from the triangular fixation pattern (Lee et al., 2019; Schwarzer et al., 2007; Stephan & Caine, 2009), with prosopagnosic patients eliciting more stochastic facial information sampling. However, less stochastic scanning patterns were found when the face displayed a positive emotion (Bate et al., 2009). Interestingly, familiar faces seemed better recognised in a 4-year-old patient with congenital prosopagnosia when including outer feature (Schmalzl et al., 2008). However, the P.S. patient with acquired prosopagnosia exhibited a saccadic response accuracy at chance level in a familiarity decision task (Ramon et al., 2018). More globally, general visual agnosia was found to show signs of implicit face recognition with a similar sampling strategy on tasks with and without a face categorisation goal (see, *e.g.*, Lê et al., 2003). All these discrepancies can be imputed to social and perceptual deficits induced by the aforementioned disorders (Arizpe et al., 2017) as well as the available decisional space (Ramon et al., 2018).

Deficits in other senses

Multi-sensory integration can occur across all senses with a modality that either enhance or biases another modality and even creates illusory effects. Therefore, deficits in other senses can significantly impact eye movements sampling strategies in face processing. It is of crucial interest to study the impact of sensory channel loss on eye movements. Crucially, auditory and visual events rarely occur in isolation, and both sensory channels have been found to strongly interact, highlighting the importance of studying hearing loss (for a review of attention and multi-sensory integration see, *e.g.*, Koelewijn et al., 2010). For instance, hearing loss was found to increase the number of fixations on the mouth while showing similar performance to hearing participants in face matching tasks (Letourneau & Mitchell, 2011; McCullough & Emmorey, 1997). Moreover, individuals with early profound hearing loss had an enhanced sensitivity to peripheral vision in certain selective tasks as compared to hearing observers (for a review, see Bavelier et al., 2006). The second contribution of this thesis extends this result to biologically relevant face recognition tasks.

Super-recogniser

On the other hand, people with an exceptional face recognition ability also exhibited different scan paths than the healthy population. In their study, Bennetts et al. (2017) evaluated the

case of O.B., a super-recogniser (SR) whose ability to recognise faces did not extend to other recognition abilities (emotion, age, or gender) or low-level visual processing: O.B. tended to fixate significantly longer on the nose than the control population. Another study investigating eight SRs also yielded this result. Additionally, the eight SRs looked more at inner features (*e.g.,* eyes, mouth and nose) and less at outer features (*e.g.,* clothes, hairs and other features than the face) than controls (Bobak et al., 2017). Altogether these results suggest that persons with the highest face recognition ability mainly fixate inner features, specifically the nose.

Cultural impact

Regarding a healthy population, a cross-cultural effect has been consistently found between Eastern Asians (EAs) and Western Caucasians (WCs). Indeed, Easterners were found to elicit fixations mostly on the nose (see, *e.g.,* Blais et al., 2008). This result was found across free-viewing, identity matching, and face recognition tasks as well as across faces from different cultures or ethnicities (*e.g.,* EA observers will elicit central fixations on both WC and EA Faces; for a recent review, see Caldara, 2017). Note that these two strategies were also found within the same culture using HMMEM (Chuk et al., 2014; Chuk et al., 2019; Tallon et al., 2021). Interestingly, the differences in fixation patterns across cultures are not impacted by social experience (Kelly, Jack, et al., 2011). EA infants' (6 and 9 months) and children's (4 to 7 years) fixation patterns both exhibited a higher performance for their own race than for other race (this effect is known as the other race effect [ORE]) and a scanning path focusing on the nose for EA faces. However, interestingly, for WC faces, fixations were directed towards the eyes (Hu et al., 2014; Liu et al., 2018, respectively). This result highlights the fragility of the sampling strategy during childhood. However, crucially, this cultural fixation bias strengthens with a child's development into adulthood (Kelly, Liu, et al., 2011; Wheeler et al., 2011; for reviews on own and other race effect, see Anzures, Quinn, Pascalis, Slater, & Lee, 2013; Anzures, Quinn, Pascalis, Slater, Tanaka, et al., 2013). The triangular pattern of fixation found on faces is ultimately not universal and varies across cultures or ethnicities (see Figure 1.3.3).

Own face

Apart from the cultural impact on fixation patterns, striking differences were found when participants had to look at their own faces compared to familiar or unfamiliar faces. First,

Figure 1.3.3

Cross-cultural effect on fixation patterns
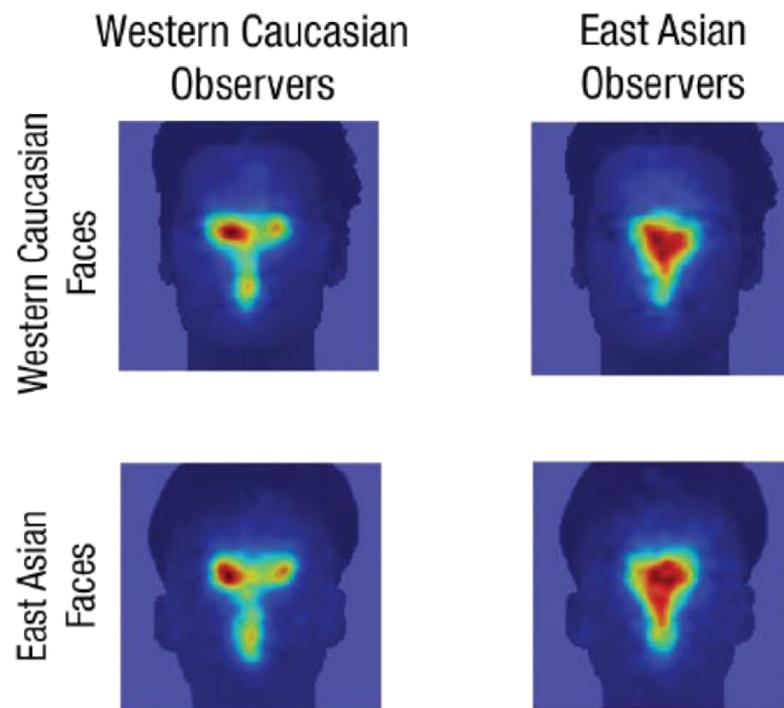
Note. Eye movements elicited by both East Asian (EA) and Western Caucasian (WC) observers on both EA and WC stimuli during a face recognition task. The coloured gradient indicates the density of the fixation (a higher amount red suggests a higher density and a higher amount of blue, denotes a lower density). The pattern of fixation is clearly different across cultures but remains the same across stimuli. Adapted from Caldara (2017).

using a morphing between a participant's own face and unfamiliar faces and two ROIs (hte upper and lower parts of the face), it was found that a closer appearance to the one's own face led to greater fixation on the upper part of the face (Chakraborty & Chakrabarti, 2018). Additionally, participants' own faces received more and longer fixations than familiar or unfamiliar faces (Hills, 2018). The advantages of identifying participants' own faces were both implicit and explicit, as found in tasks of vegetable and face categorisation (Malaspina et al., 2018). As their own faces were often viewed, they seem to have been processed in a specific way compared to others' faces.

Stimuli impact

Several manipulations of face stimuli were found to impact facial identity processing performance. Surprisingly, however, these manipulations did not impact the fixation pattern. One of the manipulations involved presenting a face upside down (face inversion). The impact of face inversion is known as the *face inversion effect* (FIE). While the inversion impacted the accuracy of a face recognition task, it did not impact either the fixations' patterns or duration across features. Thus, the FIE does not result from different fixation patterns (see, *e.g.*, Rodger et al., 2010; C. C. Williams & Henderson, 2007), but rather, in my opinion, from a mismatch with the upright, stored representation of a face. Cueing the first fixation to the left or at the centre of the stimuli, the next fixations were directed towards the eyes with upright faces and towards the mouth when inverted (Arizpe et al., 2012). Interestingly, when the cue was towards the eyes, it reduced the FIE (Hills et al., 2013). Altogether, while not resulting from aberrant aggregated eye movement behaviour, the first fixation seems to have impacted the subsequent fixations, even if a location of the first fixation on the eye reduced this effect. Another paradigm, named the *composite face paradigm*, mixes the top half of a face with the bottom half of another face. Aligning both halves results in the perception of a new face, whereas the misalignment of both halves disrupts this perception (see Young et al., 1987, for more details about this paradigm). This manipulation was found to impact the facial recognition performance but not the eye movement patterns produced (see, *e.g.*, De Heering et al., 2008; and for a review of the composite effect, see Murphy et al., 2017). Altogether, stimuli manipulation does not seem to impact eye movement patterns but rather impacts the subsequent processing.

First fixation

Beyond stimuli variations, analysing or restricting the first fixation on a specific location revealed interesting cues about the optimal position for the first fixation. When participants were free to view a face, a first fixation on the eyes led to higher accuracy as compared to a first fixation on the nose or mouth (Hills et al., 2013). Restricting the first fixation at the centre led to a longer fixation and a delayed first saccade onset. Interestingly, participants predominantly performed the next fixation on the opposite side of the first fixation (Arizpe et al., 2012). Another study found a similar result: the best performance was achieved with a fixation on the cheek—that is, on the centre-left of the face (Hsiao & Liu, 2012). This location matched the preferred landing position (see, *e.g.*, Butler et al., 2005; Hsiao & Cottrell, 2008). Altogether these results indicate that the first fixation location impacts both the facial recognition performance and the scan path. The optimal position for the first fixation seems to be located slightly to the left of the centre of a face, precisely between the cheek and the eye.

Idiosyncratic sampling

More in depth investigations into scan paths elicited by each observer revealed differences among observers and even across trials. Early observations of individual sampling strategies revealed that each observer employed a different strategy that did not necessarily reflect the triangular fixation pattern observed during a face recognition task (Walker-Smith et al., 1977). This finding has recently been confirmed in face matching (Arizpe et al., 2017) and face recognition tasks (Mehoudar et al., 2014). Moreover, these idiosyncratic sampling strategies were found to be stable over at least 18 months (Mehoudar et al., 2014) and did not impair face recognition performance (Peterson & Eckstein, 2013). Indeed, observers looking preferentially at the mouth did not exhibit a different performance, and two observers who looked at the eyes could perform differently (Peterson & Eckstein, 2013). Altogether, these studies demonstrate that the sampling of facial information is idiosyncratic. Moreover, each fixation pattern deployed by an observer is in itself optimal. In fact, forcing fixation on the preferred information for each observer yielded a better performance than for other locations (Peterson & Eckstein, 2013). Importantly, idiosyncratic sampling patterns have recently been found to be coupled with neural activity. More precisely, such optimal idiosyncratic sampling strategies have been found to be tightly coupled with neural responses, both in terms of location, (*i.e.*, preferred location will elicit stronger neural responses) and intensity (*i.e.*, between the duration of fixations and the intensity of neural responses) coupling (Stacchi et al., 2019).

In conclusion, facial information sampling is impacted by various factors, from stimuli manipulation to disorders and even by feelings exhibiting deviation from the triangular scan path. Aggregated data revealed that observers applied two main strategies: either a sampling strategy of each feature of the face or a sampling strategy with fixations directed towards the nose. Furthermore, sampling strategies are idiosyncratic, and observer strategies remain stable over time. More particularly, it seems that if observers can perform their idiosyncratic strategy according to their preferred location for the first location, then visual sampling will occur flawlessly. Regarding SR, EA, and non-hearing observers, they displayed extended parafoveal processing with fixations on the centre of the face, with SR exhibiting a high performance. Furthermore, deficits such as prosopagnosia or ASD often revealed a strategy focused on the mouth, resulting in poor performance. However, those results should be appraised with caution. As discussed in Section 1.2.4, the information sampled by the eye does not necessarily reflect information used by the brain, neither in quality nor in quantity. Precisely identifying the information used for face processing requires the use of gaze-contingent paradigms (described in Section 1.2.5), which are the subject of the next section.

1.3.2 FACIAL INFORMATION USE

Facial information use was unveiled early on with paradigms such as the Bubbles paradigm in behavioural tasks (Gosselin & Schyns, 2001). However, identifying facial information use with eye movements requires the gaze-contingent paradigm, as stated in Section 1.2.4. The most widely used gaze-contingent paradigm to study facial information use is the window paradigm and the mask paradigm. Other paradigms such as the expanding spotlight (Caldara et al., 2010) or *iHybrid* (Miellet et al., 2011) paradigm were used to a lower extent; nonetheless, they provided fruitful results (see Figure 1.2.16 for a depiction of such gaze-contingent techniques). A large variety of methods have been used to manipulate unavailable information (outside the window or in the mask). These methods include utilising an average face with a smooth transition or a uniform background with a hard border transition. Notably, the use of a hard border transition and a uniform background completely hindering global information about a face should be avoided to protect natural sampling strategies from disturbance. Indeed, as stated previously, the seminal study by McConkie and Rayner (1975) demonstrated that completely

blocking peripheral and parafoveal information results in disturbed eye movement behaviours due to a lack of pre-visualisation to program saccades (see Figure 1.3.4 for a comparison between hard and smooth transitions in the case of the moving mask paradigm and Figure 1.2.14 in Section 1.2.4 for a reminder on pre-saccadic shifts of attention). The shape of the window or the mask used were either round or oval. As explained above, since the size of the gaze-contingent paradigms is directly related to the size of the stimuli, in the following section, they are expressed from my own computation, as a percentage of the surface area of the face (height: from forehead to chin, width: from left to right cheek).

Figure 1.3.4

Comparison between hard and smooth transitions in the gaze-contingent moving mask
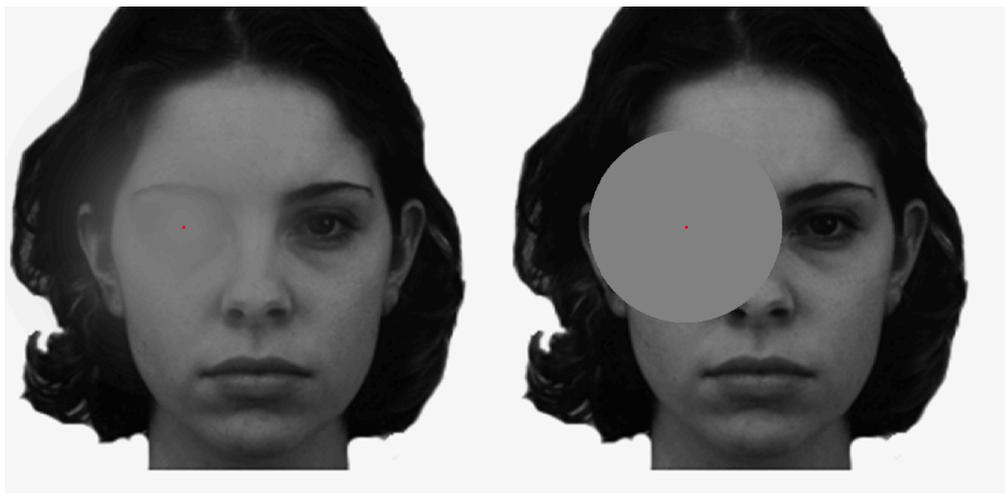
Note. Example of a gaze-contingent moving mask with a smooth (left) and hard (right) transition between the facial information hidden by the mask and the target facial information.

To my knowledge, the first experiment using the gaze-contingent window was conducted by Maw and Pomplun (2004). They used three window sizes: small (4.1° revealing 5.40% of the face), medium (5.5° revealing 9.69% of the face) and large (8.2° revealing 21.54% of the face) during a task in which participants were asked to indicate whether or not the person was famous. The authors found that the smaller the size of the window, the higher the

decrease in performance. Moreover, windows impaired the strategy that participants used in natural viewing. Indeed, they observed a greater fixation on the external features of the face, specifically the hair, with the windows. The smaller the window was, the more this behaviour was present. However, this ocular behaviour can be explained by the use of a hard border not allowing for the next saccade to be programmed and therefore requiring exploratory saccades to identify the position of the diagnostic information (McConkie & Rayner, 1975).

Using a similar range of window sizes (2°, 5° and 8°, revealing 3.39%, 21.63%, and 54.37% of facial information, respectively), Caldara et al. (2010) did not replicate this effect in a face recognition task. No differences in eye patterns were found in WC participants, although a significant decrease in performance was observed. Interestingly, EA participants exhibited strategies modified by the restriction of available information: they used a strategy similar to WCs with the smallest windows (revealing 2° or 5° facial information), while the largest window allowed for the conservation of their strategies. Therefore, a small window size comprising a feature only or less disrupted EAs strategies of central nose fixations, but not WCs' strategies.

In a series of studies, Van Belle and colleagues compared information use via both hard border window and mask paradigms during face matching tasks. Overall, they did not find any difference in eye movement patterns in healthy participants despite the decrease in accuracy observed with the gaze-contingent paradigm. The mask elicited the same performance as a natural viewing condition. Interestingly, the FIE (see the previous section for a description of the FIE) was of larger amplitude in the mask condition than in the window condition. However, both gaze-contingent conditions, which revealed approximately 35% of the face, did not differ from natural viewing (Van Belle, De Graef, Verfaillie, Rossion, et al., 2010). Importantly, increasing the distance between participants and stimuli elicited the same results. One of the studies investigated a well-known case of acquired prosopagnosia, known as P.S., with the same gaze-contingent paradigms but revealing about 50% of the face this time. The patient exhibited the same performance between the window and natural viewing but was impaired by the mask condition (Van Belle, De Graef, Verfaillie, Busigny, et al., 2010; see also Van Belle et al., 2011, for a case of pure prosopagnosia). This result was also found using the same paradigm and task with a smaller window and mask, revealing about 35% of the face, in children, who exhibited a decrease in performance with the mask condition only (Billino et al., 2018).

In these experiments, the mask remained approximately the same size covering one feature at a time. However, this systematic size of the mask does not allow for an evaluation of the potential impairment of different mask sizes. In their study, Miellet et al. (2012) solved this problem and used different mask sizes (2°, 5° and 8°, hiding 3.39%, 21.63%, and 54.37% of facial information, respectively) during a face recognition task across participants' cultures. Interestingly, WCs adapted their sampling strategy with the larger mask such that they presented no differences in fixation pattern compared with the EAs. See Figure 1.3.5 for a recapitulation of the findings elicited by both moving windows and masks with smooth transitions with regards to cultural bias.

To further investigate the change in information sampling strategy by a large mask in the EA population on the one hand and by a small window in the WC population on the other, Miellet et al. (2013) used the expanding spotlight paradigm. This paradigm revealed that the cultural visual bias found in faces related to different spatial frequency tuning. More specifically, EAs process features peripherally at a low spatial frequency due to visual acuity drops off with eccentricity in the VF. In contrast, WCs process features foveally, where information is perceived at high spatial frequencies.

Last but not least, the *iHybrid* gaze-contingent technique was used so that one identity is locally perceived and another is perceived globally (see Figure 1.2.16 for an example of a mix between Brad Pitt and William Macy). Crucially, the results revealed the flexible use of either a local strategy (sampling feature) when the first fixation landed on a feature or a global strategy (fixating the centre of the face) when the first fixation landed on the centre of the face (Miellet et al., 2011). This suggests that face processing might be more flexible than previously thought.

Altogether, these results indicate that observers could cope with a mask or blindspot covering up to 54.37% of facial information without suffering from a significant performance decrease, but with population eliciting an analytical sampling adapting their sampling strategy towards a holistic sampling strategy. However, observers were impaired by a window revealing less than 54.37% of the face, comprising a feature with its surroundings. Furthermore, this result was confirmed by one of the contributions of this thesis that specifically addresses the

Figure 1.3.5

Compilation of gaze-contingent paradigm results in face processing
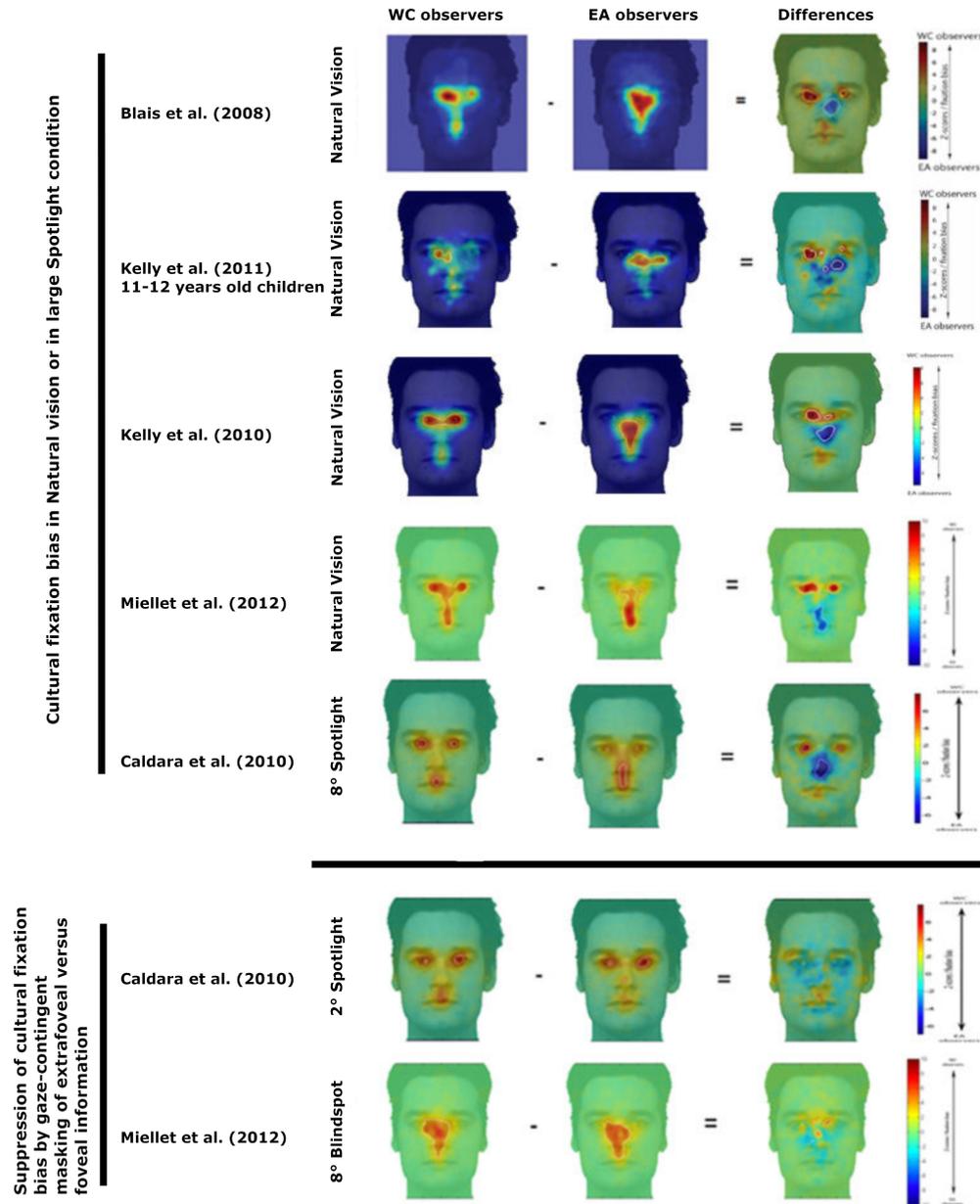
Note. Recapitulation of findings related to cultural bias using gaze-contingent moving-window paradigm and the mask paradigm. Adapted from Miellet et al. (2012).

question of the minimal quantity of information required at each fixation during face recognition. In contrast, EAs, children, and prosopagnosic patients exhibit the opposite results and suffer in the presence of smaller window sizes but not a large mask. This conclusion indicates that those populations are likely to rely more on extra-foveal information than the control adult population. Crucially, the results of gaze-contingent studies confirm the results of facial information sampling. Despite the need for a minimal amount of information, the face system flexibly switches to the most appropriate strategy in order to sample and use facial information efficiently, according to the preferred strategy and where the first fixation lands.

1.3.3 Inferences on Face Processing from Eye Movements Research

Various arguments exist regarding the way in which visual processing occurs, depending on the different field of perception. Two main processes of the visual environment were proposed: either as a whole or as parts thereof. This distinction is fundamental because, according to *Gestalt*, the whole differs from the sum of its parts. Similar debates exist in face processing, and faces were claimed to be recognised as a whole rather than through their individual parts. More precisely, a face is not thought to be perceived as a collection of features, but rather the whole face is built from individual parts (Tanaka & Simonyi, 2016).

Many hypotheses have been posed to explain face processing perception construction. Three central hypotheses emerged, and have been debated: the featural, the configural, and the holistic hypothesis. The featural hypothesis states that faces are processed according to each feature alone (eyes, mouth, and nose) and is also coined the first-order feature (see Figure 1.3.6). According to the configural hypothesis, faces are processed by the relation between the aforementioned features. These relations are also named second-order features. Finally, the holistic hypothesis states that faces are processed as a whole and that two or more features are processed simultaneously. More particularly, according to holistic processing, a combination of several first-order and second-order features are processed at once (see Figure 1.3.6, and for a review, see, *e.g.*, Piepers & Robbins, 2012).

Figure 1.3.6

Diagram of features integrated to build a representation and the possible combinations of those features

Higher-Order Feature / Attribution using Several (or all) of the Second-Order (Relational) Features

Higher-Order Feature / Attribution using a combination of First-Order and Second-Order Features

Increase in Second-Order (Relational) Information used

Second-Order (Relational) Feature Relationship between two First-Order Features e.g. spacing between the eyes

Higher-Order Feature / Attribution using several (or all) of the First-Order Features

Increase in First-Order (Isolated) Information used

First-Order (Isolated) Feature Single feature and the details pertaining to it e.g. a nose (which might be big or pointy).

Note. First-order features relate to the featural hypothesis. Second-order features are used in the configural hypothesis. Higher-order features represent the holistic hypothesis and hence serve as a loose definition of this hypothesis. Reprinted from Piepers and Robbins (2012).

63

Unfortunately, the holistic hypothesis of face processing has often been found to be too loosely defined (Richler & Gauthier, 2014) compared to the other two hypotheses. This is likely due to the numerous possible combinations of first-order and second-order features implied by the holistic hypothesis. This hypothesis is consequently unfalsifiable, as a holistic representation can explain every pattern of fixations and all disturbances in face processing by claiming a holistic processing disruption. Indeed, on the one hand, holistic face processing has been invoked from fixation patterns (Chan & Ryan, 2012; Guo, 2012) and behavioural performance in gaze-contingent paradigms as seen above (Van Belle et al., 2011; Van Belle, De Graef, Verfaillie, Busigny, et al., 2010; Van Belle, De Graef, Verfaillie, Rossion, et al., 2010). However, on the other hand, other authors claim that holistic processing is independent of fixation patterns (De Heering et al., 2008).

Eye movements were found to match facial identity representations (Bombari et al., 2009). Information use, extracted from gaze-contingent paradigms, is thus highly likely to aid in accurately identifying facial representations built for facial identity processing. The results discussed above on information use might hence be linked to facial identity representation. The construction of representations might consequently be flexibly built according to the landing of eye movements and might also be idiosyncratic. Indeed, a recent study revealed that the idiosyncratic deployment of fixation patterns is tuned with facially sensitive neural responses (Stacchi et al., 2019). As such, in my opinion, holistic processing might occur in the early stage of visual processing to gather coarse information from the first fixations. The holistic processing might pursue subsequent fixations solely in populations with either immature or impaired face processing. In contrast, healthy adults use the first fixation to grasp the gist of a face and program subsequent facial information sampling from eye movements gathering fine-grained details about the featural and configural aspects of the face. From my point of view, this processing aims to match the representation built from the information sampled by eye movements with the facial representation stored in memory to proceed to the final identification.

1.4 EYE MOVEMENTS IN SCENE PROCESSING

A scene is a view or the gist of the real-world environment from a particular vantage point. This vantage point can present a specific aspect of the spatial structure of the scene. For instance, a scene can depict either a street or a particular building on the street. Scenes typically encompass a background and multiple spatially distributed objects, providing a coherent semantic of the scene (Henderson & Ferreira, 2004; Henderson & Hollingworth, 1999). Scenes are among the most complex stimuli used in eye-tracking studies. Notably, the objects in a scene are presented in a context that depicts a more accurate real-world scenario. As such, the semantic congruence between objects and scenes can be manipulated, allowing one to tune the highlight of an object of interest in the scene, typically in visual search tasks that are widely used when investigating scene processing. The act of searching occurs on many occasions in everyday activities, for example finding one's wallet before leaving one's flat or finding one's friend in the university auditorium. Crucially, with the advent of technology, a tremendous amount of time is spent in front of one or even several screens, searching for icons, buttons or other interactive materials in order to interact with computers or phones (Eckstein, 2011). In addition to the fact that scenes more accurately reflect the real world, visual search tasks frequently represent typical tasks occurring in the real world.

The following sections focus on eye-tracking studies with scene stimuli using either free viewing or visual search. Moreover, only studies employing drawings, pictures or videos of real-world scenes are considered in the literature review. Studying eye movements in scene viewing is important, as it brings knowledge of how visual information is acquired and represented when surroundings are included and thus contextualises objects. Scenes are consequently closer to the real world than other stimuli used in visual processing.

1.4.1 SCENE INFORMATION SAMPLING

Scene processing is a seminal field of study in eye movement research: with Buswell (1935) conducting the first study on fixation location. In his study, he asked participants to look at pictures and found that persons and faces attracted more eye fixations than the background and that such behaviour appears early in scene exploration. Yarbus (1967) described a similar result in his book. However, the attractiveness of social stimuli in scene processing was only recently found to occur as early as the first fixation (Fletcher-Watson et al., 2008). Apart from the attractiveness of social stimuli, Yarbus noticed a systematic preference for location either encompassing or judged as encompassing the most useful information for scene processing. These findings led other researchers to investigate what in an image drives eye movements and whether the task has any impact (for a review, see Tatler, 2009).

By evaluating viewer ratings, researchers have confirmed the finding that eye movements land on regions with the highest informativeness. More particularly, studies decomposed a scene into a grid of regions in which each participant evaluated informativeness. This protocol empirically confirmed Yarbus' (1967) and Buswell's (1935) assumption of a correlation between eye movements and a highly informative region (*i.e.*, encompassing the most diagnostic information; Antes, 1974; N. H. Mackworth & Morandi, 1967, see Figure 1.4.1).

The efficiency of diagnostic information sampling is possible thanks to the visual system grasping the gist of a scene early on. Indeed, the meaning of a global scene (*i.e.*, scene type, general spatial layout, and the identity of some objects) is grasped by the visual system during the initial fixation (see Biederman, 1972; Biederman et al., 1982; Henderson, 1992; Intraub, 1980, 1981; Potter, 1976; Schyns & Oliva, 1994; Thorpe et al., 1996; VanRullen & Thorpe, 2001). The rapidity of the formation of the gist of a scene is likely linked to the reactivation of previously memorised related scenes (Friedman, 1979; Lampinen et al., 2001; Mandler & Johnson, 1977). Thus, the semantics of a scene can be extracted as early as the first fixation.

A spontaneous task performed when presented with a scene is known as a search. Visual

Figure 1.4.1

Stimuli segmentation for informativeness study
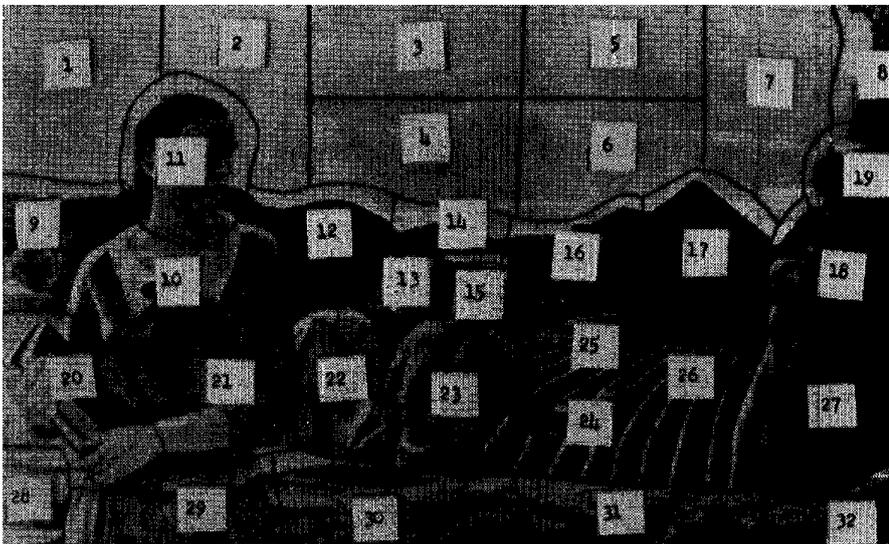
Note. Depiction of one of the segmentation used in the study of Antes (1974) for participants to judge the informativeness of each region. Adapted from Antes (1974).

search is, thus, a widely used task to study information sampling on a scene. This task involves a target that observers must find in a scene. During visual search tasks, eye movements are drawn towards the target (L. Williams, 1966), but distractors disturb the search (Findlay, 1997). In their study, Loftus and Mackworth (1978) discovered that observers could determine the consistency between an object and its scene in a single fixation. Furthermore, inconsistent objects attracted more fixations than consistent objects. However, later, other studies did not manage to replicate the consistency of participants' grasp of a scene in the first fixation, but replicated the attractiveness of gaze by an inconsistent target (De Graef et al., 1990; Henderson et al., 1999).

In these studies, it is difficult to disentangle the impact of the attractiveness of gaze due to semantic information *per se* from visual informativeness. Saliency was used to highlight the low-level characteristics of scenes and hence uncover the scenes' visual informativeness (see Section 1.2.3 for more information on saliency). Although saliency models accurately predict eye movements during free viewing of a scene (Parkhurst et al., 2002; Peters et al., 2005), they fail in more complex tasks. For instance, in the case of visual search, saliency does not seem to accurately map one's sampling strategy during this task (Henderson et al., 2007). Additionally, saliency seems to be less accurate for observers who have domain-specific knowledge (G. Underwood et al., 2009). This is not surprising, as most saliency models adopt bottom-up processes and do not consider top-down processes that impact eye movement sampling strategies. Top-down processes were only recently considered in saliency models (Voorhies et al., 2012); therefore, the modelling of top-down processes is still in its infancy, although it is promising for shedding light on the intertwining of bottom-up and top-down processes in scene searches.

Regarding differences across participants, culture was thought to influence sampling strategies. As mentioned for face processing in Section 1.3.1, EAs tend to process stimuli more globally, whereas WCs tend to process stimuli more locally. In scene processing, cultural differences were initially reported in the description of underwater scenes: WCs tended to describe focal objects, whereas EAs were found to describe the background (Masuda & Nisbett, 2001). Additionally, the sampling strategies were found to differ across groups, with EAs more often fixating the background and WCs fixating objects (Chua et al., 2005). However, several studies did replicate those results and did not find any differences in sampling strategies during scene

processing across cultures (Evans et al., 2009; Miellet et al., 2010; Rayner et al., 2009). Thus, while behavioural differences in the appreciation of a scene have been consistently found (*e.g.,* Nisbett & Masuda, 2003), they do not always impact eye movement sampling strategies.

In conclusion, information sampling during scene viewing results from a complex interplay between the low-level properties of a scene (bottom-up features) and goal-related properties (top-down features). While bottom-up saliency was found to model scenic information sampling during free viewing efficiently, it fails in more complex tasks such as visual search. This is most likely due to the lack of a model of top-down processes. Interestingly, the semantics of a scene can be grasped as early as the first fixation. This way of processing scenes suggests coarse-to-fine processing throughout the exploration. Additionally, the method for processing faces was particularly impacted by the degree of specificity of the task. However, in contrast, to face processing, culture does not seem to impact one's eye movement sampling strategy for scene processing.

1.4.2 Scene Information Use

Only a few studies have investigated the information used for scene processing, *per se*. The gaze-contingent paradigm was implemented mostly for visual search tasks. Two gaze-contingent paradigms, namely, the window and the mask paradigms, were used to investigate information use during visual search in scenes. While only a few such studies exist, they shed light on how visual information is used to conduct visual search efficiently and how the restriction of information intake impacts visual search oculomotor behaviour and performance.

The first study employing the gaze-contingent paradigm was carried out by Saida and Ikeda (1979) using moving windows to restrict available information. This study aimed to evaluate the necessary information to avoid a disruption of eye movement behaviour and processing speed through a gaze-contingent paradigm. In their study, Saida and Ikeda (1979) used windows with a visual angle of 11.40°, 7.30°, 4.60° and 3.30°, revealing 50%, 20%, 8% and 4%, respectively. They found that the impact of the gaze-contingent paradigm was relieved when

50% of the picture was revealed (*i.e.,* a window with an 11.40° visual angle). More recently, a similar study was conducted using a moving window with smooth transitions and blurred backgrounds during the visual search on real-world pictures. It evaluated the amount of information required to achieve normal eye movement behaviour and performance. On average, observers required a window of 8°, (*i.e.,* 25% of the picture size) to perform as they would natural viewing conditions (Nuthmann, 2013, see Figure 1.4.2). Altogether, this suggests that a window revealing at least 25% of information is sufficient for a person to process scenes as they would under natural viewing conditions.

Figure 1.4.2

Gaze-contingent paradigm to evaluate the perceptual span for visual search
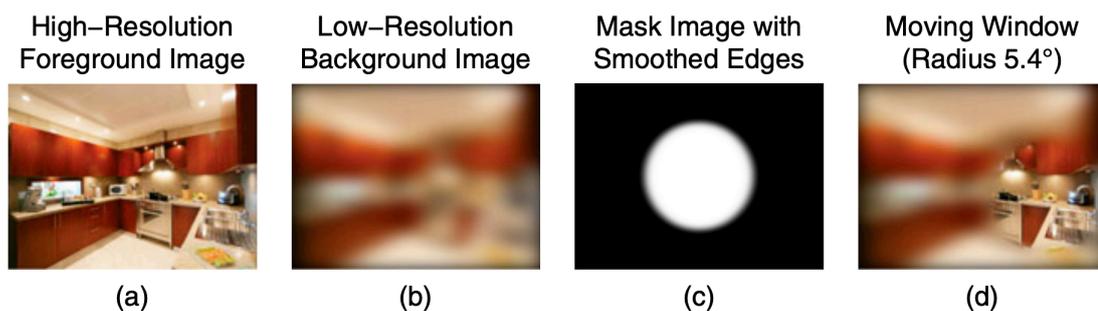
Note. Construction of gaze-contingent stimuli by Nuthmann (2013). a) Stimulus without any transformation. b) Blurring of the stimulus used as a background. c) Alpha values of the window used in the gaze-contingent moving window. d) Resulting stimulus when the three images are merged. Reprinted from Nuthmann (2013).

In another study, Foulsham, Teszka, et al. (2011) used gaze-contingent paradigms with rectangular (visual angle of 6.20°) and asymmetric moving windows oriented horizontally (bisual angle of 12.5° × 3.1°) or vertically (visual angle of 3.1° x 12.5°); interestingly, all windows revealed approximately 5% of the stimuli. Their results suggest that a contrast-adjusted background does not impact behavioural performance and eye movements, contrary to blurred or grey-scale backgrounds. This finding indicates that, as evaluated in this study, when eye movements are efficiently guided by peripheral vision, saliency is not an ideal predictor for such a process (Foulsham & Underwood, 2011). Moreover, horizontal windows have fewer impacts

on eye movements than square or vertical windows (Foulsham, Teszka, et al., 2011, see Figure 1.4.3). Horizontal windows reflect attentional demands that were also found in reading with the asymmetric perceptual span. Additionally, the spatial offset of elliptical gaze-contingent windows (subtending 12.4° horizontally and 3.1° vertically and revealing about 4% of a scene) either to the left or to the right of the fixation location revealed an eye movement bias in the direction of the offset; that is, if the windows were offset to the left, then the fixation pattern was also offset to the left, and there were more leftward saccades as compared to natural viewing (Foulsham et al., 2013). This suggests that the eyes move towards available information supporting both the idea of active vision and the importance of pre-saccadic attentional shifts to program the next saccades in scene searches. These studies, cited above, used a hard border window with an opaque background that might have impacted oculomotor behaviour, as suggested by the last study presented.

Figure 1.4.3

Rectangular and square gaze-contingent windows used in scene viewing
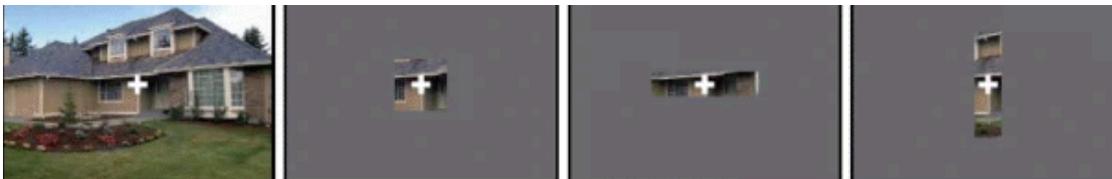
Note. Example of the windows used in the gaze-contingent study by Foulsham, Teszka, et al. (2011). Reprinted from Foulsham, Teszka, et al. (2011).

Malcolm et al. (2014) confirmed eye movement sampling strategies in scenes, where the eyes gathered the gist of the scenery first and then fine-grained details. Their study used a gaze-contingent moving window with a visual angle of about 5°, subtending roughly 7% of the scene. This window revealed that sampling from the centre was required to categorise scenes broadly. Interestingly, for more precise categorisation, eye movements sampled a wider area of the scene. This was especially evident when pictures of rooms had to be categorised with gaze distances from the centre of the scene increasing with the duration of fixations (see Fig-

ure 1.4.4).

Figure 1.4.4

Fixation patterns elicited by coarse and fine categorisation of scenes
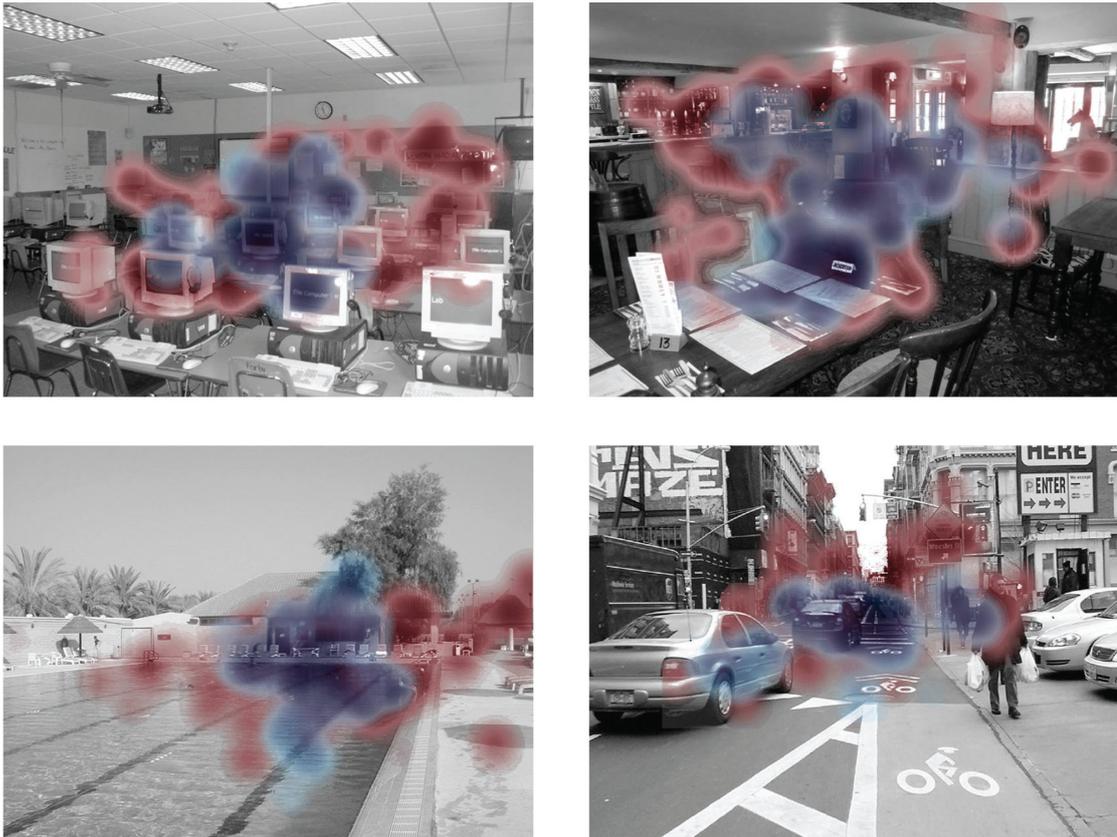
Note. Fixations elicited by participants according to a basic condition where they had to provide a broad categorisation of the scene (blue) and a subordinate condition where they had to provide a more precise categorisation of the scene (red). Overlapping areas between both conditions are depicted in purple. Reprinted from Malcolm et al. (2014).

Other studies using gaze-contingent paradigms evaluated the impact of the mask. The first study to use this paradigm was conducted by van Diepen et al. (1999), who employed line drawing pictures of real-world situations with objects and non-objects. Participants had to indicate whether a non-object was present in the scenery. The authors found that masking

foveal information with an ovoid subtending a 2.5° visual angle horizontally and a 1.9° visual angle vertically (covering about 2% of the scene) led to an increase in the duration of a fixation, with a noisy mask in particular revealing that this duration increases with processing difficulty.

Regarding the cultural effect on-scene search, it was found that both EAs and WCs were similarly impaired by masks with visual angles of 2°, 5° or 8° (subtending about 0.50%, 2% and 4.5%, respectively) as opposed to face processing. More particularly, a decrease in their performance was observed as a function of the mask size; that is, the larger the mask, the lower the performance. Furthermore, oculomotor behaviour was found to be identical across cultures. Interestingly, despite the masking of foveal information, participants tended to produce fixation towards targets as in natural viewing, whereas it would have been advantageous to process targets peripherally (Miellet et al., 2010). The absence of cultural bias was also observed in the previous section for information sampling during scene processing.

Overall, visual search requires at least 25% of the picture size to achieve natural eye movement behaviour and performance. The perceptual span for visual search might be asymmetric, as suggested by using a horizontal window. Additionally, studies have highlighted the importance of accessing low-level information in the peripheral vision to program the next saccades. However, surprisingly, the moving mask did not disturb oculomotor behaviour while impacting performance. Moreover, cultural bias was not found for scene processing for both information sampling and use. Altogether, these findings suggest a less focal sampling and thus processing of visual information for scenes than for faces. This fact could be explained by the nature of the stimuli being more complex and more contextualised than the stimuli in face processing studies.

1.4.3 INFERENCES ON SCENE PROCESSING FROM EYE MOVEMENT RESEARCH

Eye movement research has revealed that the global semantics of a scene is capture as early as the first fixation (*e.g.*, Potter, 1976). A first, coarse representation of a visual scene is thus produced from the first fixation at the beginning of scene gaze exploration. This representation does not

include details about the scene but only the nature of the scenery. Afterwards, the scene is explored according to bottom-up features (low-level characteristics), with fixations landing on salient areas and top-down features targeting the most informative regions to fulfil the task requirements. Thus, the representation will subsequently be more finely built according to the goal of the task and information feeding the visual system (Hegdé, 2008).

Despite the importance of foveation, viewed as the gathering of fine-grained information to be encoded in memory, extra-foveal processing seems especially essential in scene processing. First, it allows for the gathering of coarse information. More particularly, extra-foveal information is extracted during the pre-saccadic shift of attention during the saccade towards a target. This suggests a tight coupling of overt and covert attention during scene processing (Miellet et al., 2010)—at least, more coupled than in face processing and reading. From this view, it can be postulated that the size of the attentional spotlight and the shift in attention might vary according to the type of stimuli that observers are presented with. Then, and crucially, extra-foveal processing allows for the next saccades to be efficiently programmed towards the most informative regions in the stimuli as suggested by McConkie and Rayner (1975) and confirmed in scene processing by Foulsham, Teszka, et al. (2011).

1.5   Ecological Validity of Eye Movements in Laboratory Studies

Eye-tracking technologies have proven to provide valuable cues to the subtending process occurring during visual processing within a broad range of fields, including face perception, reading, and scene search. Although eye-tracking technologies have become precise, easy to use, and affordable, and have generated an increasing amount of research, most eye-tracking studies continue to use drawings, black-and-white photographs or impoverished representations of the real world rather than real-world stimuli. This is despite researchers, already several years ago, having raised the importance of using stimuli that are ecologically valid and as close as possible to the real world (Hayhoe & Ballard, 2005; Henderson, 2003).

The reasons for the choice to employ representations rather than real-world stimuli are

the cost of real-world ecologically valid experiments and the difficulty in controlling real-world variables impacting oculomotor behaviour. However, recent studies in face processing, for instance, have shown that the use of ecologically valid stimuli is important since eye movements differ between static (photographs) and dynamic (videos) stimuli (Richoz et al., 2018; Xiao et al., 2014; Xiao et al., 2015). This result highlights the importance of confirming the ecological validity of eye movement studies in laboratory settings using impoverished representations of the real world.

This Section first describes the methods of evaluating ecological validity. The differences between real-world and laboratory situations are subsequently described. Finally, the generalisability of the results obtained in the laboratory setting to the real world is discussed.

1.5.1 INVESTIGATION OF ECOLOGICAL VALIDITY

In the second half of the 20th century, researchers' desire to bring eye trackers into more ecologically valid settings put new constraints on the development of eye-tracking devices (Fawcett et al., 2015). However, more ecologically valid settings for eye-tracking studies were not possible until the advent of mobile eye trackers, which enabled participants to move their heads and bodies freely (Liversedge et al., 2011). Such devices contributed to the miniaturisation of eye-tracking technology. In the late 1940s, Hartridge and Thomson (1948) developed the first portable (head-mounted) eye tracker. However, it was rather bulky because of the size of the cameras recording the field of view and participants' eyes. Moreover, the setting hindered part of the participants' field of view (Fawcett et al., 2015). These constraints resulted in an impossibility to produce natural movement, and the tracker hence could not be used outside of the laboratory. Later, J. F. Mackworth and Mackworth (1958) improved portable eye tracking using techniques employed for television, thereby opening up an opportunity for researchers to conduct the first studies in the wild (N. H. Mackworth & Morandi, 1967; N. H. Mackworth & Thomas, 1962). Those studies aimed at evaluating the eye movements of drivers with a portable device that was still bulky. In the 1990s, Land (1992) developed a video-based and lightweight eye trackers allowing to move and behave seamlessly. This technology allowed researchers to

conduct studies on participants during various activities such as making tea (Land et al., 1999), playing table tennis (Land & Furneaux, 1997), and playing cricket (Land & McLeod, 2000). Since these landmark eye-tracking studies in the field of vision science, portable eye-tracking devices have been improved and used across a wide range of fields in the wild.

While portable eye-tracking technologies have become affordable, easier to use, and almost as comfortable as regular glasses, confined laboratory settings are mostly favoured over more ecologically valid possibilities. The rationale behind this preference lies in controls offered by laboratory settings, both on the stimuli and on the environment. However, it is necessary to ensure that the results obtained in the laboratory setting are ecologically valid and generalised to the real world. The video recorded by a mobile eye-tracking device in the wild can subsequently be presented in the confined space of the laboratory setting to address ecological validity. This allows researchers to assess the differences in oculomotor behaviour and behavioural performance between both conditions. However, this analysis is sound only if top-down and bottom-up processes are maintained equally. The latter process can be controlled and maintained constantly (*e.g.,* by using the same task in both conditions). The two settings should ideally be evaluated on the same participants, but this involves a second exposure to the stimuli and thus memory trace, which can influence oculomotor behaviour.

Two methods have been used to address the above mentioned first and second issues in the literature. The first method involves evaluating the same participants but on different tasks between both conditions (Foulsham, Walker, et al., 2011; Peterson et al., 2016), for example asking participants to buy something in real life and then asking the same participants to recognise their video recording in the laboratory (Foulsham, Walker, et al., 2011). While controlling for inter-individual variability, this method may introduce bias by using different tasks involving top-down processes to varying extents. The other method, while evaluating both conditions on the same task, uses different participants for the real-life versus laboratory conditions ('t Hart et al., 2009, see also the third contribution of this thesis). However, this method does not control for inter-individual variability, and such an experiment should thus favour a large sample of participants to minimise differences between groups.

In conclusion, investigation into the ecological validity of laboratory studies are far

from simple. Indeed, a comparison between laboratory and real-world settings results from a compromise on either bottom-up or top-down processes. However, even when both processes are maintained as close as possible in the two conditions, several differences remain, as described next.

1.5.2 COMPARISON BETWEEN EYE MOVEMENTS IN THE LABORATORY AND THE WILD

While investigating ecological validity is complicated and involves controlling both top-down and bottom-up processes, evaluating the differences between the laboratory setting and the wild remains crucial, since identifying these differences would allow one to estimate the generalisability of laboratory studies to the real world. The differences arise both from the setting, such as constrained head movements (*e.g.,* a chin rest or a bite bar could be used in the laboratory while in real life, the head can move freely) and from the situation in which eye movements are produced (2D static or dynamic stimulations in the laboratory against 3D in the real world). There is thus a disconnect between research investigating eye movements in the laboratory and during real-world activities (Foulsham, Walker, et al., 2011).

This disconnect arises from differences in the processes engaged in the wild versus in the laboratory setting. One substantial difference is that in real-world situations, individuals must engage several sub-processes, for example, to watch their steps to navigate or orientate their bodies to avoid obstacles and direct their gaze towards new targets (Lappi, 2015). In contrast, in laboratory settings, participants do not have to account for such uncertainties involved as they are in a controlled and protected environment, which reduces their arousal and vigilance compared to real-world settings.

Only a few studies have investigated the differences and similarities between the wild and the laboratory setting. Of those studies that have explored these discrepancies, the differences they found were mainly based on differences in eye movement dispersion in the field of view (see, *e.g.,* Foulsham, Walker, et al., 2011; 't Hart et al., 2009). More specifically, a central bias was observed in stimuli, likely due to different low-level features between the real-world

and laboratory settings. Indeed, it was found that saliency models poorly predict real-world eye movements, whereas they successfully predict eye movements in laboratory settings (Foulsham, Walker, et al., 2011). Additionally, observers differed in the proportion of their gaze directed towards a nearby path and distant objects. Walking observers directed their gaze towards near paths more often than watching observers, who more frequently fixated distant objects (see Figure 1.5.1). However, these differences could also be imputed to methodological issues involving different top-down processes, as some studies used different tasks between the laboratory and real-world conditions. Regarding the similarities, as suspected from the early

Figure 1.5.1

The proportion of fixation on items in the laboratory and in the wild
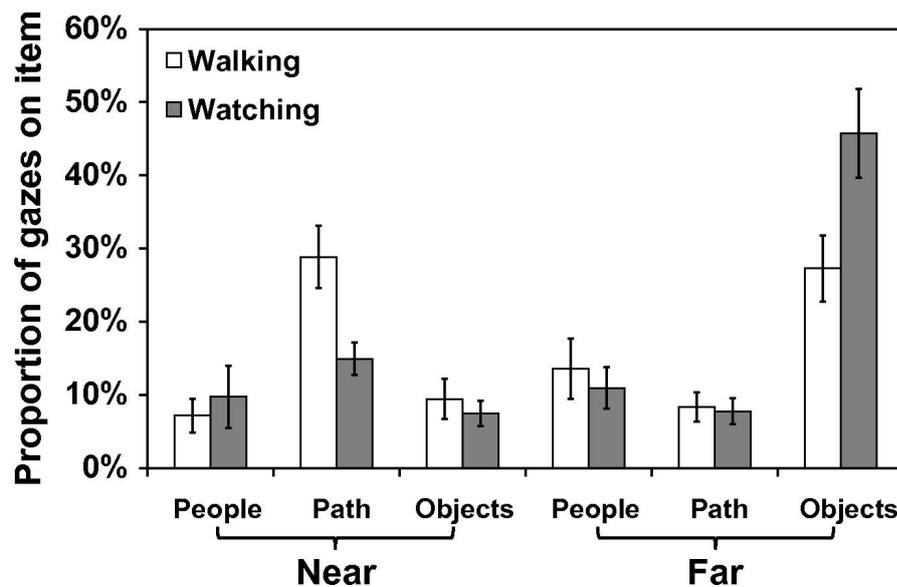
Note. Proportion of participants' gaze directed towards near or far people, paths or objects for walking and watching observers. Reprinted from Foulsham, Walker, et al. (2011).

work of Yarbus, it was found that social stimuli attracted participants' gaze to the same extent in both conditions. Furthermore, the triangular scan path observed in face processing was consistently found to occur in both conditions (Peterson et al., 2016). More precisely, we found that

other stimuli, such as readable or actionable stimuli (*e.g.*, doorknob) were also found to attract participants' gaze to the same extent in both conditions. Interestingly, the same proportion of observers' gaze was directed towards proximal persons and objects, both in the laboratory and in the real world (Foulsham, Walker, et al., 2011). The third contribution of this thesis sheds more light on such results.

In conclusion, similarities in the attraction of gaze towards social stimuli between the real-world and laboratory settings seem consistent across studies. However, there are no such consistencies for the differences between the two settings. These discrepancies might arise from uncontrolled top-down processes or different low-level information usage. Crucially, these differences likely arise from sub-tasks performed during the real-world condition, such as obstacle avoidance during walking (Lappi, 2015, 2016). Such differences in processing are challenging to control when evaluating the ecological validity of laboratory studies and they are further discussed in the next section.

1.5.3   THE GENERALISABILITY OF EYE-TRACKING STUDIES TO THE REAL WORLD

The generalisability of results from laboratory experiments to the wild is not as valid as researchers might hope. The validation of a specific behaviour requires that behaviour to be quantitatively and qualitatively of the same nature both in the wild and in the laboratory for a sound generalisation (Lappi, 2015). This implies cautiously controlling for top-down and bottom-up processes. However, studying the differences between the laboratory and the real-world condition also induces subtle differences arising from the use of an eye-tracking apparatus. Despite persisting differences in processes between both conditions, the complete generalisability of results would be possible only with a complete understanding of how our representations of the visual world are updated by the use of body, head and eye motions (Ballard, 1991; Tatler & Land, 2011). This will allow researchers to precisely model and account for the differences between the two experimental conditions using both portable eye tracking and desktop eye tracking. This section describes the two steps required to generalise eye tracking in the laboratory setting to the wild. The first step involves determining the remaining differences between

laboratory-based and real-world eye-tracking studies. The second step consists of evaluating the potential differences between eye-tracking studies in the real-world and natural behaviour in the wild (see Figure 1.5.2, for an illustration of these steps). These two steps are elaborated next.

Figure 1.5.2

Illustration of the approach to validate laboratory-based eye-tracking studies ecologically
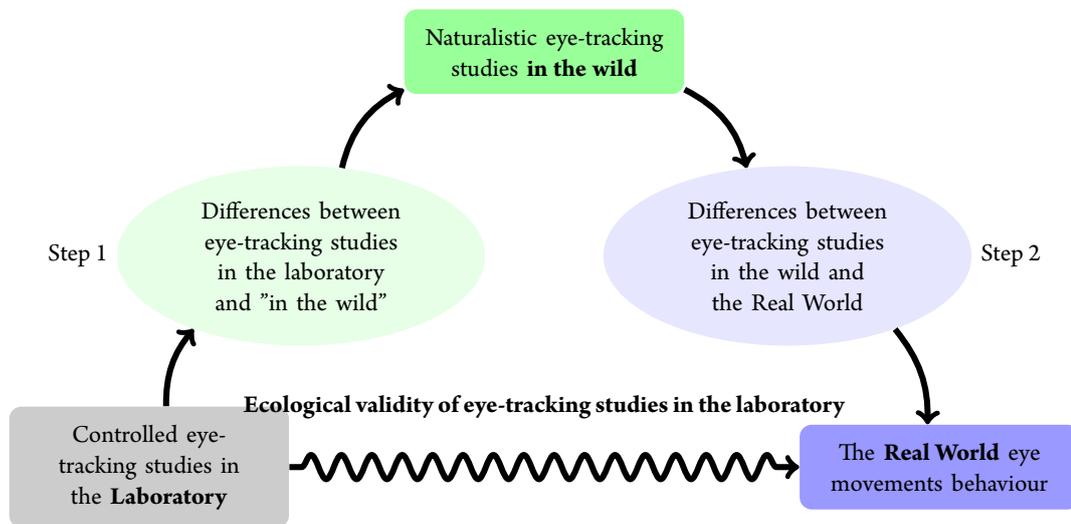
Note. Illustration of the steps to validate eye-tracking laboratory studies ecologically: from eye-tracking studies in the laboratory and naturalistic eye-tracking studies "in the wild" to eye movement behaviour in the real world. Inspired by Lappi (2016).

The first step to validate the generalisability of laboratory research to the wild is to compare eye-tracking studies in both the laboratory and the wild with ETG constraints. This was already described in the previous section. In brief, comparing eye-tracking laboratory studies to the real world with wearable eye-tracking devices reveals mainly differences in top-down processes (see Foulsham, Walker, et al., 2011; 't Hart et al., 2009). These differences are induced by sub-tasks conducted in the real world arising from the interaction between the observer and their environment (Lappi, 2015, 2016), such as watching the next step while walking.

Other differences arise from the reference surface. One's real-world gaze behaviour might not be classifiable similarly to that of eye-tracking studies in the laboratory. Indeed, directing one's gaze results from a complex interplay of head and eye motions and thus might not match a common typology. Additionally, laboratory studies often take monocular recordings, even though (as seen in Section 1.2.1) both eyes do not necessarily converge entirely; that is, there is no complete overlap between each VF.

Other crucial differences between eye-tracking studies in laboratory settings and those in the wild were found when analysing setting differences more formally. First, in a 3D scene, the head and body are maintained in a common frame of reference (FOR). As a result, the gaze location in a 3D scene accurately reflects its projection on a 2D reference surface represented by the display. In contrast, in the wild, the head and body are free to move. The gaze resulting from the combination of eye, head and body motions is characterised by the location of the point of regard (POR: motion of the eye) and the orientation of the VF (motion of the head and the body). In naturalistic eye-tracking studies, the gaze location is estimated by the projection of the point of fixation (POF) on the VF of the ETGs represented by the POR. The POR on the VF of the ETGs consequently does not accurately represent the POF (see Figure 1.5.3). Altogether, this evidence suggests differences in the typology of eye movements found between naturalistic eye-tracking studies and laboratory studies. However, comparing the typology of eye movements observed in both conditions remains a commonly used method.

Notable differences are induced by the recording of eye movements in a constrained laboratory setting compared to natural behaviour in the wild. Although portable eye-tracking devices minimise these differences between the laboratory and the real world, less apparent differences are induced by ETGs. First, even if a camera captures the field of view, it usually does not record the observer's entire VF. Binocular vision has an extent of about 180° horizontally and 150° vertically (Spector, 1990). In contrast, modern eye-tracking cameras capture eye movements at an average extent of 82,70° horizontally and 62,67° vertically (see Table 1.5.1).

This reduced gaze-tracking field, as compared to the VF of view, is accompanied by two

Figure 1.5.3

Depiction of the differences between eye-tracking studies in the laboratory and in the wild
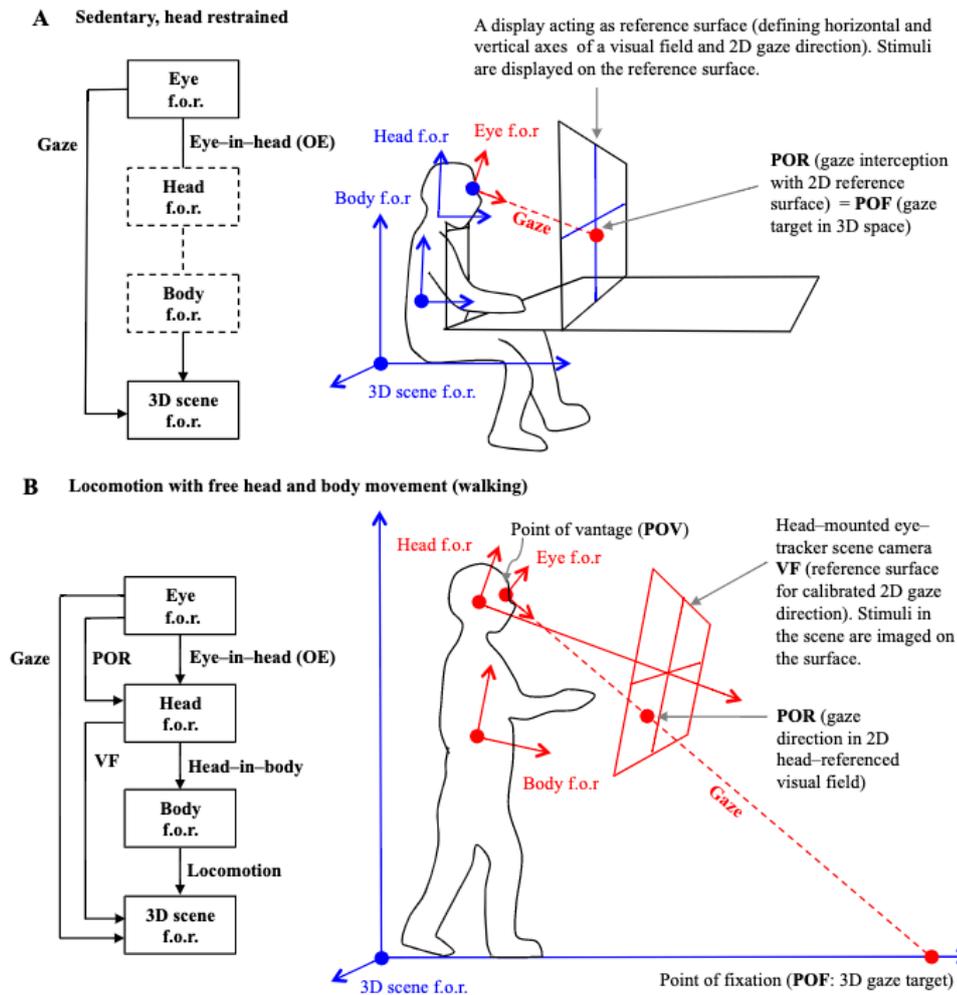
Note. Schematic depiction of eye-tracking studies in the laboratory and in the wild. Each different part is shown with its frame of reference (f.o.r.). The red colour indicates an f.o.r. that can move, and the blue colour indicates a fixed f.o.r. A) A typical laboratory setting with the head and body restrained. B) A typical setting during eye-tracking studies in the wild, where the head and body are free to move. The combined motion of the eye, head and body characterise the point of fixation (POR) and the orientation of the visual field (VF). Crucially, the point of fixation (POF) is projected on the eye-tracking VF but does not match the POF. Adapted from Lappi (2015).

Table 1.5.1

Range of eye-tracking fields observed for different wearable eye-tracker manufacturers

		Gaze-tracking Field (in degree of visual angle)	
		Horizontally	Vertically
Manufacturer	Product Name		
Arrington Research	BSU07	44	20
Ergoneers	Dikabilis	180	130
ISCAN	OmniView	90	60
SMI	Eye-tracking Glasses 2	80	60
SR Research	EyeLink II	20	18
Tobii Pro	Tobii Pro Glasses 2	82	52
Mean (SD)		82.67 (54.69)	56.67 (40.65)

Note. Size of the gaze-tracking field of different manufacturers' wearable eye-tracking devices in degree of visual angle. Average horizontal and vertical gaze-tracking fields and standard deviation are listed in the last row. Adapted from Cognolato et al. (2018).

problems. First, eye movements that are produced outside of the gaze-tracking field will not be recorded. Second, stimuli that could have been processed peripherally will not be rendered in the video as long as the head is not directed towards it. As such, the recording of eye movements in the real world does not precisely reflect natural oculomotor behaviour and the perception of the real-world environment. Another crucial point is that ETGs transpose the 3D world into a 2D plane. This transposition implies that the position of the eye recorded by the eye tracker and rendered on the video captured in the real world might not represent the exact real-world eye movement location. More specifically, the POR taken on the virtual plane does not match the POF in the wild (Lappi, 2015, 2016, see Figure 1.5.3). These device-related differences are considered when studying differences between real-world and laboratory settings. Finally, even though research has provided a large number of studies regarding how head, body and eyes motions are involved in the update of our representations of the visual world, we are far from a complete understanding of such a complex framework. For this reason, research should endeavour to minimise the differences between the wild and eye-tracking studies either in the laboratory or in the real world (see Figure 1.5.4 for a summary of the issues encountered in the validation of laboratory-based eye-tracking studies).

When researchers aim to generalise their results to the real world, they must ensure that their laboratory experiments are as close as possible to real-world conditions. Then, even if differences in the interplay of bottom-up and top-down processes between the wild and the laboratory persist, results should be generalised to the maximum extent offered by the laboratory setting. While ETGs minimise the differences between the laboratory and the wild, subtle differences remain. As such, eye-tracking studies are not fully generalisable to the real world. However, interestingly, when processing specific stimuli, the eye movement behaviour observed in the wild converges towards the behaviour observed in the laboratory. Thus, results from laboratory studies generalise for static stimuli processing but not for active tasks involving sub-tasks related to active behaviour.

Figure 1.5.4

Illustration of the issues encountered through the ecological validation of laboratory-based eye-tracking studies
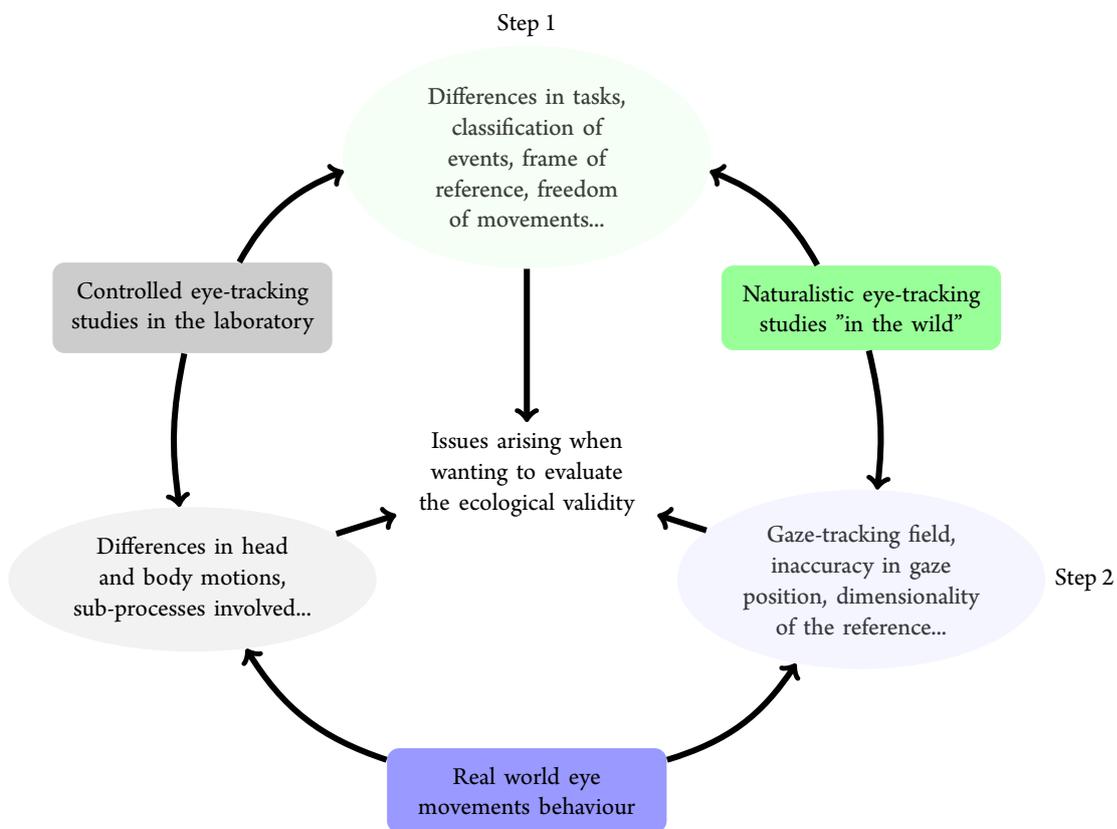
Note. Illustration of the issues encountered in the steps to validate eye-tracking studies ecologically: from eye-tracking studies in the laboratory, and naturalistic eye-tracking studies in the real world to real-world eye movement behaviour. Adapted from Lappi (2016).

2

Experimental Contribution

E YE movements have provided fruitful insight into the cognitive process during various
tasks. While the sampling of information regarding both faces and scenes has been ex-
tensively studied, information use, uncovered by the gaze-contingent paradigms, has been stud-
ied to a lower extent. Additionally, eye-tracking studies have predominantly been conducted in
confined laboratories; however, the ecological validity and the possibility to generalise labora-
tory results to the real world remain critical open questions.

Another critical question addressed approximately 50 years ago was the evaluation of
the minimal amount of information required at each fixation to reach a normal reading performance—
the perceptual span in reading—that was evaluated through the use of the gaze-contingent

paradigm (McConkie & Rayner, 1975). This led to a better understanding of attention and eye movements control in reading. Moreover, the modulation of the perceptual span in reading by several constraints or empowerment was thoroughly studied. Knowing the spatial extent of information intake at each fixation allowed researchers to build more appropriate models of reading. Following this idea, one of the goals of this thesis is to evaluate the perceptual span for faces—the *facespan*. While the facespan is a critical concept and might allow studies and models to be fine-tuned, studying facial information intake modulation under several conditions could also shed light on the face-processing system.

Modulation conditions can occur through multi-sensory integration, with a sensory modality enhancing or biasing another sensory modality. Crucially, auditory and visual events have been found to strongly interact, thus highlighting the importance of studying hearing loss (for a review of attention and multi-sensory integration, see, *e.g.*, Koelewijn et al., 2010). Studying hearing loss could hence shed light on visual processing in hearing people and the impact of hearing loss on the visual modality. Accordingly, another goal of this thesis is to study the modulation of information intake in the particular case of non-hearing people.

A growing amount of literature highlights the importance of studying dynamic stimuli in face-related research (see, *e.g.*, Richoz et al., 2018; Xiao et al., 2014; Xiao et al., 2015) and more generally, the importance of using ecologically valid stimuli (Hayhoe & Ballard, 2005; Henderson, 2003) to draw conclusions about the cognitive processes underlying eye movements. Researchers are currently working to achieve the foregoing goals exclusively in laboratory settings. Following these suggestions, the broader question of the ecological validity of eye-tracking studies was raised and assessed in an active navigation context. Indeed, previous studies did not reach a consensus on the ecological validity of eye tracking conducted in the laboratory. Studies lack control over both top-down and bottom-up processes.

The following sections briefly describe the three experimental contributions of this thesis. These contributions are included in Chapters 3, 4, and 5. Note that the scientific papers included in the aforementioned Chapters were further edited to correct for remaining grammar and spelling mistakes as well as to comply with the seventh edition of the American Psychological Association publication guidelines (American Psychological Association, 2020). Finally,

the main findings, their impact, and the possibilities to expand on this thesis are discussed in Chapter 6.

2.1   THE QUANTITY OF FACIAL INFORMATION USE

The first contribution, included in chapter 3, is an answer to the following question:

What is the minimal quantity of information processed at each fixation during face recognition?

As explained in Chapter 1, the information sampled by the eyes differs from the information used by the brain because of the visual acuity drop-off with eccentricity and the decoupling of visual attention. However, most research focuses on information sampling and not on information use. Therefore, the information actually used in various tasks remains to be clarified.

Concerning face processing, several studies have employed the gaze-contingent moving-window paradigm to identify the information used. For the most part, these studies evaluated the difference in the information used across participants and conditions. Crucially, in reading, the minimum amount of information required to achieve a normal reading speed was already identified more than 50 years ago. However, the minimum amount of information at each fixation that is necessary to achieve normal face recognition has not yet been identified. This is likely because the discrete stimuli represented by letters are more easily manipulated methodologically than the continuous information contained in face stimuli.

The present study utilised the gaze-contingent technique to restrict the available facial information and deliver target information into a fixed window, moving with eye movements. We ensured that the location of facial attributes was preserved, as a complete blockage of the

spatial location of facial attributes disrupts the programming of saccades. To do this, the information available in the opening of the moving-window gaze-contingent paradigm decayed gradually with eccentricity. In addition, the target facial information was masked by an average face allowing one to locate the position of the facial attributes. To identify the minimum amount of information required for normal facial recognition, we compared different aperture sizes to natural vision performance.

2.2   THE MODULATION OF THE QUANTITY OF FACIAL INFORMATION USED

The second contribution, included in chapter 4, is an answer to the following question:

Are facial information intake and use modulate by hearing loss?

As explained in Chapter 1, eye movements are directed through the integration of several sensory channels, such as hearing and vision, and interaction with the world. Therefore, the lack of a sensory modality makes it possible to evaluate the contribution of this sensory modality to a specific ability. Moreover, the deprivation of one modality sometimes results in the improvement of another, thanks to brain plasticity. For example, in the context of vision, it has been observed that hearing loss leads to greater extra-foveal sensitivity in non-hearing persons compared to hearing individuals.

Similar effects have been found in the treatment of faces. For example, De Heering et al. (2012) evaluated face recognition abilities in non-hearing and hearing participants and found a larger FIE in non-hearing participants but similar performances across groups. In contrast, H. He et al. (2016) found a reduced FIE in non-hearing participants. These inconsistent results must be clarified. Additionally, these studies focused on the sampled facial information and not the facial information used. Therefore, it remains to be clarified whether hearing loss results in different facial information use.

In this study, we also used a gaze-contingent moving-window paradigm. However, this time, the opening changed according to the duration of the fixation at a rate of 1° every 25 ms; that is the longer the duration of the fixation, the larger the aperture (Miellet et al., 2013). This paradigm enables researchers to analyse the duration of fixations and the amount of information required conjointly. As before, information on the location of facial attributes was preserved, thanks to an average face as a mask covering the target face. In addition, a smooth transition between the aperture and the mask meant that participants would not be disturbed by the gaze-contingent paradigm. Finally, this time, we used a delayed matching task by varying the orientation of the faces (upright vs inverted). This allowed us to study the effect of inversion. Indeed, despite previous studies investigating the FIE, they found either a reduced or an increased FIE in the non-hearing population. Thus, we also hope to clarify this issue.

2.3   THE ECOLOGICAL VALIDITY OF EYE-TRACKING STUDIES

The last contribution, presented in chapter 5, is an answer to the following question:

To what extent can the results of eye-tracking studies be generalised to the real world?

In vision-related research, several calls have been made for the scientific community to use ecologically valid stimuli as close to the real world as possible. However, to date, most vision research uses fixed, black-and-white stimuli (for low-level control purposes) that poorly reflect the vivid and colourful aspects of the real world. This raises the question as to whether the results found with such stimuli can be generalised to the real world.

Despite the few studies that have attempted to investigate this problem, the question remains open. Indeed, most of these studies suffer from methodological biases invoking differences between the real world and the laboratory, such as using two different tasks for these two conditions.

To overcome these problems, we carried out a study as close as possible to reality. We recorded the eye movements of two groups of participants. One group (the Walkers) had to walk in real life while their eye movements were recorded via ETGs. The other group (the Watchers) had to watch the recorded videos from the point of view of the Walkers, but in laboratory setting. In this way, the eye movements could be directly compared, and the differences and similarities between the two groups extracted. In addition, we applied a dynamic saliency algorithm to consider the temporal characteristic of the videos. Finally, this temporal dimension was further investigated using an algorithm to extract motion maps.

This page intentionally left blank.

Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise.

John Tukey

3

The Facespan—the perceptual span for face recognition

The Facespan—the perceptual span for face recognition

Michael Papinutto
Eye and Brain Mapping Laboratory (iBMLab),
Department of Psychology, University of Fribourg,
Fribourg, Switzerland

Junpeng Lao
Eye and Brain Mapping Laboratory (iBMLab),
Department of Psychology, University of Fribourg,
Fribourg, Switzerland

Meike Ramon
Eye and Brain Mapping Laboratory (iBMLab),
Department of Psychology, University of Fribourg,
Fribourg, Switzerland

Roberto Caldara
Eye and Brain Mapping Laboratory (iBMLab),
Department of Psychology, University of Fribourg,
Fribourg, Switzerland

Sébastien Miellet
Eye and Brain Mapping Laboratory (iBMLab),
Department of Psychology, University of Fribourg,
Fribourg, Switzerland
School of Psychology, University of Wollongong,
Wollongong, New South Wales, Australia
Bournemouth University,
Faculty of Science & Technology,
Cognition and Cognitive Neuroscience Research Centre,
Bournemouth, UK

3.1 Abstract

In reading, the perceptual span is a well-established concept that refers to the amount of information that can be read in a single fixation. Surprisingly, despite extensive empirical interest in determining the perceptual strategies deployed to process faces and an ongoing debate regarding the factors or mechanism(s) underlying efficient face processing, the perceptual span for faces – the facespan – remains undetermined. To address this issue, we applied the gaze-contingent Spotlight technique implemented in an old-new face recognition paradigm. This procedure allowed us to parametrically vary the amount of facial information available at a fixated location in order to determine the minimal aperture size at which face recognition performance plateaus. As expected, accuracy increased non-linearly with spotlight size apertures. Analyses of Structural Similarity comparing the available information during spotlight and natural viewing conditions indicate that the facespan—the minimum spatial extent of preserved facial information leading to comparable performance as in natural viewing—encompasses 7° of visual angle in our viewing conditions (size of the face stimulus: 15.6°; viewing distance: 70cm), which represents 45% of the face. The present findings provide a benchmark for future investigations that will address if and how the facespan is modulated by factors such as cultural, developmental, idiosyncratic or task-related differences.

3.2   INTRODUCTION

Face processing is a socially and biologically crucial feat achieved with high proficiency by the human visual system. An abundance of empirical studies has addressed various aspects of face processing, including the categorization of gender, age, race or facial expression, as well as face recognition and identification (*e.g.* Blais et al., 2008; Ebner et al., 2010; Kelly, Liu, et al., 2011; McClure, 2000; Meissner & Brigham, 2001). Isolating the very nature of the information used to achieve such visual categorizations is a major challenge in this field.

Recordings of oculomotor behavior have provided a valuable source of information regarding the relationship between the facial information sampled and observers' behavior. For instance, the robust and systematic triangular pattern of fixations exhibited during face perception (Henderson et al., 1990; Hsiao & Cottrell, 2008; Peterson & Eckstein, 2012; Van Belle, De Graef, Verfaillie, Rossion, et al., 2010; Walker-Smith et al., 1977; Xivry et al., 2008; Yarbus, 1967) was long considered invariant and universal. More recent investigations, however, have revealed that individuals' fixation patterns are affected by various factors, such as task demands and experience.

Preferential sampling of the eye region has been reported during face identity processing, whereas a shift towards lower face regions has been observed when the categorization of facial expression is of interest (*e.g.* Henderson et al., 2005; Malcolm et al., 2008). Such findings indicate a direct relationship between task demands and information diagnosticity. This notion receives further support from a recent study in which fixation patterns could be used to determine the task performed by observers' (Kanan et al., 2015), as well as findings obtained using the response classification technique "Bubbles" (Gosselin & Schyns, 2001; Schyns et al., 2002). In short, these studies demonstrate that sampled and used face information is consistently determined by the task performed. More fundamentally, this line of research indicates that , in cognitive terms, visual information should not be described purely in terms of low-level properties, but rather needs to be considered as diagnostic for a given task and observer.

Evidence suggesting experience-dependent modulation of oculomotor behavior stems

from both cross-cultural, as well as learning studies. For example, we have previously reported differential oculomotor patterns during face recognition as a function of the culture of the observers: East-Asian observers exhibit a central fixation bias, while and Western-Caucasian observers exhibit for an eye-mouth bias (Blais et al., 2008; Kelly, Liu, et al., 2011), both being consistently displayed despite stimulus inversion (Rodger et al., 2010). However, in studies using gaze-contingent techniques, we have demonstrated that despite using diverse gaze scan-paths, observers from both cultures rely on the same diagnostic features to perform face recognition (*i.e.*, the eyes and the mouth) with comparable levels of performance (Caldara et al., 2010; Miellet et al., 2012; Miellet et al., 2013). Thus, in both cultures, observers preferentially use different face information sampling strategies: Westerners favor a local strategy by sampling foveal information, whereas Easterners exhibit a global strategy to use the same diagnostic information. Finally, observers sample facial information differently from personally familiar and unfamiliar faces, with fixations expressed during processing of familiar individuals being relatively more distributed across all facial features (Van Belle, De Graef, Verfaillie, Rossion, et al., 2010).

While collectively these studies have enhanced our understanding of facial information sampling, they do not directly provide information regarding quantity of information use. For instance, observers may flexibly use information located within either the fovea, or extra-foveally during face identification (Miellet et al., 2011), depending on the landing position of their first fixation. Hence, the following important question remains unanswered: What is the quantity of information processed at each fixation during face recognition?

This question finds its analogue in the field of reading: "How much can be read in a single fixation?" (Woodworth, 1938, p.721).[1] McConkie and Rayner (1975) addressed this question of the perceptual span in reading using an elegant gaze-contingent moving-window paradigm. The authors reported prolonged reading times when the information outside of the window / fixation location, *i.e.* parafoveal information, was altered. Moreover, by parametrically varying the size of the gaze-contingent window, they were able to determine the minimal aperture size affording for normal reading performance as observed under unconstrained con-

---

[1]See also McConkie and Rayner, 1975, p.578, who improved upon the operationalization of this question by asking "How far into the periphery are specific aspects of the visual stimulus typically acquired and used during fixations in reading?"

ditions.[2] Recently, this approach has also been employed in the field of visual scene processing (Nuthmann, 2013).

A different but related question was addressed by Y. He et al. (2015) and Näsänen and Ojanpää (2004) who investigated the number of faces that can be recognized in a single fixation during a search task. Their research question differs from ours in that Y. He et al. (2015) and Näsänen and Ojanpää (2004) used the number of faces as metric of the visual span for face recognition, whereas we considered the information within a single face during recognition for the facespan.

In the present study we sought to determine the currently unknown perceptual span for faces – the facespan. Following the aforementioned reasoning, we aimed to specify the spatial extent across which facial information is accrued within a single fixation during face recognition. In other words we asked: What is the minimum quantity (in terms of spatial extent) of information needed at a fixation to achieve normal face recognition performance? To this end we employed a gaze-contingent moving-window technique, adapted for face stimuli, presented in the context of an old-new face recognition task. The facial information available to observers was restricted by using the Spotlight technique (for details see *e.g.*, Caldara et al., 2010; Miellet et al., 2013) with Gaussian apertures dynamically centred on their fixations, the size of which varied parametrically. Our goal was to ascertain the spotlight size that would allow subjects to reach a performance similar to that observed during natural face viewing. We reasoned that benchmarking the facespan in this manner would provide the basis for investigating how information intake is impacted by factors such as those related to the stimulus, observer, or task-dependency.

---

[2]For English the perceptual span extends from 3 characters to the left of fixation (approximately the beginning of the fixated word) to 14 characters to the right of fixation (McConkie & Rayner, 1975). Importantly in active reading, involving eye-movements, the span is mainly constrained by higher-level processing limitations rather than low-level visual constraints (Miellet et al., 2009). Note that a number of studies (Kwon & Legge, 2012; Legge et al., 2001; D. Pelli et al., 2007; Yu et al., 2007) support the notion of a visual span largely limited by low level processing. However, crucially, these studies used rapid presentations and identification of meaningless trigrams in absence of eye-movements (Rapid Serial Visual Presentation). The processes involved might have been different from those present in active reading of meaningful sentences or texts, for instance visual remapping and contextual influence. Thus these studies are likely to underestimate the higher-level influences in play in natural reading while overestimating the low-level influences. The concept of visual span presented in those studies might differ in nature from the perceptual span measured in active reading of sentences (see Legge, 2006).

3.3 Methods

3.3.1 Participants

Two hundred and twenty young adults (173 females, M=21.47, SD=2.50) from the University of Fribourg, Switzerland, participated in this study. All participants had normal or corrected vision, were Western-Caucasian, and received course credit for participation. All participants gave informed consent; the protocol was approved by the ethical committees of the Department of Psychology of Fribourg University, Switzerland.

3.3.2 Materials

Stimuli were obtained from the Karolinska Directed Emotional Faces (KDEF, Lundqvist et al., 1998) and the Asian Face Image Database (AFID, Bang et al., 2001) and consisted of 84 Eastern-Asians and 84 Western-Caucasians identities containing equal numbers of males and females. The images were 688 pixels in size vertically and 702 pixels horizontally, subtending 15.3° and 15.6° of visual angle, respectively. Face stimuli were aligned with respect to the eye and mouth positions, normalized for luminance, and no distinctive external features or facial hair. Images were viewed at a distance of 70 cm, which is representative of a natural distance during human interactions (Hall, 1966), and were presented on a 1920x1080 pixels grey background displayed on a ViewPixx/3D LCD monitor (120 Hz refresh rate).

3.3.3 Apparatus

Eye movements were recorded at a sampling rate of 1000 Hz with the SR Research Desktop-Mount EyeLink 2K eye tracker (with a chin and forehead rest), which has an average gaze position error of about .5° and a spatial resolution of .01°. The eye tracker had a linear output over

the range of the monitor used. Although viewing was binocular, only the dominant eye was tracked. The experiment was implemented in Matlab (R2009b, The MathWorks, Natick, MA, USA), using the Psychophysics toolbox (PTB-3, Kleiner et al., 2007; D. Pelli, 1997) and Eye-Link Toolbox extensions (Cornelissen et al., 2002; Kleiner et al., 2007). Calibrations of eye fixations were conducted at the beginning of the experiment using a nine-point fixation procedure as implemented in the EyeLink API (see EyeLink Manual) and using Matlab software. Afterwards, calibrations were validated with the EyeLink software and repeated when necessary until reaching an optimal calibration criterion. At the beginning of each trial, participants were instructed to fixate a cross at the centre of the screen to perform a drift correction. If this exceeded 1° a new calibration was performed to ensure optimal recording quality.

3.3.4 Spotlight

We used 11 spotlight aperture sizes from 9° to 19° of visual angle by step of 1° (see Figure 3.4.1A). These apertures were centred dynamically on the observers' fixations. The target face was presented at fixation location and was progressively masked by an average face template with retinal eccentricity. More specifically, at the centre of the spotlight, the alpha channel had a value of zero, corresponding to complete transparency of the average template and full access to the target face. The alpha channel value increased with distance from the fixation location according to a Gaussian function. For each spotlight, the standard deviation of the corresponding Gaussian mask was 10.73% of the spotlight size. Values below .004 were set to 0, corresponding to complete opacity of the average template and no information from the target face. We used progressive masking to avoid both extra-foveally attracting participants' attention and abruptly dividing facial features by a hard aperture border. However, this progressive masking also involves challenges in terms of quantifying the information preserved within the Gaussian aperture. In the section 3.4.3 "Data-driven reconstruction of the facespan" we discuss how we overcame these challenges by introducing a measurement of the facespan that goes beyond the mere diameter of the gaze-contingent window. This was done by taking into account acuity drop-off, fixation locations and similarity between the target face and the average template. The information outwith the spotlight was an average face composed from all stimuli from databases. This average face did not provide any useful information for recognition, but allowed observers to program

natural saccades. In terms of gaze-contingent display updating, it took 1 ms to receive a sample from the eye tracker, less than 1 ms to calculate texture including the background and the Gaussian mask, and 2-57 ms to refresh the screen. Thus, the display was updated on average every 9 ms (between 3 and 58 ms), which eliminated any subjective impression of flickering.

3.3.5   Procedure

Each observer performed both natural viewing and spotlight conditions. For the spotlight condition, each observer was randomly assigned one of the eleven spotlight sizes. Participants started with the calibration procedure described in the apparatus section (section 3.3.3). They were then presented with a training session to familiarize them with the gaze-contingent moving-window display and informed about the experimental procedure. This consisted of two blocks (natural viewing, and spotlight condition), each involving presentation of a series of faces to be learned and subsequently recognized. Each of the blocks was divided into two sub-blocks (one per ethnicity – Western-Caucasian and East-Asian), resulting in a total of 4 sub-blocks. In each sub-block participants had to learn 14 face identities (7 females), which randomly displayed neutral, happy or disgusted facial expressions. After a 30 seconds break, they were presented with a series of 28 faces (14 faces from the learning phase and 14 new faces; 14 females) and instructed to indicate as quickly and as accurately as possible whether each face had already been presented in the learning phase or not by pressing keys on the keyboard with the index of their left and right hand. Reaction times and accuracy were collected and analysed for the purpose of this experiment; stimulus presentation duration was terminated by subjects' responses.

Facial expressions differed between learning and recognition phases in order to ensure identity learning, rather than trivial image matching. Each trial started with the presentation of a central fixation cross allowing the experimenter to check that the calibration was still accurate (toleration of an error of a maximum of .5 degrees of visual angle). If the calibration was not sufficiently accurate, a new one was performed. Therefore, calibration was validated prior to each trial. This calibration checking was followed by a final central fixation cross used as the drift correction. Afterwards, a face was presented at a random location on the computer screen

in order to avoid anticipatory oculomotor strategies.

3.3.6 Data Analyses

Behavioural performance measures included reaction times (RT) and A-Prime (A, Stanislaw & Todorov, 1999). Participants whose face recognition performance did not exceed chance level were discarded (14% of the sample tested; significance threshold determined at 61% using a bootstrapping approach). Moreover, trials for which participants exhibited RT larger than 2.5 standard deviations of their average RT were considered outliers and discarded from analysis. Individuals Cohen's d effect sizes (Cohen, 1988) were then calculated to assess the impact of each spotlight size on performance.

Saccades and fixations were determined based on angular velocity using a custom algorithm. Individual saccade velocity thresholds were set by expert experimenters and ranged between 30 and 80°/s (M=40.16, SD=9.20). When the velocity threshold exceeded 100°/s, participants were discarded (cf. Holmqvist et al., 2011) as having noisy eye-movement data (15% of the sample tested). Fixations that were spatially (<0.3°) and temporally (<20ms) too close were merged. Previous studies have not revealed any effect of the experimental phase (learning vs. recognition), correctness of the answer (correct vs. incorrect trials), or race of face stimuli ((Western-Caucasians vs. Eastern-Asians) Blais et al., 2008; Caldara & Miellet, 2011; Miellet et al., 2012). In light of these findings and to ensure compatibility with the common procedure in the field, only the eye-movements of trials with correct responses were analysed and data from Eastern-Asian and Western-Caucasian face stimuli were collapsed (see Appendix, Figure 3.9.1 for demonstration of the lack of effect of stimulus race). Variables describing the general oculomotor behaviour (number of fixations and fixation durations) were computed. Statistical fixation maps were computed with the *iMap* toolbox (version 4.01, Caldara & Miellet, 2011; Lao et al., 2017; Miellet et al., 2014). *iMap4* performed pixel-wise Linear Mixed Models (LMM) on the smoothed fixation maps with subject as a random effect, and condition (spotlight vs. natural viewing) as a fixed effect. A spatial cluster test based on bootstrapping was then used to assess statistical significance of the linear contrasts between the natural viewing and spotlight

conditions. The trimmed mean of beta values i.e. the model coefficients of the fixed effect (in this case they are the conditional mean difference of fixation duration in seconds) within the face region in the fixation maps was then extracted for each spotlight size in order to assess the impact of size on the distribution of fixations.

Finally, stimuli from the candidate spotlight size (smallest spotlight size leading to performance similar to natural viewing) were compared to natural viewing stimuli. The goal here was to assess how much information was preserved in the spotlight condition compared to natural viewing when considering observers' fixations. Indeed, it is not straightforward to infer the perceptual span for face recognition based on our gaze-contingent manipulation. We needed to quantify how much information (in terms of spatial extent) in the critical spotlight was identical to the natural viewing condition. In order to address this question, we needed to consider four challenges. First, the spotlight is a Gaussian aperture so the target and average faces information blend progressively. Second, visual acuity drops off with retinal eccentricity. Thus, small extra-foveal variability in high spatial frequencies between the spotlight and natural viewing stimuli might not be captured by the visual system. Third, the similarity between spotlight and natural viewing stimuli might depend on how dissimilar a given target face (presented centrally) is from the average face (displayed extra-foveally). Finally, the similarity between spotlight and natural viewing stimuli might depend on the actual fixation location.

To address these challenges, we used a procedure that aimed at mimicking early constraints of the visual system (e.g., acuity drop-off with eccentricity), while also considering targets' typicality (how similar to the average face a specific target face is), as well as fixation locations. First, for each participant and stimulus we convolved a retinal filter (Targino Da Costa & Do, 2014) on the spotlight of interest and natural viewing stimuli according to fixation locations in the critical spotlight condition (see Figure 3.4.2 for the facespan reconstruction pipeline). The retinal filter parameters were distance to the screen (700 mm in our viewing conditions), stimulus size in pixels and the lossy parameter ($\Delta = 25$, chosen to include visual perceptual lost according to human sight). We then used the *Structural SIMilarity index* (SSIM, Wang et al., 2004) to quantify, independently for each target face, the similarity between spotlight and natural viewing stimuli after retinal filtering at each of the fixation locations in the critical spotlight condition. The *SSIM* uses luminance, contrast and structure of two images in order to

assess similarity of both images pixel by pixel. In the next step, we used the *pixel-test* (Chauvin et al., 2005) to assess significance in the *SSIM* maps corresponding to each target face and fixation location. The *pixel-test* (Chauvin et al., 2005), which is based on the *Random Field Theory* (RFT), allowed to highlight the information that was significantly preserved between spotlight and natural viewing conditions (pixels that were significantly most similar). We used the following parameters for the *pixel-test* as recommended by Chauvin et al. (2005): *sigma*=20, cluster test *threshold*=2.7, *p*=.05. The Random Field Theory provides the probability of observing a cluster of pixels exceeding a threshold in a smooth Gaussian random field while taking into account the spatial correlation inherent to the data set. More details about *SSIM* and the *pixel-test* can be found in Wang et al. (2004) and Chauvin et al. (2005), respectively. We finally centred and averaged, across stimuli, observers and fixations, the areas of information preserved by the critical spotlight manipulation (significant *SSIM* according to pixel-test after retinal filtering on spotlight and natural viewing stimuli). The rationale of this approach is that the amount of preserved information, corresponding to the minimal spotlight from which additional information does not improve performance, is the facespan.

3.4 RESULTS

3.4.1 BEHAVIOUR

First, we investigated the individual effect sizes (Cohen's d) of spotlight compared to natural viewing on performance. Large values indicate that, for a given participant performance advantage (in terms of A-prime) for natural viewing as compared to the spotlight condition is larger than the variability between trials irrespective of condition. In contrast, small effect sizes indicate that the difference between viewing conditions for a given participant is not much larger (or smaller for effect sizes < 1) than the variability between trials across conditions. An effect size which does not significantly differ from 0 indicates that performance does not different across viewing conditions.

We observed that effect sizes decreased parametrically with increasing spotlight size. However, spotlights always impacted on performance regardless of their size. A-posteriori analyses revealed that log-average luminance was higher for stimuli in the natural viewing, compared to spotlight conditions (Natural Viewing= .27; Spotlight= .39). In contrast, spotlight conditions did not differ from each other in terms of luminance (M= .39, SD= .002). The difference in luminance between natural viewing and spotlight conditions might be due to our use of the alpha-blending function from the Psychotoolbox (Screen('BlendFunction')). Thus, the fact that the effect of spotlight on performance was always significant might arise from differences in luminance between spotlight and natural viewing conditions. Importantly, the critical spotlight that we wanted to identify does not necessarily correspond to an identical performance between natural viewing and spotlight condition. It rather corresponds to the spotlight size from which additional information available in larger spotlights does not further improve performance. Performance data was fitted with various models from the simplest linear regression to sigmoidal or polynomial models. A 2 degree polynomial model offered the best fit (R^2= .94, $F(8,11)$= 60, p< .001) with the effect size of RT depending on spotlight size (Figure 3.9.2). The effect sizes of A' were best described by a sigmoidal model (R^2= .79, $F(7,11)$= 126, p< .001).

The first plateau of the sigmoidal model indicates that the effect sizes for A' remained constant for spotlight sized 9° to 14°, indicating that until 14° additional information does not noticeably increase performance. The second plateau of the sigmoidal model shows smaller effect sizes for performance (A'). Crucially, this second plateau also indicates that, from a 17° spotlight size onward, additional information available in larger spotlights did not further improve performance.

3.4.2 EYE MOVEMENTS

Contrasts between the spotlight and the natural viewing conditions performed individually for each spotlight size on the number of fixations and fixation durations (paired t-tests, Bonferroni corrected) did not reveal any consistent differences. In contrast, differential fixation maps generated by *iMap4* (Lao et al., 2017) revealed differences in the spatial distribution of fixation

Figure 3.4.1

Recapitulation of the facespan gaze-contingent window sizes, impact of those windows on the behaviour and the fixation patterns

A

B

C
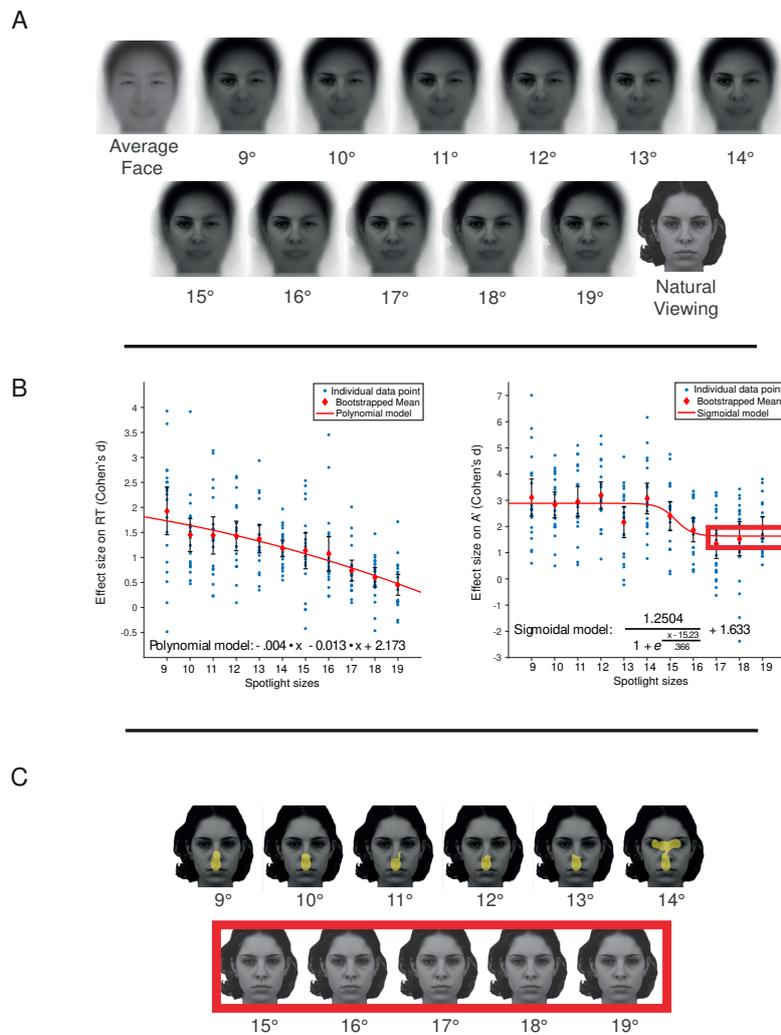
Note. (A) The average face used as mask; an example stimulus with different spotlight sizes centered on the left eye and under natural viewing. (B) Impact of spotlights on RT (left) and A' (right). The performance plateau in accuracy is highlighted by the red rectangle. (C) Effect of spotlights on fixation pattern. Highlighted in yellow are the areas fixated longer for the spotlight compared to natural viewing condition. The rectangle highlights the absence of a significant effect of spotlight (compared to natural viewing) on fixation patterns.

durations between the spotlight and natural viewing conditions for spotlight sizes ranging from 9 to 14°. This range of spotlights exhibits areas with significantly longer fixation durations in the spotlight than in the natural viewing condition. Crucially *iMap4* does not show any significant differences across the whole stimulus space between spotlight and natural viewing conditions for spotlights of or larger than 15° (Figure 3.4.1 C).

3.4.3   DATA-DRIVEN RECONSTRUCTION OF THE FACESPAN

As noted above, eye movements exhibited during natural viewing and spotlight conditions did not differ for spotlights equal to or larger than 15°. Moreover, A' indicated that facial information contained by spotlights larger than a 17° Gaussian aperture did not lead improved performance. In short, from eye movement patterns and performance, it appears that, on average, face information outside of a 17° Gaussian aperture around the fixation location is not used for face recognition as tested in the present experiment. As described in the data analyses in section 3.3.6, we quantified how much information (in terms of spatial extent) in the spotlight was identical to the natural viewing condition, for a 17° spotlight. The average area of preserved information was contained in a circle of 7° degree of visual angle.[3]

In summary, performance and fixation patterns revealed that face information outside a 17° Gaussian aperture is not used for face recognition. *SSIM* after retinal filtering and *RFT* indicate that face information projecting in the central 7° of the 17° Gaussian aperture was preserved compared to natural viewing. Thus, we conclude that the average perceptual span for face recognition (as tested in the current experiment), the facespan, is 7° of visual angle given the viewing conditions used in our paradigm.

---

[3]Note that the facespan estimate is robust to systematic manipulations of the pixel-test parameters, spanning the values found in the literature. The parameters manipulation had a minimal influence on the estimated facespan whose values ranged from 6.48 to 7.29 degrees of visual angle (42 to 47% of the face)

Figure 3.4.2

Pipeline of the facespan reconstruction
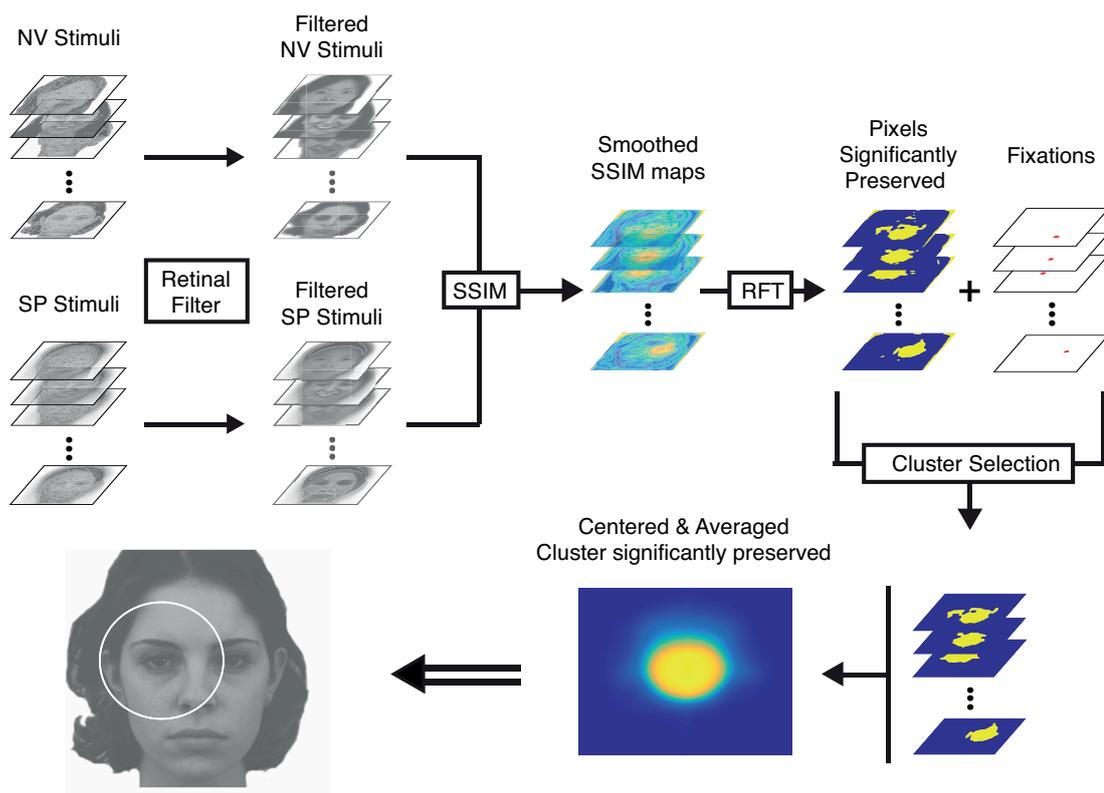
Facespan: 7° / 45% of the face

Note. Pipeline of the facespan reconstruction. On the top left, stimuli from recognition phases for both spotlight and natural viewing were passed through a retinal filter. To assess their similarity, *SSIM* was used and resulted in *SSIM* maps. Afterwards, the pixel-test *RFT* was used to assess the area significantly preserved from natural viewing in the 17° spotlight. Clusters of interest were then selected, centred and averaged. The averaged cluster significantly preserved was estimated to contained 7° of information, representing the facespan.

3.5   DISCUSSION

In this study we aimed to determine the perceptual span for faces – the facespan. To this end we employed a gaze-contingent paradigm in which the quantity of available information on the target face—the spotlight or aperture size—varied parametrically. Importantly, we reconstructed the available information inside these apertures in a data-driven fashion. This allowed us to quantify the extent of information preserved in the spotlight compared to natural viewing conditions, while taking into account the spotlight size, acuity drop-off, fixation locations and dissimilarity between the target faces and the average face template.

With regards to observers' performance and oculomotor behavior during the old/new recognition task, our findings suggest that the facespan – the information extent required at each fixation for normal face recognition – is 7° of visual angle given our experimental condition. For a fixation located on the nose, this perceptual span covers almost the entire area of internal features; for a fixation positioned on a different feature, it encompasses that feature along with its surrounding information (for visual representation see Figure 3.4.2). Another way of expressing the facespan is in terms of percentage of the face size, which in this study represents 45% of the face. Interestingly this value is very similar to that reported by Kwon et al. (2016). In a study manipulating spatial-frequency content and gaze-contingent window size, these authors found that at least 50% of facial information was required for observers to exhibit efficient face recognition performance. The similarity between the percentage of necessary information reported by Kwon et al. (2016) and the facespan we measured here is remarkable especially considering the numerous differences between both studies: face sizes (4° and 2° vs. 15.5° respectively), tasks (identification vs. recognition), familiarity (famous vs. unfamiliar faces) and presence of extra-foveal information (homogeneous mask vs. average face template mask). The most critical difference between both studies is that Kwon et al. (2016) measured the revealed information across several fixations (*cf.* Miellet et al., 2013), whereas the facespan represents the information necessary at a single fixation.

Concerning observers' oculomotor behavior, fixation patterns varied as a function of aperture size for spotlights of 9° - 14° as demonstrated in Figure 3.4.1C; beyond this aperture

size, no further changes were observed. This is consistent with the idea that access to face information from a wider visually central region progressively allows for a more typical oculomotor behavior (see also Miellet et al., 2012). Interestingly, subjects' performance increased once fixation patterns were stable, supporting the idea that typical eye-movements are crucial to achieve optimal face recognition (Henderson et al., 2005). Finally, performance plateaued once at least an individual feature was visible at each fixation.

With a face displayed at a distance used during natural human interaction (about 80 cm), a facespan of 7° allows the covering of the entire face within three fixations that would be directed towards these facial features. The perceptual span for face recognition is in line with the perceptual span for reading (McConkie & Rayner, 1975) and for visual scene search (Nuthmann, 2013). Indeed, the facespan falls between these two perceptual spans (the perceptual span for reading: 4.5°; the perceptual span for face recognition: 7°; the perceptual span for scene search: 8°). This observation could be explained in terms of the density of to-be-processed information under each of those tasks, which is highest for text and lowest for scenes. This idea, that the perceptual span depends on information density, is supported by findings in reading. For instance, in Chinese where each character carries more linguistic information than in English, the perceptual span is much smaller (1 character to the left of the fixated character and 3 characters to its right according to Inhoff & Liu, 1998). Using the gaze-contingent parafoveal magnification technique, Miellet et al. (2009) demonstrated that the perceptual span is more accurately defined in terms of number of characters (quantity of linguistic information), rather than visual angle. This hypothesis is also consistent with evidence showing that the perceptual span for reading is modulated by text difficulty (Henderson et al., 1990), as well as reader characteristics, such as expertise (Rayner, 1986; Rayner et al., 2010), age (Rayner, 1986), and disorder (Rayner, 1983). From this perspective, information is not considered in absolute terms but rather follows the concept of diagnostic information and thus varies depending on viewing conditions, strategies and task at hand (see, for instance, Oliva & Schyns, 1997).

Following this idea, we emphasize that the facespan should not be considered as an absolute quantity. Inasmuch as the perceptual span for reading is not absolute, but instead flexible, the facespan reported here should be considered as an average benchmark obtained under the aforementioned specific viewing conditions and task. Mirroring previous observations for the

perceptual span in reading, we suggest that the facespan may vary as a function of factors including, but not limited to e.g. stimulus or observer characteristics, and task demands. Indeed, previous research in face processing has demonstrated cultural (Blais et al., 2008; Caldara et al., 2010; Miellet et al., 2012; Miellet et al., 2013) , idiosyncratic (Kanan et al., 2015; Mehoudar et al., 2014; Peterson & Eckstein, 2013; Ramon et al., 2016) and familiarity (Van Belle, De Graef, Verfaillie, Rossion, et al., 2010) biases in information sampling. Moreover, initial fixation location on the face also impacts on sampling strategy (Miellet et al., 2011). These findings suggest that the facespan might also be flexibly modulated by such factors. However, further studies are necessary to understand precisely when and how the facespan is modulated.

We suggest that characterizing the facespan at the individual level, as well as across cultures may broaden our understanding of these well-established, systematic differences in perceptual biases and processing. Specifically, we propose a direct relationship between individual preferences in information sampling (global vs. local) and an individual's facespan, which in turn determines the facial representation generated based on the visual input. In other words, we propose that observers exhibiting global sampling strategies (*i.e.* favoring fixations on the center of the face) will exhibit a broader facespan, which allows to sample diagnostic feature (eyes and mouth) from central fixation. In contrast, observers exhibiting local sampling strategies (*i.e.* favoring fixations on eyes and mouth) will show a spatially more constrained facespan. As demonstrated in Miellet et al. (2013), the sampling strategies determines the information available to the brain.

3.6   CONCLUSION

In the present study, we determined the perceptual span for face recognition – the facespan. Our data shows that given our viewing conditions, observers require information encompassing 7° of visual angle to exhibit face recognition performance comparable to that observed during natural viewing. Importantly, similarly to the perceptual span in reading, the facespan should not be considered as fixed, but rather flexible. Our findings provide a benchmark for information sampling during face processing, which will leverage further investigations into intra- and

inter-individual differences and how the perceptual span for face processing is modulated by constraints such as culture, development or task.

3.7   ACKNOWLEDGEMENTS

3.8   REFERENCES

Bang, S., Kim, D., & Choi, S. (2001). Asian face image database.

Blais, C., Jack, R. E., Scheepers, C., Fiset, D., & Caldara, R. (2008). Culture shapes how we look at faces. *PLoS ONE, 3*(8), e3022. https://doi.org/10.1371/journal.pone.0003022

Caldara, R., & Miellet, S. (2011). Imap: A novel method for statistical fixation mapping of eye movement data. *Behavior Research Methods, 43*(3), 864–878. https://doi.org/10.3758/s13428-011-0092-x.

Caldara, R., Zhou, X., & Miellet, S. (2010). Putting culture under the "spotlight" reveals universal information use for face recognition. *PLoS ONE, 5*(3), e9708. https://doi.org/10.1371/journal.pone.0009708.

Chauvin, A., Worsley, K. J., Schyns, P. G., Arguin, M., & Gosselin, F. (2005). Accurate statistical tests for smooth classification images. *Journal of Vision, 5*(9), 659–667. https://doi.org/10.1167/5.9.1.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Erlbaum.

Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The eyelink toolbox: Eye tracking with matlab and the psychophysics toolbox. *Behavior Research Methods, Instruments, & Computers, 34*(4), 613–617. https://doi.org/10.3758/BF03195489

Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). Faces—a database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior Research Methods, 42*(1), 351–362. https://doi.org/10.3758/BRM.42.1.351.

Gosselin, F., & Schyns, P. G. (2001). Bubbles: A technique to reveal the use of information in recognition tasks. *Vision Research, 41*(17), 2261–2271. https://doi.org/10.1016/S0042-6989(01)00097-9

Hall, E. T. (1966). *The hidden dimension*. Doubleday. https://doi.org/10.1017/CBO9781107415324.004.

He, Y., Scholz, J. M., Gage, R., Kallie, C. S., Liu, T., & Legge, G. E. (2015). Comparing the visual spans for faces and letters. *Journal of Vision, 15*(8), 1–13. https://doi.org/10.1167/15.8.7.

Henderson, J. M., Ferreira, F., Rayner, K., & Pollatsek, A. (1990). Effects of foveal processing difficulty on the perceptual span in reading: Implications for attention and eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(3), 41–429. https://doi.org/10.1037/0278-7393.16.3.417.

Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. *Memory & Cognition, 33*(1), 98–106. https://doi.org/10.3758/BF03195300.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures.* OUP Oxford.

Hsiao, J. H., & Cottrell, G. (2008). Two fixations suffice in face recognition. *Psychological Science, 19*(10), 998–1006. https://doi.org/10.1111/j.1467-9280.2008.02191.x.

Inhoff, A. W., & Liu, W. (1998). The perceptual span and oculomotor activity during the reading of chinese sentences. *Journal of Experimental Psychology: Human Perception and Performance, 24*(1), 20–34. https://doi.org/10.1037/0096-1523.24.1.20

Kanan, C., Bseiso, D. N. F., Ray, N. A., Hsiao, J. H., & Cottrell, G. W. (2015). Humans have idiosyncratic and task-specific scanpaths for judging faces. *Vision Research, 108,* 67–76. https://doi.org/10.1016/j.visres.2015.01.013.

Kelly, D. J., Liu, S., Rodger, H., Miellet, S., Ge, L., & Caldara, R. (2011). Developing cultural differences in face processing. *Developmental Science, 14*(5), 1176–1184. https://doi.org/10.1111/j.1467-7687.2011.01067.x.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in psychtoolbox-3? *Perception, 36.* https://doi.org/10.1068/v070821.

Kwon, M., & Legge, G. E. (2012). Spatial-frequency requirements for reading revisited. *Vision Research, 62,* 139–147. https://doi.org/10.1016/j.visres.2012.03.025.

Kwon, M., Liu, R., & Chien, L. (2016). Compensation for blur requires increase in field of view and viewing time. *PLOS ONE, 11*(9), e0162711. https://doi.org/10.1371/journal.pone.0162711.

Lao, J., Miellet, S., Pernet, C., Sokhn, N., & Caldara, R. (2017). Imap4: An open source toolbox for the statistical fixation mapping of eye movement data with linear mixed modeling. *Behavior Research Methods, 49*(12), 559–575. https://doi.org/10.3758/s13428-016-0737-x

Legge, G. E. (2006). *Psychophysics of reading in normal and low vision.* CRC Press.

Legge, G. E., Mansfield, J. S., & Chung, S. T. (2001). Psychophysics of reading. xx. linking letter recognition to reading speed in central and peripheral vision. *Vision Research, 41*(6), 725–743.

Lundqvist, D., Flykt, A., & Öhman, A. (1998). The karolinska directed emotional faces (kdef).

Malcolm, G. L., Lanyon, L. J., Fugard, A. J., & Barton, J. J. (2008). Scan patterns during the processing of facial expression versus identity: An exploration of task-driven and stimulus-driven effects. *Journal of Vision, 8*(8), 2–2.

McClure, E. B. (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychological bulletin, 126*(3), 424.

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics, 17*(6), 578–586. https://doi.org/10.3758/BF03203972.

Mehoudar, E., Arizpe, J., Baker, C. I., & Yovel, G. (2014). Faces in the eye of the beholder: Unique and stable eye scanning patterns of individual observers. *Journal of Vision, 14*(7), 1–11. https://doi.org/10.1167/14.7.6.

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law, 7*(1), 3–35. https://doi.org/10.1037/1076-8971.7.1.3.

Miellet, S., Caldara, R., & Schyns, P. G. (2011). Local jekyll and global hyde: The dual identity of face identification. *Psychological Science, 22*(12), 1518–1526. https://doi.org/10.1177/0956797611424290.

Miellet, S., He, L., Zhou, X., Lao, J., & Caldara, R. (2012). When east meets west: Gaze-contingent blindspots abolish cultural diversity in eye movements for faces. *Journal of Eye Movement Research, 5*(2). https://doi.org/10.16910/JEMR.5.2.5

Miellet, S., Lao, J., & Caldara, R. (2014). An appropriate use of imap produces correct statistical results: A reply to mcmanus. *Perception, 43*(5), 451–457. https://doi.org/10.1068/p7682.

Miellet, S., O'Donnell, P. J., & Sereno, S. C. (2009). Parafoveal magnification: Visual acuity does not modulate the perceptual span in reading. *Psychological Science, 20*(6), 721–728. https://doi.org/10.1111/j.1467-9280.2009.02364.x.

Miellet, S., Vizioli, L., He, L., Zhou, X., & Caldara, R. (2013). Mapping face recognition information use across cultures. *Frontiers in Psychology, 4*(February), 34. https://doi.org/10.3389/fpsyg.2013.00034

Näsänen, R., & Ojanpää, H. (2004). How many faces can be processed during a single eye fixation? *Perception, 33*(1), 67–77. https://doi.org/10.1068/p3417

Nuthmann, A. (2013). On the visual span during object search in real-world scenes. *Visual Cognition, 21*(7), 803–837. https://doi.org/10.1080/13506285.2013.832449.

Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges? evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology, 34*(1), 72–107. https://doi.org/10.1006/cogp.1997.0667.

Pelli, D. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437–442. https://doi.org/10.1163/156856897X00366.

Pelli, D., Tillman, K. A., Freeman, J., Su, M., Berger, T. D., & Majaj, N. J. (2007). Crowding and eccentricity determine reading rate. *Journal of Vision, 7*(2), 1–36. https://doi.org/10.1167/7.2.20.

Peterson, M. F., & Eckstein, M. P. (2012). Looking just below the eyes is optimal across face recognition tasks. *Proceedings of the National Academy of Sciences, 109*(48), 3314–3323. https://doi.org/10.1073/pnas.1214269109.

Peterson, M. F., & Eckstein, M. P. (2013). Individual differences in eye movements during face identification reflect observer-specific optimal points of fixation. *Psychological Science, 24*(7), 1216–1225. https://doi.org/10.1177/0956797612471684.

Ramon, M., Miellet, S., Dzieciol, A. M., Konrad, B. N., Dresler, M., & Caldara, R. (2016). Super-memorizers are not super-recognizers. *PloS One, 11*(3), e0150972. https://doi.org/10.1371/journal.pone.0150972.

Rayner, K. (1983). Eye movements, perceptual span, and reading disability. *Annals of Dyslexia, 33*(1), 163–173. https://doi.org/10.1007/BF02648003.

Rayner, K. (1986). Eye movements and the perceptual span in beginning and skilled readers. *Journal of Experimental Child Psychology, 41*(2), 211–236. https://doi.org/10.1016/0022-0965(86)90037-8.

Rayner, K., Slattery, T. J., & Belanger, N. N. (2010). Eye movements, the perceptual span, and reading speed. *Psychonomic Bulletin and Review, 17*(6), 834–839. https://doi.org/10.3758/PBR.17.6.834.

Rodger, H., Kelly, D. J., Blais, C., & Caldara, R. (2010). Inverting faces does not abolish cultural diversity in eye movements. *Perception, 39*(11), 1491–1503. https://doi.org/10.1068/p6750

Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show me the features! understanding recognition from the use of visual information. *Psychological Science, 13*(5), 402–409. https://doi.org/10.1111/1467-9280.00472.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137–149. https://doi.org/10.3758/BF03207704.

Targino Da Costa, A. L. N., & Do, M. N. (2014). A retina-based perceptually lossless limit and a gaussian foveation scheme with loss control. *IEEE Journal of Selected Topics in Signal Processing, 8*(3), 438–453.

Van Belle, G., De Graef, P., Verfaillie, K., Rossion, B., & Lefevre, P. (2010). Face inversion impairs holistic perception: Evidence from gaze-contingent stimulation. *Journal of Vision, 10*(5), 1–13. https://doi.org/10.1167/10.5.10.

Walker-Smith, G. J., Gale, A. G., & Findlay, J. M. (1977). Eye movement strategies involved in face perception. *Perception, 6*(3), 313–326.

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing, 13*(4), 600–612. https://doi.org/10.1109/TIP.2003.819861.

Woodworth, R. S. (1938). *Experimental psychology*. Holt.

Xivry, J.-J. O., Ramon, M., Lefèvre, P., & Rossion, B. (2008). Reduced fixation on the upper area of personally familiar faces following acquired prosopagnosia. *Journal of Neuropsychology, 2*(1), 245–268. https://doi.org/10.1348/174866407X260199.

Yarbus, A. L. (1967). *Eye movements and vision*. Plenum Press.

Yu, D., Cheung, S.-H., Legge, G. E., & Chung, S. T. (2007). Effect of letter spacing on visual span and reading speed. *Journal of vision, 7*(2), 2–2.

3.9   APPENDIX

3.9.1   APPENDIX A

Figure 3.9.1

Behavioural performance in function of face ethnicity
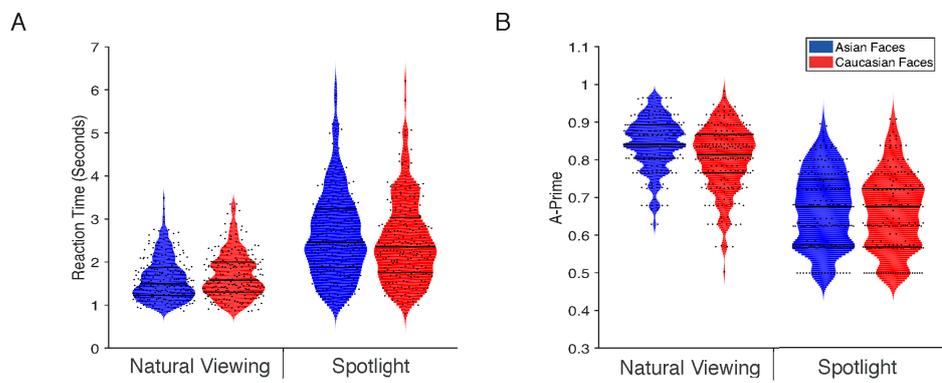
Note. Behavioural performance in function of face ethnicity. (A) Reaction time in function of the viewing condition. (B) A-Prime in function of the viewing condition. Black lines represent the median (thick black line) and the 25th and 75th percentile (thin black line).

3.9.2   Appendix B

Figure 3.9.2

Impact of the spotlight on fixation pattern
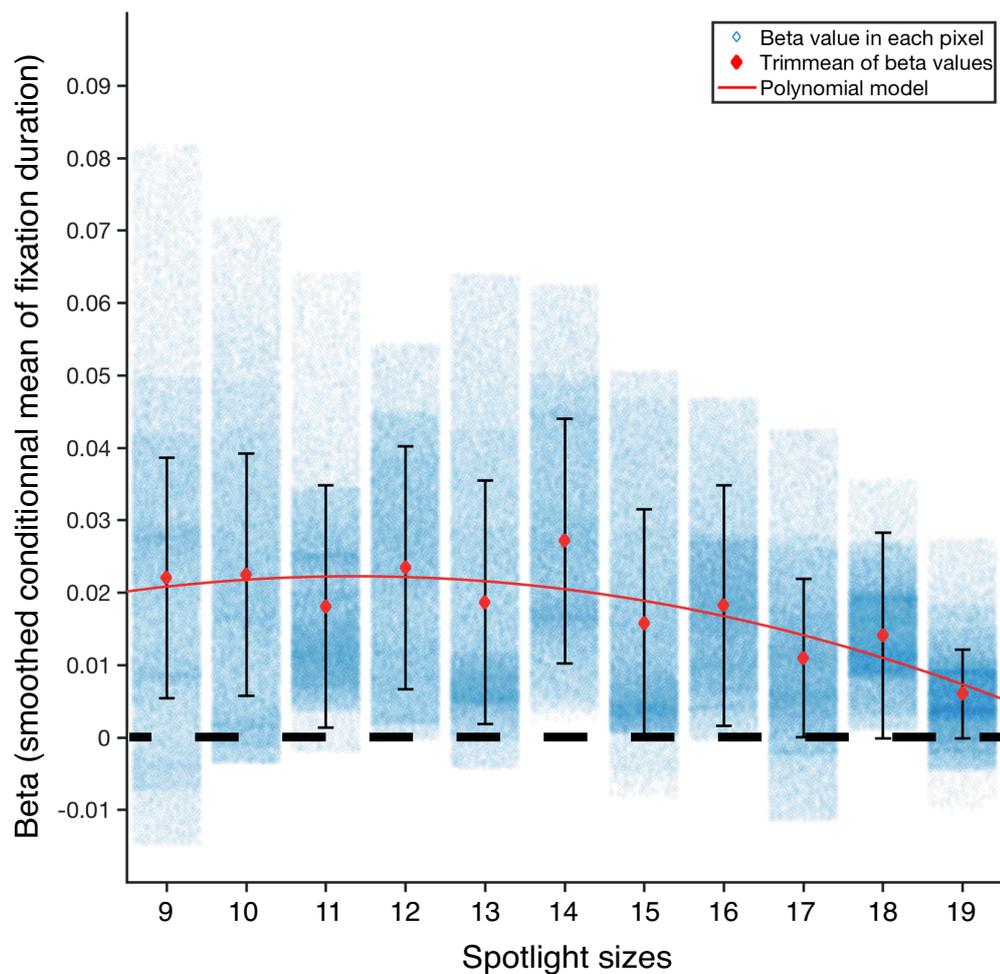
Note. Impact of spotlight on fixation pattern (see Figure 5.2.1C) represented by the trimmean of beta values within the mask on the face region (see *iMap4* procedure for more information). Beta values of each pixel inside the mask are represented by diamond, trimmean are represented by red filled diamond and are surrounded by confidence interval (bold line). Finally, the fitting of a second degree polynomial model is represented by the bold red line.

3.9.3   Appendix C

Figure 3.9.3

Pooled variance across natural viewing and the spotlight
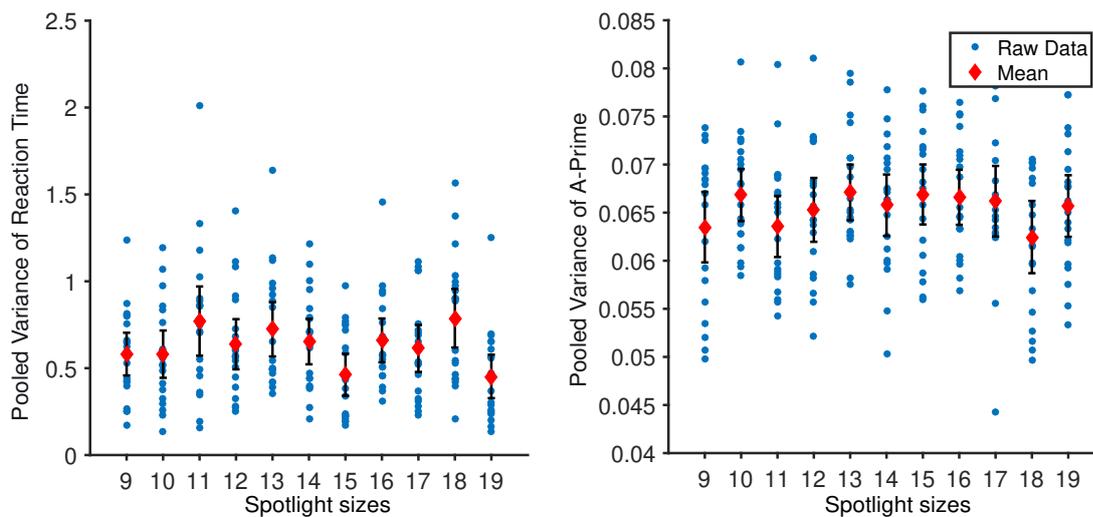
Note. Pooled variance across conditions (Natural Vision and Spotlight) used to compute Cohen's d for both Reaction Time (A) and A-Prime (B). These two figures show that the variance is similar across spotlight sizes. Pooled variance of each participants are represented by filled circles, mean are represented by red filled diamond and are surrounded by confidence interval (bold line).

I would rather be vaguely right than very wrong.

Nassim Taleb

4

Non-hearings have an amplified facial information use

4.1 Abstract

Eye movements are a crucial feat to identify the visual world. Despite eye movements studies shedding light on visual information sampling and use in many field, the extent to which visual sampling is modulated and shaped by non-visual information remains difficult to control and assess. A particular fate of nature might help to achieve this aim: the occurrence of hearing loss. Indeed, research has shown that early and profound hearing loss enhances the sensitivity and efficiency of the visual sensory channel in non-hearing individuals, resulting in a larger peripheral visual attention as compared to the hearing population. However, whether such perceptual bias extends to visual sampling strategies deployed during biologically relevant face recognition tasks remains to be clarified. To shed light on this question, we recorded eye movements of non-hearing and hearing observers during a delayed matching task with upright and inverted faces. Non-hearing observers showed a preferential central fixation pattern while hearing controls exhibited a focused fixation pattern just below the eyes. Interestingly, the non-hearing observers were not impaired by face inversion and did not change their sampling strategies. To assess whether those particular fixation strategies result in a larger information intake, a gaze-contingent design parametrically and dynamically modulating the quantity of information available at each fixation—the *Expanding Spotlight* was used. From visual information used reconstruction with a retinal filter, our results revealed that unlike hearing participants, non-hearing observers used larger information intake. This visual sampling strategy was robust and effective to inverted face; cancelling the face inversion effect. Altogether, our data shows that the face system is flexible and might tune to distinct strategies as a function of visual and social experience shaping information intake.

4.2 INTRODUCTION

Living organisms are in a constant state of information exchange with their surroundings. Efficiently perceiving and understanding information from different sources are, hence, vital for survival. To achieve this aim, specialized sensory organs adapted, thanks to evolution, to transform specific physical energy into sensory information (Thewissen & Nummela, 2008). For example, while visual information is its main source of information, human beings rely on multisensory signals from visual, auditory, equilibrium, olfactory, gustatory, and somato-sensory information and make use of these sensory inputs to build representations of the world (Stein, 2012). Notably, different sensory channels are not processed in isolation; they are integrated non-linearly to form a unique and unified representation. One example depicting such multisensory integration is the McGurk effect, where the perception of the same auditory signal could be different in the presence or absence of visual stimuli (McGurk & MacDonald, 1976). Indeed, even when a unique sensory channel is focused the representation of the stimulus is rarely uni-modal and rather is intertwined at the brain level especially in higher-level visual cognition tasks such as face recognition (*e.g.,* Blank et al., 2011).

Consequently, because of multi-sensory integration, understanding the impact of a perception channel loss could unveil crucial information upon the lost channel as well as other sensory channels. More specifically, how hearing loss modulates visual perception, the information sampling strategy, and the internal representation of the visual world. For example, one of the notable differences observed in the individuals with early profound hearing loss is their enhanced sensitivity to peripheral vision in certain selective tasks as compared to their hearing counterpart (for a review, see Bavelier et al., 2006). Such impact has been demonstrated using various visual tasks with low-level visual stimuli (*i.e.,* non-naturalistic visual stimuli like shapes or gratings; Buckley et al., 2010; Codina et al., 2011; Lore & Song, 1991; Proksch & Bavelier, 2002; Stivalet et al., 1998) resulting in the non-hearing observers paying more attention towards the peripheral visual field than the hearing controls. Moreover, these differences seem to be genuinely related to the hearing loss, but do not appear to be imputed to other long-term behavioural changes such as observers with sign language ability (*i.e.,* signers). For instance, by including hearing native signers and/or non-hearing non-signer as additional control,

studies found that sign language use is not sufficient to explained these attentional differences (Bosworth & Dobkins, 2002; Dye et al., 2009; Fine et al., 2005; Neville & Lawson, 1987; but see Dye et al., 2016). However, early exposure to the sign language seems to impact on motion perception with a greater recruitment brain areas in the left hemisphere as compared to typically hearing non-signers (Bavelier et al., 2001). Altogether those results suggest that hearing loss enhance peripheral processing while signing improve motion perception. However, the interplay of signing and hearing loss impact on face perception remains to clarify.

Regarding other higher-level visual cognition tasks, such as face recognition, several impacts of hearing loss were also found. For example, De Heering et al. (2012) evaluated face recognition abilities of non-hearing and hearing observers with two different tasks: an inversion-matching task and a face-composite task. The non-hearing observers have shown a larger Face inversion effect (FIE; Yin, 1969), while responding slower when matching inverted faces than the hearing controls. However, both groups showed similar performance in the composite face task. A finer-grained difference between non-hearing and hearing observers on FIE was reported by H. He et al. (2016). Using a face dimension task, they parametrically manipulated the configural and featural changes of a pair of faces by extending or by diminishing distance between features. They reported that non-hearing signers showed a reduced FIE as compared to hearing non-signers when the changes in dimension occurred in the mouth region. Importantly, as with tasks using simple low-level stimuli, these differences in face-processing between non-hearing and hearing participants might also be related to the practice of the sign language. Indeed, for signers, the face carries additional grammatical and syntactic information besides emotional expressions (Liddell, 2003). Thus, the communication using sign language might have a positive impact on face recognition requiring more attention, especially in the peripheral region (see, *e.g.*, Dye et al., 2016). For example, early studies showed that non-hearing signers are better than hearing non-signers at face matching, while their performance for non-face objects are comparable (Arnold & Murray, 1998). Moreover, Bettger et al. (1997) showed that both non-hearing and hearing native signers are better than hearing non-signers at discriminating face stimuli under various conditions. In both studies, the performance of hearing signers is closer to non-hearing signers than hearing non-signers. These increased performances in signers have been related to better discrimination of slight differences in facial features (McCullough & Emmorey, 1997). While the behavioural advantage could be attributed to a specific

perceptual tuning as a result of sign language practice, it could also result from a difference in behavioural response strategy. For example, Stoll et al. (2018) reported that both non-hearing and hearing signers responded slower but more accurately than hearing non-signers, indicating a speed-accuracy trade-off in face recognition for signers. Moreover, using a Bayesian hierarchical drift-diffusion model, they showed that signers are faster to accumulate information from faces than non-signers (*i.e.*, higher drift-rate) while displaying a more conservative decision strategy (*i.e.*, higher response threshold). However, it remains unclear whether such differences result from a difference in facial information intake and how non-hearing observers make use of information sampling to recognise faces.

To clarify facial information use in non-hearing observers, we recorded the eye movement of non-hearing signers and hearing non-signers while they were performing a delayed matching task with both upright and inverted faces. Additionally, to determine the quantity and quality of the perceived visual information, besides Natural Viewing, we applied a gaze-contingent design: the Expanding Spotlight (Miellet et al., 2013). Using a Gaussian aperture centred at the fixation location and dynamically expanding as a function of the fixation duration, allowing us to assess what and how much facial information is intake.

4.3 MATERIAL AND METHODS

4.3.1 PARTICIPANTS

Fifty-one Western Caucasian hearing non-signers—34 females—aged between 18 and 33 ($M = 21.56$, $SD = 3.48$) participated in the current study. All participants had a normal or corrected-to-normal vision. Seven participants were excluded due to noisy eye-tracking recording, and two participants due to their low behavioural performance (*i.e.*, below 50% in either hit or correct rejections during Natural Viewing conditions), resulting in 42 hearing participants. The hearing participants were from the University of Fribourg, Switzerland. Thirty-three Western early non-hearing signers—17 females—aged between 18 and 32 ($M = 22.46$, $SD = 3.10$) par-

ticipated in the current study. All non-hearing participants had severe to profound hearing loss (dB loss > 70) from birth or the first years of life and were all native or early American Sign language (ASL) signers (before the age of 5). All non-hearing participants had a normal or corrected-to-normal vision. We excluded 5 participants due to noisy eye-tracking recording and 3 participants due to their low behavioural performances (*i.e.*, below 50% in either hit or correct rejections during Natural Viewing conditions), resulting in 24 non-hearing participants as the final sample size. We recruited the non-hearing participants from the National Technical Institute for the Deaf, in Rochester, New York. This study was approved by the local Ethical Committees in both institutes. All participants gave written informed consent and received course credit for their participation from their affiliated institute.

4.3.2   STIMULI

Face Stimuli were obtained from the KDEF database (Lundqvist et al., 1998) and consisted of 56 unique WC identities (equal numbers of females and males). For each identity we selected three images from this database: two depicting one a basic emotion and the other displaying a neutral expression. Facial expressions were counter-balanced across all identities. Stimuli were aligned on eyes and mouth positions and were normalized for overall luminance. External features of the face stimuli were removed by superimposing a uniform contour (for example, see Figure 4.3.1). The stimulus subtended about 15° of visual angle vertically and horizontally at a 70 cm viewing distance.

4.3.3   EXPANDING SPOTLIGHT

We used a gaze-contingent technique previously developed by Miellet et al. (2013) to quantify the visual information sampling during face task: the *Expanding Spotlight*. This gaze-contingent paradigm makes use of a dynamic Gaussian aperture actively centred on the participant's gaze location. The area outside of the Gaussian aperture was masked by an average face of all the

Figure 4.3.1

Experimental procedure to study facial information use and FIE in non-hearing
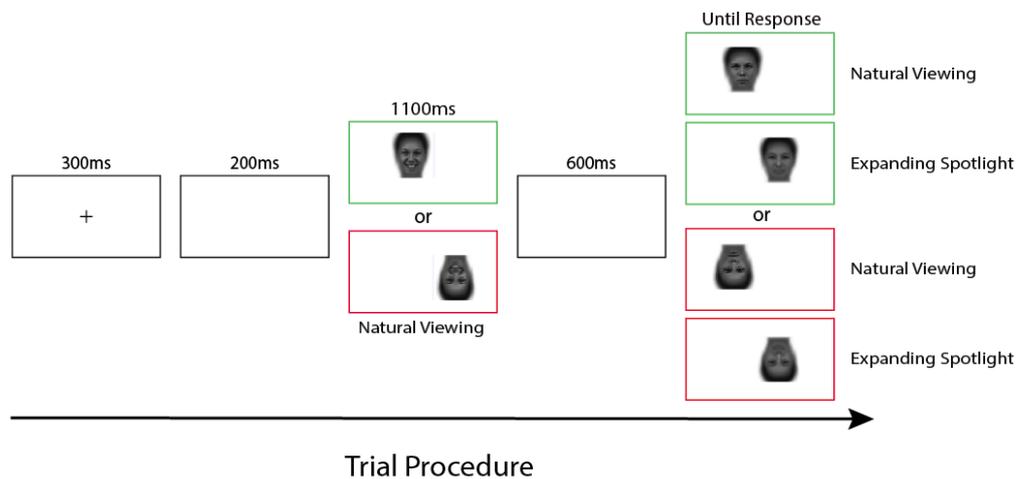
Note. Experimental procedure used in the current study. Each trial started with a 300 ms fixation cross and a 200 ms blank screen. Then, faces to learn will be displayed either upright (green framed) or inverted (red framed) for 1100 ms. After a blank screen of 600 ms, either the same face or a different one displayed until response. This face was presented with natural vision or with the Expanding Spotlight.

stimuli used in the present experiment. The average face mask allowed to program next sac-
cades without providing any task-related information. Importantly, in the current experiment,
the Gaussian aperture expanded at a rate of $\frac{1}{12}$ degree of visual angle per millisecond of fixa-
tion duration. The window was reset to its initial size after a saccade. The expansion rate was
chosen based on two main considerations that were validated empirically in a previous study
(Miellet et al., 2013). First, it is sufficiently fast to keep the fixation duration in the average
range observed during face-processing under Natural Viewing conditions, *i.e.*, about 300 ms.
Within this range, the spotlight reaches the size of about 14 degree of visual angle. Thus, fixat-
ing the centre of the face, a Gaussian aperture of this size allows simultaneous processing of the
eyes and mouth. Secondly, the expansion rate is slow enough to provide sufficient sensitivity
to highlight differences in extent of facial information intake. For instance, at least 100 ms is
required for the spotlight to expand over the area of a feature (*e.g.* an eye; for a demonstration,
see Miellet et al., 2013).

4.3.4   Apparatus

Stimuli were presented using the Psychophysics Toolbox Version 3 (PTB-3; Kleiner et al., 2007;
D. Pelli, 1997) in Matlab R2010a (The MathWorks, Natick, MA, USA). The face stimuli were
presented on a 27-inch CRT screen with 1920x1080 pixels resolutions at a refresh rate of 100
Hz at a 70 cm viewing distance. Eye movements were recorded via Eyelink Toolbox exten-
sions (Cornelissen et al., 2002), using an SR Research Desktop-Mount EyeLink 2K eye tracker,
which has an average gaze position error of about .25° and a spatial resolution of .01°. Only the
dominant eye was tracked at a sampling rate of 1000Hz. A chin/forehead rest ensured stable
head positioning.

4.3.5   Procedure

The experiment procedure is depicted in Figure 4.3.1. Participants performed a two-alternative
forced-choice delayed-matching task. Each trial began with a central fixation of 300 ms followed

by a blank screen of 200 ms. Then, a face was presented for 1100 ms on a random screen location in order to avoid the central tendency of the first fixation. Importantly, initial faces were always presented under natural viewing, *i.e.*, without the expanding spotlight paradigm. After a 600 ms blank screen, a target face was presented until a response with the expanding spotlight (spotlight condition) or in natural viewing (natural viewing condition). Participants' task was to decide as rapidly and accurately as possible whether the probe depicted the same, or a different identity as the previously presented face. To avoid trivial image matching, the initial and target faces always conveyed a different facial expression.

Each participant completed two blocks: one with target faces presented under natural viewing and the other with target faces under the Expanding Spotlight. The blocks' order was assigned randomly across participants. At the beginning of each black, there was a short training block of 8 trials to familiarise participants with the task and the gaze-contingent paradigms in the spotlight condition. The identities presented throughout practice trials were not included in the subsequent testing. The testing block contained 96 trials, resulting in a total of 192 trials. Each block encompassed the same number of target face matching the initial face and differing from the initial face. The same number of faces was presented in upright and inverted conditions. The initial and target face stimuli were always presented in the same orientation. Thus, we implemented a 2 x 2 within-participant design including the factors viewing condition (Natural Viewing or Expanding Spotlight) and orientation (upright or inverted). A nine-point calibration was conducted before each block; additional calibrations were performed until reaching an optimal recording quality.

4.3.6   Data Analyses

Behavioural Analysis

For each participant, we computed the averaged reaction time (M_{RT}), accuracy (M_{ACC}), D-Prime (D') and A-prime (A', Stanislaw & Todorov, 1999) for each condition. These behavioural measures were then fitted with Generalized Hierarchical Linear Models. A full 2 (Group) × 2

(Viewing conditions) × 2 (face orientation) model was used, with each participant modelled as a random grouping effect (*i.e.*, repeated measure) using Generalized Estimating Equations in statsmodels python package. Bounded response (*i.e.*, accuracy rate and A') was fitted using the same linear equation with a logit link function. Hypothesis testing of the main effects and interaction were computed using ANOVAs.

To achieve a better estimation of the face inversion effect, we fitted a Bayesian linear mixed model with random intercept and slope for accuracy rate and reaction time (denoted as Y in equation 4.1).

$$\begin{aligned}
Y \sim{} & 1 + Group + ViewingCondition + FaceOrientation + Group \cdot ViewingCondition + \\
& Group \cdot FaceOrientation + ViewingCondition \cdot FaceOrientation + Group \cdot \\
& ViewingCondition \cdot FaceOrientation + (ViewingCondition + FaceOrientation + \\
& ViewingCondition \cdot FaceOrientation | participant)
\end{aligned}$$

$$(4.1)$$

This model accounted more accurately for the individual differences by modelling the random intercept of the within-participant condition. The Bayesian Generalized Linear mixed models were fitted in Bambi python package using the PyMC3 backend (Capretto et al., 2020). We performed statistical inference by sampling from the posterior using the No U-Turn Sampler (NUTS; Hoffman, Gelman, et al., 2014). Four Markov Chain Monte Carlo (MCMC, see, *e.g.*, Metropolis et al., 1953) chains were run with 2000 samples each and with the default NUTS sampler setup. The first 1000 samples were used to tune the mass matrix and step size for NUTS. These samples were subsequently discarded, leaving a total of 4000 samples for each model parameter. Model convergence was diagnosed by computing Gelman and Rubin's convergence, checking whether there is any divergent sample that has been returned from the sampler, and visually inspecting the mixing of the traces (see, *e.g.*, Gelman et al., 2013). Using the posterior samples, we computed the estimation of the FIE for each group of observers in both Natural Viewing and Expanding Spotlight conditions.

EYE MOVEMENTS PREPROCESSING

Raw eye-tracking data was preprocessed using a custom algorithm based on angular velocity. It employs threshold parameters for saccade and fixation detection similar to that of the Eyelink software (saccade velocity threshold = 30° per second; saccade acceleration threshold = 4000° per second).

Spatial modelling of fixations was performed using *i*Map4, a Matlab toolbox for the statistical fixation mapping of eye movement data (Caldara & Miellet, 2011; Lao et al., 2017). It applies a pixel-wise Linear Mixed Model (LMM) on the smoothed fixation maps and implements non-parametric statistics based on resampling as hypothesis testing. In brief, fixations duration of each participant from each trial were projected into two dimensional space for each fixation location. Fixation duration maps were then smoothed with a Gaussian Kernel function of Full Width at Half Maximum (FWHM) at 1° visual angle. The resulting three dimensional matrix (trials × x-Size × y-Size) was then normalized per trial to have a maximum value of 1. We then computed the conditional mean for each condition within each participant. We first conducted a multivariate similarity analysis between different conditions and groups using a representative distance matrix technique based on principal component analysis and Mahalanobis distance. Then, we fitted a linear mixed model with the smoothed fixation maps as the response variable (see equation 4.2)

$$NormalizedFixationDuration(x, y) \sim 1 + Group + ViewingCondition + FaceOrientation +$$
$$Group \cdot ViewingCondition + Group \cdot FaceOrientation + ViewingCondition \cdot$$
$$FaceOrientation + Group \cdot ViewingCondition \cdot FaceOrientation + (1|participant)$$

$$with \ 1 \leqslant x \leqslant xSize \ and \ 1 \leqslant y \leqslant ySize$$

$$(4.2)$$

This is, hence, a 2 (group) × 2 (viewing conditions) × 2 (face orientation) full-factorial mixed model with each participant as the random intercept. Importantly, only the fixation data during correct target face identification was modelled here. Null hypothesis significance testing on main effects and interactions were performed on the contrast of estimated model coefficients using bootstrapped clustering procedure.

FACIAL INFORMATION USE RECONSTRUCTION

To further demonstrate the differences between fixation strategies and visual information use, we took advantage of the eye movement data collected in the Expanding Spotlight condition. Indeed, the Expanding Spotlight parametrically and dynamically modulates the quantity of information available at each fixation (Miellet et al., 2013). Thus, The available visual information at different facial features (left eye, right eye, the nose, or the mouth) was estimated by computing the average fixation duration at each region using a 50*50 pixel region of interest (ROI) and reconstructing the states of the expanding spotlight. Finally, we further reconstructed the total available information at each fixation in both upright and inverted conditions and across groups using a retinal filter. The reconstruction procedure followed the same algorithm described in Miellet et al. (2013).

4.4   RESULTS

4.4.1   BEHAVIOUR

Behavioral result across conditions were displayed in Figure 4.4.1. We found significant main effects of both orientation and viewing condition for the four behavioral measurement (M_{RT}: $F(1,256) = 10.81$, $p=.001$ and $F(1,256)=13.52$, $p < .001$, respectively; M_{ACC}: $F(1,256) = 8.09$, $p = .005$ and $F(1,256) = 5.91$, $p = .016$; A': $F(1,256) = 6.73$, $p = .010$ and $F(1,256) = 6.94$,

p = .009; D': *F*(1,256) = 7.19, *p* = .008 and *F*(1,256) = 5.50, *p* = .020). Participants' performance for matching upright faces (M_{RT}: 1352 ms, 95% CI = [1265, 1460]; M_{ACC}: 76.80%, 95% CI = [75.30, 78.30]) was better than inverted faces (M_{RT}: 1410 ms, 95% CI = [1317, 1517]; M_{ACC}: 72.20%, 95% CI = [70.80, 73.70]). Participants also performed better in the Natural Viewing condition (M_{RT}: 1118ms, 95% CI = [1075, 1162]; M_{ACC}: 77.60%, 95% CI = [76.00, 79.10]), compared to the Expanding Spotlight condition (M_{RT}: 1644ms, 95% CI = [1531, 1770]; M_{ACC}: 71.40%, 95% CI = [70.10, 72.70]). The differences between the non-hearing and hearing observers were not significant (maximum main effect of groups across the four measurement: F_{max}(1,256)=2.53, p > .05). However, the performance estimation for the non-hearing participants (M_{RT}: 1548ms, 95% CI = [1393, 1721]; M_{ACC}: 73.00%, 95% CI = [71.30, 74.90]) are generally lower than the hearing controls (M_{RT}: 1286 ms, 95% CI = [1233, 1339]; M_{ACC}: 75.30%, 95% CI = [74.00, 76.70]). No other effects from the ANOVAs were significant.

Given that we have an unbalanced number of participants between the two groups, to better estimate the behavioural performance of non-hearing observers and Hearing control across the different conditions, we applied a Bayesian Linear Mixed effect model with an intercept and slope model. The model coefficients from such models are often much easier to fit and better regulated with weakly informed prior using a Bayesian approach. In our case, the Bayesian LMM return a very similar result compare to the Frequentist approach (with only participant intercept), with the exception that we also found a weak group difference on mean reaction time between the non-hearing and Hearing participants. Non-hearing observers respond slower than the Hearing controls (ΔRT_{NH-H} = 109ms, [-12, 231] 95% highest posterior density).

Importantly, using the full posterior distribution, we quantified the FIE by computing the differences between upright and inverted conditions for the estimated accuracy rate in both Natural Viewing and Expanding Spotlight conditions. Interestingly, the non-hearing observers displayed smaller face inversion effect in both Natural Viewing condition (*uprightinverted* contrast of logit regression coefficients: FIE_{NH} = 0.27, 95% CI=[0.07, 0.46], FIE_{H} = 0.44, 95% CI=[0.27, 0.61]), and in Expanding Spotlight condition (FIE_{NH} = -0.01, 95% CI=[-0.18, 0.18],

Figure 4.4.1

Behavioral results of hearing and non-hearing participants under the Expanding Spotlight
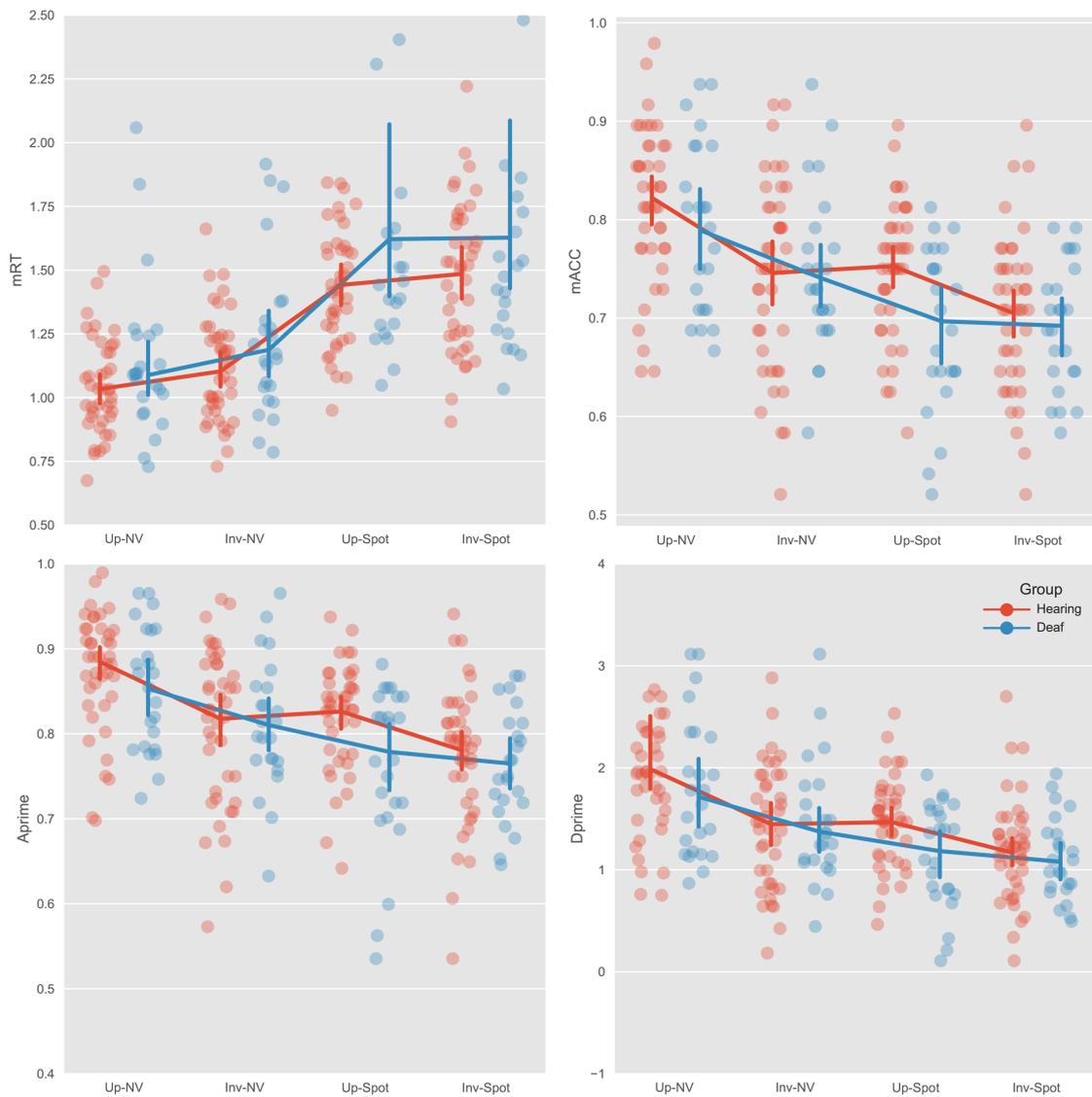
Note. Behavioral results in hearing and non-hearing observers. Every single dot represents one observation. Blue indicates non-hearing participants, red indicates hearing participants. The line plot shows the conditional mean and the error bars show a 95% bootstrapped confidence interval of the estimation.

FIE_H = 0.23, 95% CI=[0.08, 0.38], see Figure 4.4.2 a)). More specifically, the non-hearing observers do not show the Face inversion effect in the spotlight condition: the behavioural performance is equally impaired under the spotlight condition for both upright and inverted face stimuli.

4.4.2 EYE MOVEMENTS

The mean fixation duration maps are shown in Figure 4.4.3 a). The average fixation pattern indicates a classical T-shape where the participant's gaze distributed mostly on the eyes and mouth. Interestingly, as shown in the representative distance matrix (Figure 4.4.3 b)), the spatial fixation distribution, as represented by the fixation maps, are highly similar between non-hearing and hearing observers, with the exception in the inverted condition (top right corner of the distance matrix).

Fixation mapping using iMap4 shows that there is a significant group effect and FIE (Figure 4.4.4 a)). Non-hearing observers showed a more distributed fixation pattern compare to hearing controls ($F_{min}(1,256)$ = 3.88, $p < .05$ cluster corrected and $F_{max}(1,256)$ = 13.55, $p < .05$ cluster corrected). Moreover, both groups of observers fixated more on the nose and mouth area during the viewing of inverted faces ($F_{min}(1,256)$ = 3.88, $F_{max}(1,256)$ = 29.16, p $< .05$ cluster corrected). Following the significant main effect, we computed the fixation map contrast between upright and inverted conditions to better quantify the face inversion effect on fixation patterns in both groups. Interestingly, as shown in Figure 4.4.4 b), significant strategy change is only presented in the hearing controls (significant cluster indicated in black line), but not in the non-hearing observers.

At the local maxima (*i.e.*, around the eye regions in Figure 4.4.4 b)), the Hearing observers showed a fixation contrast of 4.69, 95% CI=[1.69, 7.70] between upright and inverted faces, whereas the non-hearing observers showed a fixation contrast of 2.59, 95% CI=[-1.38,

Figure 4.4.2

Face inversion effect in hearing and non-hearing observers
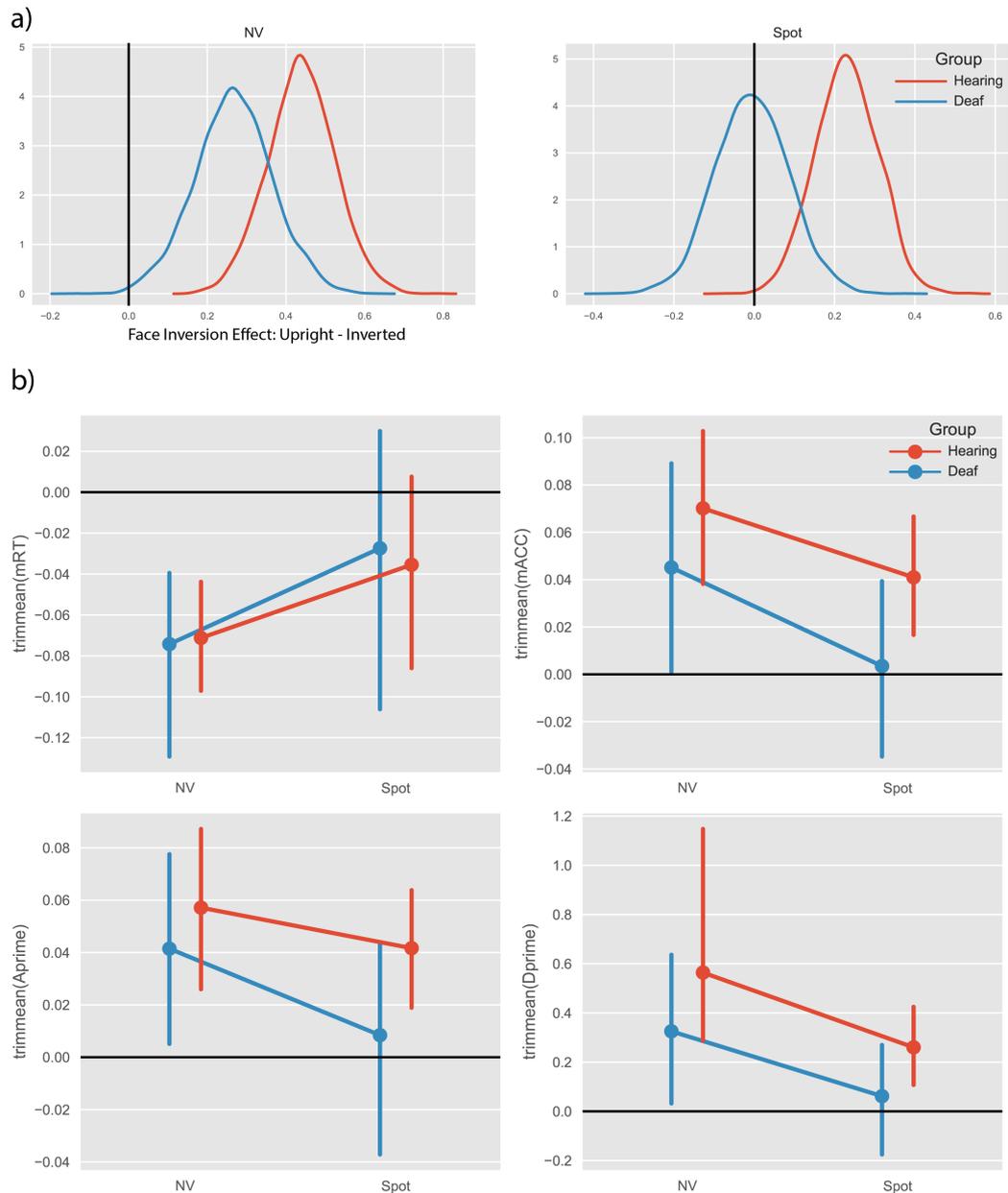
Note. Face inversion effect in hearing and non-hearing observers. a) FIE of accuracy estimated by performing contrast on the posterior estimation. b) bootstrapped estimation of FIE across the different behavioural measurements.

140

Figure 4.4.3

Descriptive eye movements results for hearing and non-hearing observers

a)

b)
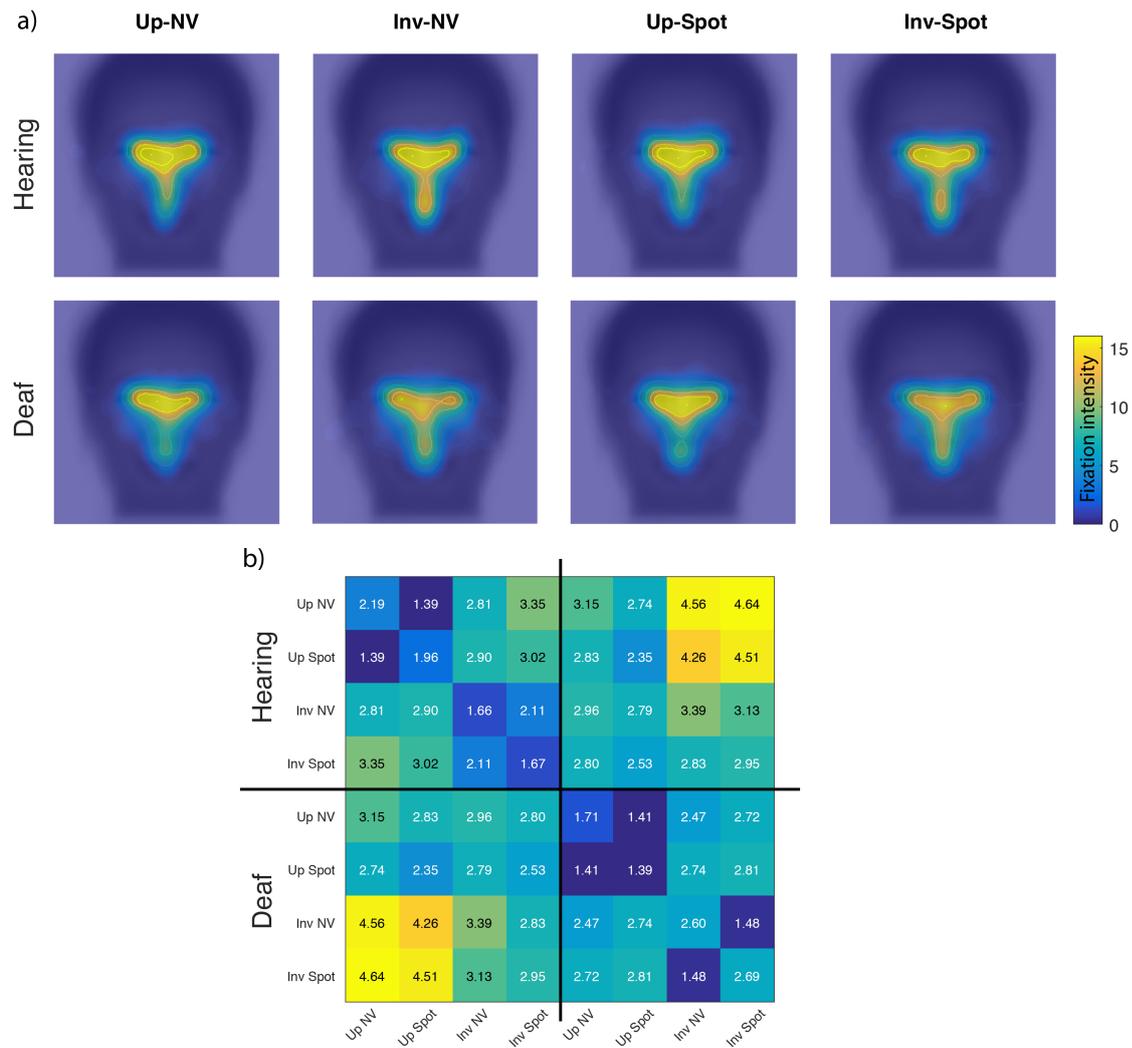
Note. Descriptive eye movements results. a) normalized fixation duration map for both groups. (b) representative distance matrix computed on the multivariate Mahalanobis distance between two conditions. The lower value shows a smaller multivariate distance (*i.e.*, more similar).

Figure 4.4.4

Fixation mapping result using iMap4 in hearing and non-hearing observers
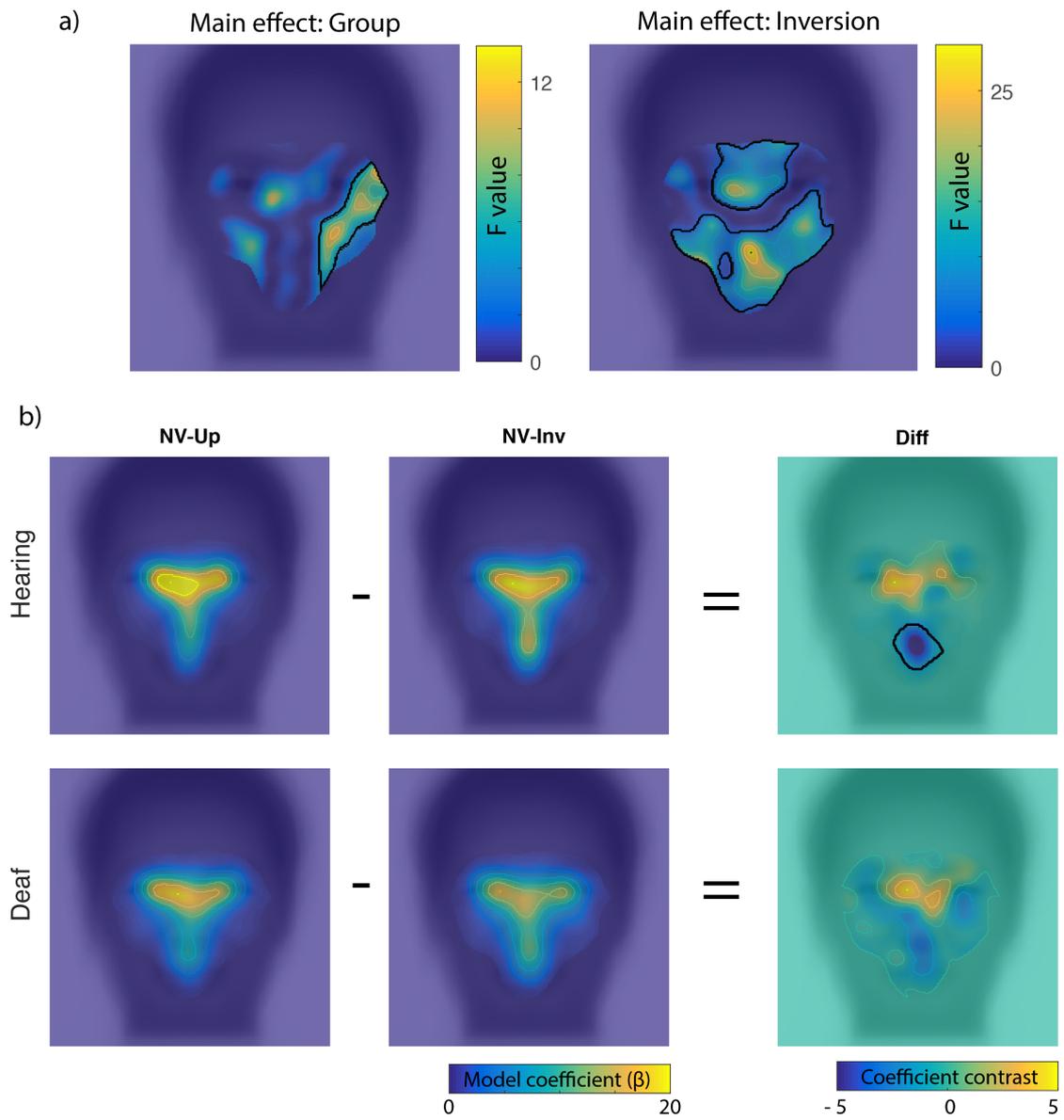
Note. Fixation mapping result using *i*Map4 in hearing and non-hearing observers. a) significant main effect of Group and Orientation. b) linear contrast between upright and inverted conditions for both deaf and hearing observers. Black lines delineate significant differences between conditions.

6.57]. At the local minimal around the mouth region, the hearing observers showed a fixation contrast of -6.12, 95% CI=[-8.81, -3.44] and the non-hearing observers showed a fixation contrast of -1.94, 95% CI=[-5.50, 1.61] (see Figure 4.4.5).

Figure 4.4.5

Eye movements FIE for hearing and non-hearing observers
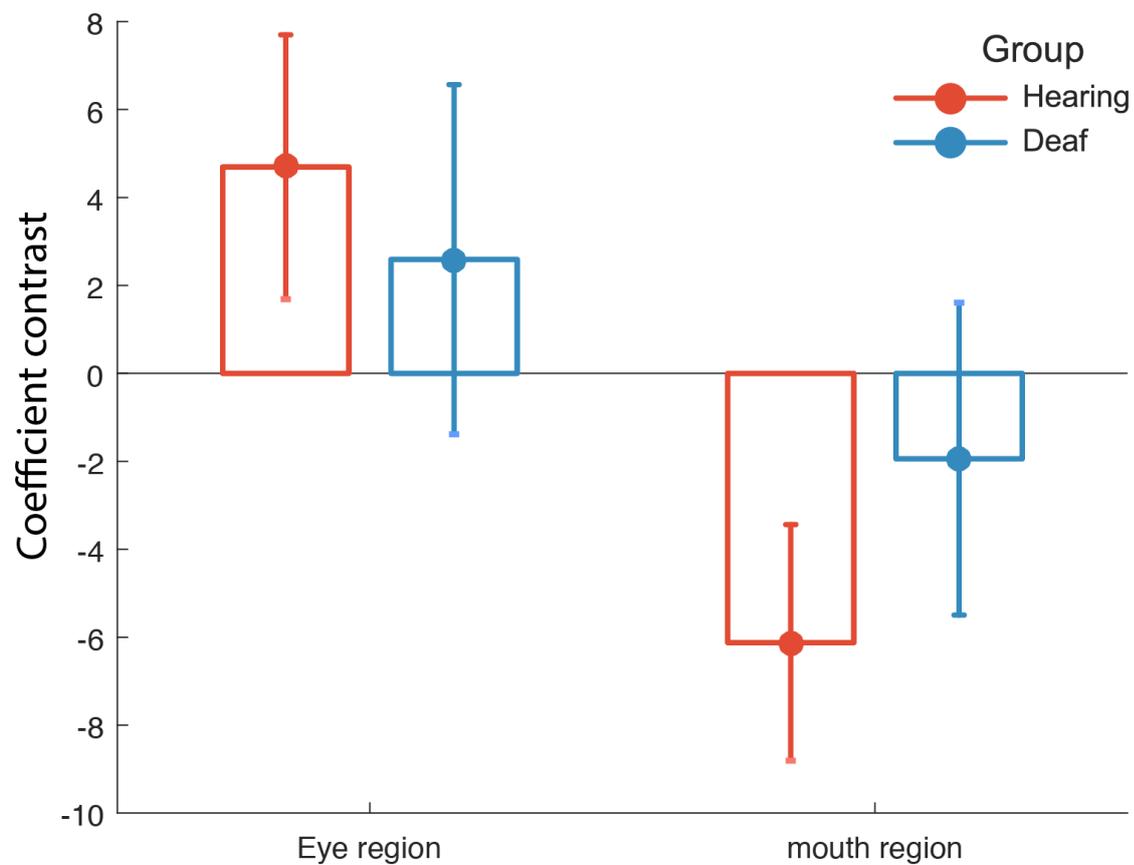
Note. Fixation strategy changes between the viewing of upright and inverted faces (*i.e.*, eye movement FIE). Bar plot shows point estimation at the local maximal and minimal of the contrast of the model coefficient (differences map in Figure 4.4.4b)). Error bar shows 95% CI.

4.4.3 Facial Information Use Reconstruction

The difference between fixation strategy revealed that non-hearing participants display longer fixation duration (418ms, 95% CI=[401.70, 434.90]) overall compare to hearing controls (375 ms, 95% CI=[365.00, 384.50]). Moreover, both groups display longer fixation duration in the upright compare to the inverted condition (upright: 408ms, 95% CI=[394.80, 422.60]; inverted: 384ms, 95% CI=[371.90, 396.80]). As a demonstration, Figure 4.4.6 a) shows the fixation duration distribution for the upright and inverted faces when the fixation is located on the left eye. Figure 4.4.6 b) shows the available visual information during Expanding Spotlight condition quantified according to the fixation duration. While the difference of perceptual span between non-hearing and hearing observers is comparable for both upright (non-hearing – hearing: 47.50 ms, 95% CI=[41.00, 55.50]) and inverted condition (non-hearing – hearing: 39.20 ms, 95% CI=[32.60, 45.40]); the non-hearing participants show a larger fixation duration change between the upright and inverted condition (*uprightinverted* contrast: 28.20 ms, 95% CI=[27.10, 30.90]) than the hearing controls (*uprightinverted* contrast: 19.90 ms, 95% CI=[18.60, 20.80]).

Using the fixation duration estimated from the Expanding Spotlight condition in Figure 4.4.6, conjointly with the fixation location map recorded during the Natural Viewing condition, we can estimate the available visual information during the face identification task in both upright and inverted conditions. As shown below, both groups of observers extracted a very similar quality of visual information, with the non-hearing observers extracting slightly more visual information. Moreover, the quantity of visual information is nearly identical between upright and inverted conditions (see Figure 4.4.7).

144

Figure 4.4.6

Averaged information used at each fixation by hearing and non-hearing observers
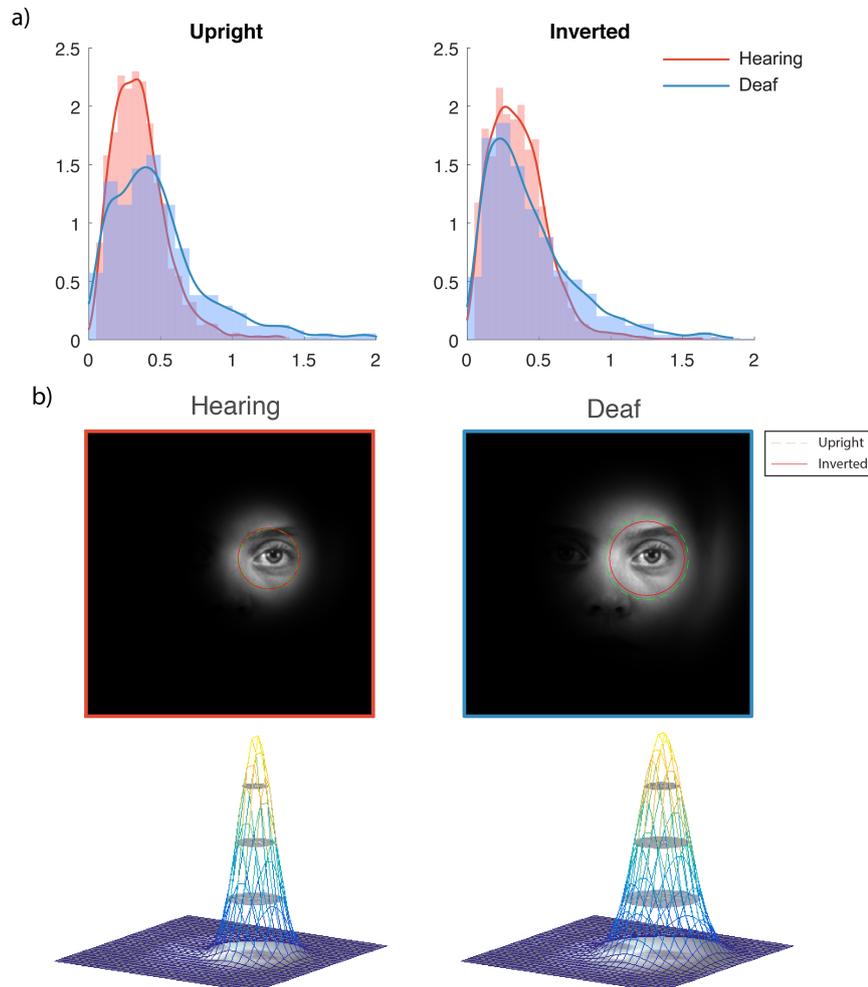
Note. Visual information at each fixation. a) Smoothed histogram shows the distribution of fixation duration at the left eye for both non-hearing and hearing observers in upright and inverted conditions. b) available visual information during Expanding Spotlight condition quantified according to the fixation duration. Hearing participants are presented on the left panels and non-hearing participants are presented on the right. Within the visual field demonstration, the green dash line indicates upright condition and the red solid line indicates the inverted condition.

Figure 4.4.7

Reconstruction of the facial information used at each fixation by hearing and non-hearing observers with upright or inverted faces
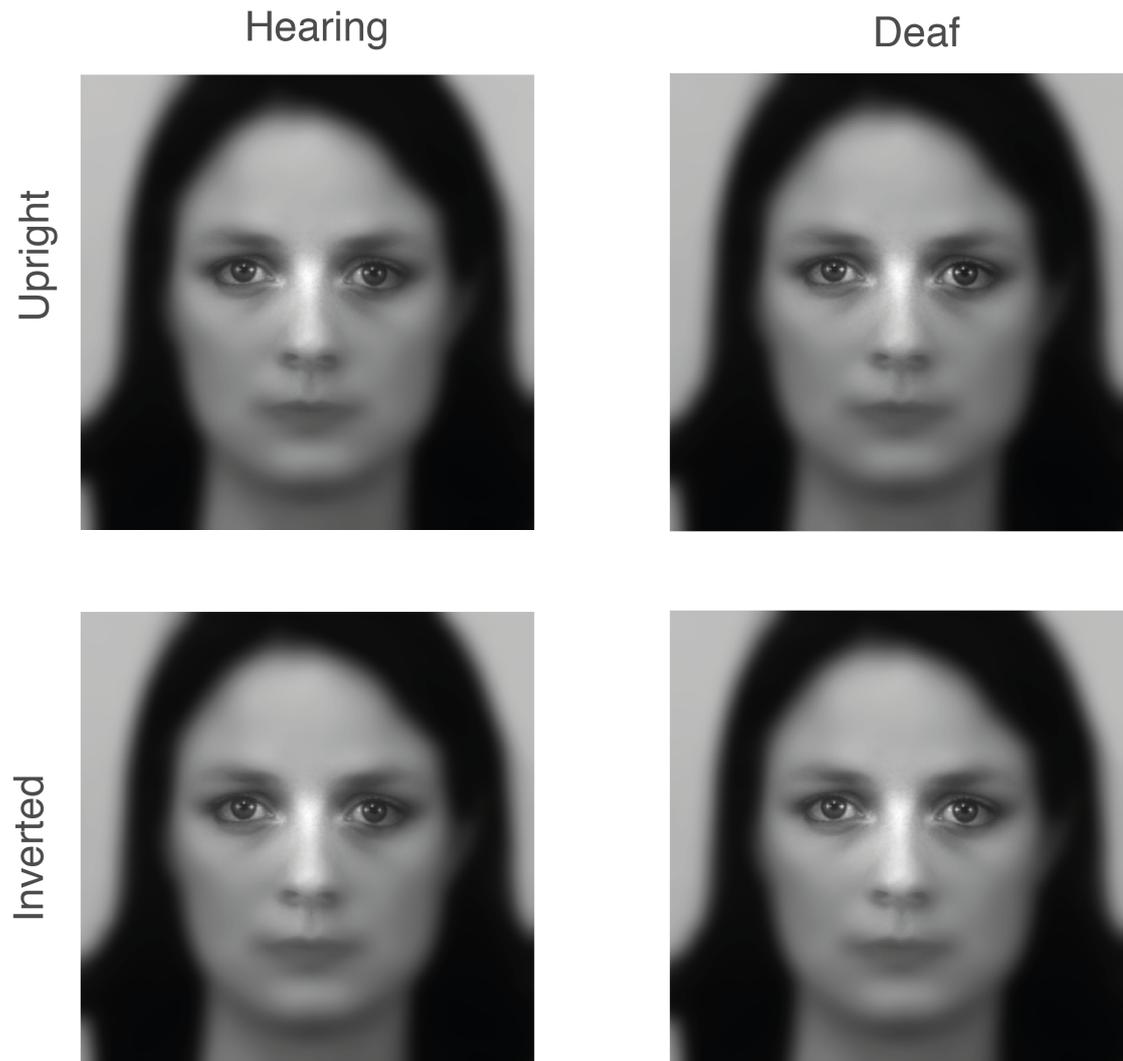
Note. Visual reconstruction of sampled information from one face stimulus using a retinal filter in hearing and non-hearing observers with either upright or inverted faces.

4.5   DISCUSSION

In the present study, we aimed to investigate facial information use in non-hearing observers during a face recognition. By recording the eye movement of non-hearing and hearing participants while they performed a delayed-matching task of upright and inverted faces, we showed that non-hearing observers fixate similar facial features to hearing observers in upright Natural Viewing, with slightly more spread scan paths. However, non-hearing participants were less affected by face inversion, as they show a smaller reduction of response accuracy in the inverted condition than the hearing observers, *i.e.*, non-hearing participants did not exhibited a FIE. Moreover, they do not change their fixation strategy between upright and inverted conditions, whereas hearing participants tend to fixate more at the mouth of the inverted faces. Importantly, by using an Expanding Spotlight paradigm, we found that non-hearing participants systematically use more peripheral information from the face and display on average longer fixation durations than hearing controls, indicating an enlarged visual span, for both upright and inverted faces. Overall, these result indicate that non-hearing individuals may be using a more robust eye movement and behavioural strategy, with a tendency to sample a greater quantity of facial information, during face identification tasks.

Two studies had previously also reported behavioural differences between non-hearing and hearing observers in FIE (De Heering et al., 2012; H. He et al., 2016). However, the behavioural measures reported in these two studies were not the same. De Heering et al. (2012) reported a greater reaction time difference between the upright and inverted condition in non-hearing than hearing observers, whereas H. He et al. (2016) reported a smaller D' differences in non-hearing participants. Altough they did not report statistics on the reaction time, H. He et al. (2016) highlighted non-hearing observers to respond slightly faster than hearing observers with both upright and inverted faces. It is thus crucial to note the task differences here, as similar task-dependent enhancement or impairment effect of hearing loss had previously been reported at least for visual tasks using low-level visual stimuli (Bavelier et al., 2006). In the current experiment, our task is comparable to the ones used in H. He et al. (2016), as participants saw two face stimuli in sequence and had to determine whether the face identities were the same. Accordingly, we found the FIE in identification performance (*e.g.*, D') to be smaller for non-hearing

participants as compared to hearing observers. Moreover, we found that non-hearing observers were slower to respond on average. However and crucially, no FIE differences were shown between the two groups. Altogether, these behavioural patterns are in line with the previously reported strategy differences between signers (both non-hearing and hearing) and non-signers (Stoll et al., 2018), highlighting, hence, the importance to report all behavioural measurements.

A couple of previous studies also reported fixation differences between non-hearing and hearing observers. For instance, early studies showed that non-hearing signers fixate more near the mouth during conversations, as the mouth region rest in the centre of the signing space (De Filippo & Lansing, 2006). By comparing non-hearing signer and hearing non-signers, Letourneau and Mitchell (2011) reported that non-hearing observers evenly fixated on the top and bottom half of the face in both identity and expression recognition tasks, whereas hearing observers distributed their fixation differently. We also observed a similar pattern in the current experiment, as our non-hearing participants fixated more evenly on the facial features (eyes and mouth) than hearing observers. Moreover, hearing controls change their fixation strategy between upright and inverted conditions, by fixating more towards the mouth in the inverted condition. In contrast, the inversion effect on fixation patterns was much weaker in non-hearing participants. This result indicated that the fixation strategy in the non-hearing observers is more robust than those of hearing observers.

Observers with enhanced attention towards the peripheral vision, such as participants from an East Asian culture, often display a consistent fixation strategy bias towards the centre of the face (*e.g.*, Blais et al., 2008; for a review, see Caldara, 2017). Moreover, using gaze-contingent paradigm it has been demonstrated that the central fixation bias in East Asian observers is driven by a specific perceptual tuning sampling different spatial frequency information from the same facial features. More particularly, they preferentially sample low spatial frequency information from the eyes and mouths through central fixations (Caldara et al., 2010; Miellet et al., 2012; Miellet et al., 2013). However, in the case of hearing loss, our results revealed that non-hearing participants display a triangular fixation pattern towards eyes and mouth as typically observed in Western Caucasian participants, despite their enhanced attention to peripheral vision. Moreover, by using the expanding spotlight gaze-contingent technique gradually revealing facial information, we found that non-hearing observers have a larger

perceptual span than hearing observers. Crucially, the visual field reconstruction using a retinal filter further demonstrates that non-hearing observers uniformly required more visual information from all facial features during face recognition.

Altogether behavioural and eye movements results revealed that hearing loss while lacking the sensory channel for audition results in a visual channel enhanced. However, this visual information intake enhancement seems to require a greater processing time while ensuring the robustness of visual information sampling towards altered stimuli, such as inverted faces.

One of the main limitations of the current study is that we only tested non-hearing signers and hearing non-signers. Thus it is unclear whether the differences reported here are caused by the practice of sign language of hearing loss. Indeed, sign language plays a central role in the non-hearing community and has a profound impact on the everyday visual experience of non-hearing individuals. While it might be difficult to completely disentangle the effect of sign language from the effect of hearing loss, future studies that include additional hearing signer observers will shed light on this question.

4.6   CONCLUSION

In conclusion, we reported an eye movement study assessing non-hearing signer and hearing non-signer participants in a face-matching task with upright and inverted stimuli. Our result shows that both behavioural response and eye movement patterns are more stable between the upright and inverted faces in non-hearing than hearing observers. Moreover, using an Expanding Spotlight paradigm, we showed that non-hearing participants uniformly requires more facial information intake resulting in a larger visual span than the hearing observers. Combining with the behavioural result, it suggests that non-hearing observers employ a conservative and robust strategy. Future studies should be conducted with additional hearing signer controls to investigate further how hearing loss and practising sign language jointly shaped this visual behavioural strategy.

4.7   REFERENCES

Arnold, P., & Murray, C. (1998). Memory for faces and objects by deaf and hearing signers and hearing nonsigners. *Journal of psycholinguistic research*, *27*(4), 481–497.

Bavelier, D., Brozinsky, C., Tomann, A., Mitchell, T., Neville, H., & Liu, G. (2001). Impact of early deafness and early exposure to sign language on the cerebral organization for motion processing. *Journal of Neuroscience*, *21*(22), 8931–8942.

Bavelier, D., Dye, M. W., & Hauser, P. C. (2006). Do deaf individuals see better? *Trends in cognitive sciences*, *10*(11), 512–518.

Bettger, J. G., Emmorey, K., McCullough, S. H., & Bellugi, U. (1997). Enhanced facial discrimination: Effects of experience with american sign language. *Journal of Deaf Studies and Deaf Education*, 223–233.

Blais, C., Jack, R. E., Scheepers, C., Fiset, D., & Caldara, R. (2008). Culture shapes how we look at faces. *PLoS ONE*, *3*(8), e3022. https://doi.org/10.1371/journal.pone.0003022

Blank, H., Anwander, A., & Von Kriegstein, K. (2011). Direct structural connections between voice-and face-recognition areas. *Journal of Neuroscience*, *31*(36), 12906–12915.

Bosworth, R. G., & Dobkins, K. R. (2002). Visual field asymmetries for motion processing in deaf and hearing signers. *Brain and cognition*, *49*(1), 170–181.

Buckley, D., Codina, C., Bhardwaj, P., & Pascalis, O. (2010). Action video game players and deaf observers have larger goldmann visual fields. *Vision research*, *50*(5), 548–556.

Caldara, R. (2017). Culture reveals a flexible system for face processing. *Current Directions in Psychological Science*, *26*(3), 249–255. https://doi.org/10.1177/0963721417710036

Caldara, R., & Miellet, S. (2011). Imap: A novel method for statistical fixation mapping of eye movement data. *Behavior Research Methods*, *43*(3), 864–878. https://doi.org/10.3758/s13428-011-0092-x.

Caldara, R., Zhou, X., & Miellet, S. (2010). Putting culture under the "spotlight" reveals universal information use for face recognition. *PLoS ONE*, *5*(3), e9708. https://doi.org/10.1371/journal.pone.0009708.

Capretto, T., Piho, C., Kumar, R., Westfall, J., Yarkoni, T., & Martin, O. A. (2020). Bambi: A simple interface for fitting bayesian linear models in python.

Codina, C., Buckley, D., Port, M., & Pascalis, O. (2011). Deaf and hearing children: A comparison of peripheral vision development. *Developmental science*, *14*(4), 725–737.

Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The eyelink toolbox: Eye tracking with matlab and the psychophysics toolbox. *Behavior Research Methods, Instruments, & Computers*, *34*(4), 613–617. https://doi.org/10.3758/BF03195489

De Filippo, C. L., & Lansing, C. R. (2006). Eye fixations of deaf and hearing observers in simultaneous communication perception. *Ear and hearing*, *27*(4), 331–352.

De Heering, A., Rossion, B., & Pascalis, O. (2012). Early deafness increases the face inversion effect but does not modulate the composite face effect. *Frontiers in psychology*, *3*, 124.

Dye, M. W., Hauser, P. C., & Bavelier, D. (2009). Is visual selective attention in deaf individuals enhanced or deficient? the case of the useful field of view. *PloS one*, *4*(5), e5640.

Dye, M. W., Seymour, J. L., & Hauser, P. C. (2016). Response bias reveals enhanced attention to inferior visual field in signers of american sign language. *Experimental brain research*, *234*(4), 1067–1076.

Fine, I., Finney, E. M., Boynton, G. M., & Dobkins, K. R. (2005). Comparing the effects of auditory deprivation and sign language within the auditory and visual cortex. *Journal of cognitive neuroscience*, *17*(10), 1621–1637.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. CRC press.

He, H., Xu, B., & Tanaka, J. (2016). Investigating the face inversion effect in a deaf population using the dimensions tasks. *Visual Cognition*, *24*(3), 201–211.

Hoffman, M. D., Gelman, A. et al. (2014). The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res., 15*(1), 1593–1623.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in psychtoolbox-3? *Perception, 36.* https://doi.org/10.1068/v070821.

Lao, J., Miellet, S., Pernet, C., Sokhn, N., & Caldara, R. (2017). Imap4: An open source toolbox for the statistical fixation mapping of eye movement data with linear mixed modeling. *Behavior Research Methods, 49*(12), 559–575. https://doi.org/10.3758/s13428-016-0737-x

Letourneau, S. M., & Mitchell, T. V. (2011). Gaze patterns during identity and emotion judgments in hearing adults and deaf users of american sign language. *Perception, 40*(5), 563–575.

Liddell, S. (2003). *Grammar, gesture, and meaning in american sign language.* Cambridge.

Lore, W. H., & Song, S. (1991). Central and peripheral visual processing in hearing and non-hearing individuals. *Bulletin of the Psychonomic Society, 29*(5), 437–440.

Lundqvist, D., Flykt, A., & Öhman, A. (1998). The karolinska directed emotional faces (kdef).

McCullough, S., & Emmorey, K. (1997). Face processing by deaf asl signers: Evidence for expertise in distinguishing local features. *Journal of Deaf Studies and Deaf Education,* 212–222.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746–748.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics, 21*(6), 1087–1092.

Miellet, S., He, L., Zhou, X., Lao, J., & Caldara, R. (2012). When east meets west: Gaze-contingent blindspots abolish cultural diversity in eye movements for faces. *Journal of Eye Movement Research, 5*(2). https://doi.org/10.16910/JEMR.5.2.5

Miellet, S., Vizioli, L., He, L., Zhou, X., & Caldara, R. (2013). Mapping face recognition information use across cultures. *Frontiers in Psychology, 4*(February), 34. https://doi.org/10.3389/fpsyg.2013.00034

Neville, H. J., & Lawson, D. (1987). Attention to central and peripheral visual space in a movement detection task. iii. separate effects of auditory deprivation and acquisition of a visual language. *Brain research, 405*(2), 284–294.

Pelli, D. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437–442. https://doi.org/10.1163/156856897X00366.

Proksch, J., & Bavelier, D. (2002). Changes in the spatial distribution of visual attention after early deafness. *Journal of cognitive neuroscience, 14*(5), 687–701.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137–149. https://doi.org/10.3758/BF03207704.

Stein, B. E. (2012). *The new handbook of multisensory processing.* Mit Press.

Stivalet, P., Moreno, Y., Richard, J., Barraud, P.-A., & Raphel, C. (1998). Differences in visual search tasks between congenitally deaf and normally hearing adults. *Cognitive Brain Research, 6*(3), 227–232.

Stoll, C., Palluel-Germain, R., Caldara, R., Lao, J., Dye, M. W., Aptel, F., & Pascalis, O. (2018). Face recognition is shaped by the use of sign language. *The Journal of Deaf Studies and Deaf Education, 23*(1), 62–70.

Thewissen, J. G., & Nummela, S. (2008). *Sensory evolution on the threshold: Adaptations in secondarily aquatic vertebrates.* Univ of California Press.

Yin, R. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81*(1), 141.

5

Watchers Do Not Follow the Eye Movements of Walkers

Watchers do not follow the eye movements of Walkers

M. Papinutto[a,b,*], J. Lao[a], D. Lalanne[b], R. Caldara[a]

[a] Eye and Brain Mapping Laboratory (iBMLab), Department of Psychology, University of Fribourg, Switzerland
[b] Human-IST Institute, Department of Informatics, University of Fribourg, Switzerland

A R T I C L E   I N F O

5.1 Abstract

Eye movements are a functional signature of how the visual system effectively decodes and adapts to the environment. However, scientific knowledge in eye movements mostly arises from studies conducted in laboratories, with well-controlled stimuli presented in constrained unnatural settings. Only a few studies have attempted to directly compare and assess whether eye movement data acquired in the real world generalize with those in laboratory settings, with same visual inputs. However, none of these studies controlled for both the auditory signals typical of real-world settings and the top-down task effects across conditions, leaving this question unresolved. To minimize this inherent gap across conditions, we compared the eye movements recorded from observers during ecological spatial navigation in the wild (the Walkers) with those recorded in laboratory (the Watchers) on the same visual and auditory inputs, with both groups performing the very same active cognitive task. We derived robust data-driven statistical saliency and motion maps. The Walkers and Watchers differed in terms of eye movement characteristics: fixation number and duration, saccade amplitude. The Watchers relied significantly more on saliency and motion than the Walkers. Interestingly, both groups exhibited similar fixation patterns towards social agents and objects. Altogether, our data shows that eye movements patterns obtained in laboratory do not fully generalize to real world, even when task and auditory information is controlled. These observations invite to caution when generalizing the eye movements obtained in laboratory with those of ecological spatial navigation.

5.2 INTRODUCTION

The human visual system is a complex and sophisticated machine that allows human beings to effectively process the environment and extract useful information for adapted spatial navigation and social interactions. The ocular motor system plays a critical role by producing a fine-tuned combination of muscle movements to orientate gaze to regions of interest, via a sequence of fixations and saccades feeding the visual system with diagnostic information. However, it remains unclear what the precise top-down and bottom-up mechanisms are that drive eye movements and their fine-tuned interplay to perceive and process the visual environment.

Since the very first eye-tracking studies, it was clearly demonstrated that eye movements do not land randomly on the visual input space, but rather reflect an efficient, near optimal, sampling of diagnostic information (*e.g.*, Buswell, 1935; C. H. Judd, 1905a; Stratton, 1902; Yarbus, 1967). In face perception research, for example, eye movement studies revealed that eye movements land on faces' diagnostic information (Gosselin & Schyns, 2001; Schyns et al., 2002), which flexibly adjusts on task constraints (*e.g.* Geangu et al., 2016; Jack et al., 2009; Kanan et al., 2015), information quantity (Caldara et al., 2010; Miellet et al., 2012; Papinutto et al., 2017; for a review see Caldara, 2017) and quality (Miellet et al., 2011), culture (Blais et al., 2008; Kelly, Liu, et al., 2011; Miellet et al., 2013; Rodger et al., 2010), and other higher-level effects such as context and prior knowledge. Interestingly, it has been very recently demonstrated that such idiosyncratic fixation patterns finely tune face sensitive neural responses (Stacchi et al., 2019). However, the processing of scenes, which inherently involve more variable inputs, is by far more complex and less understood.

Many laboratory studies have clearly shown that visual scene processing results from the combination of eye movements guided by low-level saliency information (*e.g.*, color, luminance, contrast and intensity—(Itti & Koch, 2000, 2001; Koch & Ullman, 1985; for a review, see Borji & Itti, 2013) and top-down cognitive processes using prior knowledge and expectations (*e.g.* Malcom & Henderson, 2009; Rao et al., 2002; Zelinsky, 2008). In ecological spatial navigation contexts, eye movements land on low-level salient information to rapidly filter visual scenes, as significantly predicted by bottom-up image-driven saliency models (*e.g.*, Peters et al.,

2005; for a recent review, see Riche & Mancas, 2016a, 2016b). However, ecological spatial navigation also heavily involves top-down processes, which are more difficult to take into account in laboratory experiments. Indeed, in laboratories, participants are in steady and protected environments and often processing pre-defined visual scenes. Nevertheless, more recently, such processes have started to be accounted for by neuro- and computer scientists implementing simultaneously bottom-up and top-down factors in the saliency models (*e.g.,* Voorhies et al., 2012). The results are very promising but predicting eye movement patterns of the visual sampling of scenes remains one of the greatest challenges in the understating of human vision. It is thus important to investigate the contribution of each mechanism in controlled laboratory settings. But, perhaps, it is also even more important to validate whether laboratory results generalize to real world situations in the case of ecological spatial navigation.

Only a few studies have attempted to shed light on the differences between laboratory and real-world settings. In the real world, walking participants are more likely to exhibit larger saccades due to free head movement (Stahl, 1999) and have predicted smaller fixation durations (Foulsham, Walker, et al., 2011). There are also differences in terms of sensitivity to saliency ('t Hart et al., 2009), as well as in motion effects. In fact, such motion effects impact on where eye movements are directed, eliciting differences in fixation locations between laboratory and real-world conditions (Hillstrom & Yantis, 1994; Lappi, 2015, 2016). While these experiments highlight the differences between real-world and laboratory settings, it should also be noted that many similarities eye movement patterns occur in these conditions. For example, similar fixation patterns on foveated faces were found (Peterson et al., 2016) and both groups exhibited similar central bias (Foulsham, Walker, et al., 2011).

While these studies shed light on the differences and similarities between laboratory and real-world conditions, they rarely evaluated the sensitivity to saliency, motion or other predictive models of eye movements. Additionally, they might include confounded effects due to differences in tasks (Foulsham, Walker, et al., 2011), visual conditions (Peterson et al., 2016), or acoustical information (conventionally there is no acoustical information provided in laboratory settings despite this channel feeds the visual system— Coutrot & Guyader, 2014; Coutrot et al., 2012). Other aspects might also induce differences, such as head movements in real-world conditions (Pelz et al., 2001; 't Hart et al., 2009), technical issues due to parallax errors in

mobile eye tracking (Evans et al., 2012) and differences in reference frame. In fact, the point of fixation in laboratory conditions reflects precisely the fixation location on the computer screen whereas in real-world conditions the point of fixation assessed by the eye-tracking glasses is a reference gaze point drew on a virtual plane (usually the camera recording) artificially drawn in the 3D space (see Figure 5.2.1—Lappi, 2015, 2016).

Figure 5.2.1

Schematic representation of differences in fixation location when using eye-tracking glasses in the wild
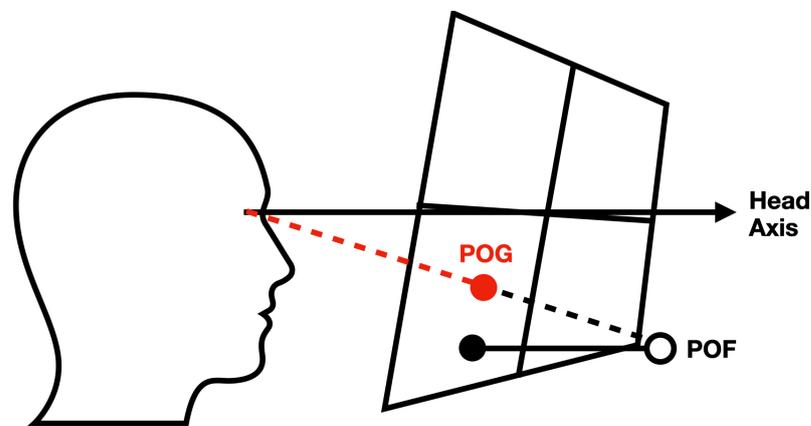
Note. Schematic representation of differences in reference frame in mobile eye tracking. The black arrow represents the orientation of the glasses and is crossing the virtual plane. On this plane, there are two dots: the red one indicates the point of gaze (POG) and the black one is the projection of the point of fixation (POF, empty point). Adapted from Lappi (2015).

As a consequence, differences found between laboratory and real-world conditions might be induced by the experiment itself rather than representing genuine differences between those settings. Other variations could arise from top-down processes that are more engaged in real-world activities and lead to more active sampling strategies than those deployed in laboratory conditions (Hayhoe et al., 2012; Pelz et al., 2000). Indeed, routine real-world activities involve a series of subtasks, such as obstacle avoidance or the planning of the next foot-

steps by looking two steps ahead during real-world walking (Hollands & Marple-Horvat, 2001; Marigold & Patla, 2007; Matthis et al., 2018). Moreover, real-world conditions allow multi-sensory integrations, whereas laboratory conditions usually provide unimodal visual information with simple cognitive tasks. This could be problematic, as multi-sensory integrations conjointly with anticipatory strategies convey multiple environmental indices to precisely guide eye movements on diagnostic information for navigation (Lappi, 2015, 2016).

Hence, a fixation landing position in a real-world condition results from a complex combination of multi-sensorial information integration and anticipatory mechanisms. Such combinations make the rationale of a fixation location in real-world condition difficult to assess. On the contrary, laboratory studies allow a fine tuning of experimental conditions in order to constrain and control the amount and the type of information available for navigation. Such control over information availability allows to precisely evaluate the rationale behind a fixation location while underestimating the impact of other sources of information available in the real world.

Altogether, these studies critically suggest that it remains unclear whether eye movements used in laboratory conditions are similar to those deployed in real-world conditions. In addition, it remains to be determined whether the contribution of top-down and bottom-up processes to guide eye movements is comparable between both conditions. Therefore, it remains to be clarified whether the results and conclusions obtained from eye movements laboratory studies can be generalized to the real world.

To address these issues, we directly compared the eye movements deployed by observers in the real world (the Walkers) with those obtained by observers in laboratory (the Watchers). To avoid potential confounds driven by differences in top-down processes across experiments (i.e., natural vs laboratory settings), we instructed the Walkers and the Watchers to perform the same cognitive task. The Walkers were instructed that questions about the walking path will be asked at the end of the walk. The Watchers performed the very same task with the same visual and auditory inputs obtained from the Walkers. Raw data was preprocessed with a data-driven method based on a common angular speed threshold (75th percentile), allowing us to categorize eye movements according to the inherent idiosyncratic differences of observers, as

well as the technical differences driven by the different eye trackers used in the wild and the laboratory. We then used gaze maps from both conditions and the *Normalized Scan path Saliency* score (NSS), which is a score that indicates whether an eye movement landed on salient region or not (Bylinskii et al., 2016). Conjointly with a leave-one-out procedure, this allowed us to probe the comparability between real-world and laboratory eye movements, as well as the sensitivity towards bottom-up saliency and scene motion. Moreover, to validate the robustness of our results, we ensure that the group results were consistently reproduced across the Walkers' videos and across the first video watched by the Watchers. To the best of our knowledge, these data-driven and robust approaches were not previously used in the evaluation of ecological eye movements acquired in the wild. Such approach also offers the possibility to thoroughly identify differences and similarities across both the laboratory and real-world conditions.

5.3 METHODS

5.3.1 PARTICIPANTS

Twelve participants—11 females—aged from 20 to 29 years old (M= 22.83, SD= 2.79) took part in the real-world condition; this group will be subsequently called the Walkers. Twenty participants—14 females—aged from 19 to 37 years old (M= 23.65, SD= 4.73) were tested in the laboratory; this group will be subsequently called the Watchers. All participants were students at the University of Fribourg, Switzerland for <2 years to ensure medium knowledge of the walking path. We evaluated potential participants' knowledge of the campus before the experiment on a scale from "0" (I don't know this path) to "10" (I know this path very well). We only tested participants with an answer below 8. The Walkers had a knowledge of 4.20 on average (SD= 1.87) and the Watchers had a knowledge of 4.05 on average (SD = 3.19) and neither group differed in path knowledge, $t(28) = 0.14$, $p > .05$; $BF_{10} = .36$. If the Watchers experienced a severe motion sickness, they were discarded. A total of 4 subjects reported severe motion sickness and were dropouts during the experiment. All participants had normal or corrected-to-normal vision. Participants received course credits for completing the experi-

ment. All participants gave oral informed consent and the protocol was approved by the ethical committees of the Department of Psychology of Fribourg University, Switzerland.

5.3.2   ROUTE

The route consisted of a walking path of about 200 meters inside the building of the Department of Psychology of the University of Fribourg, Switzerland. This route encompassed corridors, stairs and doors which were surrounded by offices, posters, and billboards (see Figure 5.3.1). Participants were tested during the working hours leading to high probabilities of encountering people inside the building.

Figure 5.3.1

Representation of the walking path of the Walkers
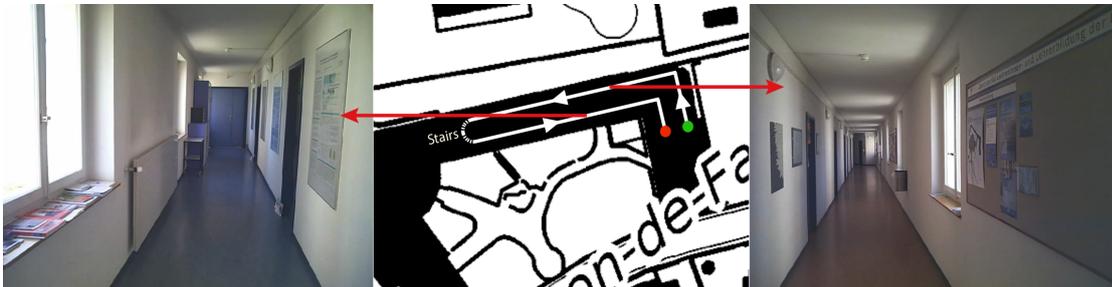
Note. Representation of the walking path that Walkers took. Left and right pictures show capture frame from the environment at two different stages of the path.

5.3.3   MATERIALS

We extracted the scene videos from the mobile eye tracker. The videos were cut to have the same starting and ending location. We also discarded 2 videos where the walking speed was

too fast or too slow (about 17% of the total number of videos), resulting in 10 videos of varying duration (M= 3 min 48 sec, SD= 23 sec) and idiosyncratic walking pace. We then presented the videos with monophonic sounds on a ViewPixx 3D monitor (1440 x 1080-pixel resolution), subtending 44.80° horizontally and 33.60° vertically of visual angle. All videos were presented at a distance of 50cm which is the minimal distance allowed to record eye movements with the SR Research Desktop-Mount EyeLink 2K eye tracker. These distances and sizes were chosen so that the presentation in the laboratory resembled as much as possible the real-world condition. The closest ratio between visual angle size allowed by our setting was: 1.43 vertically and 1.38 horizontally.

5.3.4   APPARATUS

Eye movements in the real-world condition were recorded with SMI eye-tracking glasses 2.0 (ETG) at a sampling rate of 60 Hz. ETG has a tracking range of 60° vertically and 80° horizontally, and an average gaze position error radius of about .5°. Although viewing was binocular, only the dominant eye was tracked. We calibrated the eye glass at the beginning of the experiment using a three-point fixation procedure as implemented in the SMI API (see SMI Manual). Calibrations were validated visually by the experimenter and repeated until reaching an optimal calibration. The external scene camera in the SMI ETG recorded First Person Point of View (POV) videos at 30 frames per second (fps). The video resolution was 720 pixels vertically and 960 horizontally, evaluated to sustain 48° and 62° of visual angle, respectively.

In the laboratory condition, eye movements were recorded at a sampling rate of 1000 Hz with the SR Research Desktop-Mount EyeLink 2K eye tracker (with a chin and forehead rest), which has an average gaze position error of about .5° and a spatial resolution of .01°. Although viewing was binocular, only the dominant eye was tracked. The experiment was implemented in MatLab (The MathWorks, Natick, MA, USA), using the Psychophysics toolbox (PTB-3) (Kleiner et al., 2007; D. Pelli, 1997) and EyeLink Toolbox extensions (Cornelissen et al., 2002; Kleiner et al., 2007). Videos were presented with their acoustical background through speakers on each side of the screen in order to avoid differences due to the influence of sound

(Coutrot & Guyader, 2014; Coutrot et al., 2012). Calibrations of the eye tracker were conducted at the beginning of the experiment using a nine-point fixation procedure and repeated when necessary until reaching an optimal calibration criterion. At the beginning of each video, participants were instructed to fixate at the central fixation cross. If the eye drift was more than 1°, a new calibration was launched to ensure an optimal recording quality.

5.3.5 PROCEDURE

Both groups completed a socio-demographic survey. The Walkers were instructed to walk through the path provided on a printed sheet (as described above). The Walkers did not carry this sheet with them. The building has the same structure at all floors and participants managed to navigate properly across the planned path. When a problem occurred, the participants were instructed to pursue their route normally and ask their potential questions out loud. The Watchers were instructed to watch the POV videos of the Walkers. Participants were informed that questions about the route would be asked at the end of the experiment. This was done to ensure that all participants (real-world and laboratory conditions) actively attended to the environment.

We then set up and calibrated the eye-tracking device (the SMI ETG for the Walkers and the SR Research EyeLink 2K for the Watchers). In the real-world condition, the Walkers walked through the path at their usual walk pace, whereas in the laboratory condition the Watchers watched the 10 POV videos with sounds in a random order, with breaks between each video if needed.

At the end of the experiment, participants were told that no questionnaire needed to be fulfilled, were thanked, and dismissed. On average the experiment in the real-world condition lasted about 15 minutes and the laboratory condition lasted for about 50 minutes on average.

5.3.6 DATA ANALYSIS

Raw eye movements were used as well as preprocessed data. Sampling rates were matched across conditions. Raw eye movements data was then preprocessed to extract fixations and saccades with a custom algorithm using a velocity threshold based on 75th percentiles of the velocity. Fixations that were too close spatially (<0.3°) and temporally (<20ms) were merged. We then computed the number of fixations, fixation duration, saccade number, saccade amplitude, and saccade orientation for each participant to quantify the general oculomotor behavior. To compare these eye movement characteristics between the Walkers and the Watchers, we performed Kolmogorov-Smirnov tests on the probability density functions.

To compare whether the Watchers and the Walkers looked at the same locations, we used the Normalized Scan path Saliency score (NSS) as it offers a good balance between false positives and false negatives (Bylinskii et al., 2016). The NSS score evaluates the correspondence between the normalized saliency map and the gaze. More specifically, the chance level is at 0, a negative score of NSS indicates anti-correspondence whereas a positive score of NSS suggests correspondence between eye movements and saliency map. Moreover, an NSS score above 1 indicates that the eye movements rely significantly on the normalized saliency map when it is; a score below 1 indicates that eye movements did not rely significantly on the saliency map. In the current study, the NSS score was used to compute the match between normalized maps and eye movement using 3 types of maps as saliency maps in the NSS algorithm: The Watchers gaze maps, the saliency maps and the motion maps.

Regarding the evaluation of the match between the Watchers and the Walkers, the NSS scores were computed for each frame, each video and each participant (both the Walkers and the Watchers) using a leave-one-out procedure on the Watchers and raw gaze data as used by Dorr and colleagues (For a justification of this method, see Dorr et al., 2010). This technique allowed us to compare eye movements of a Walker or a Watcher considering the eye maps of the rest of the Watchers. The NSS scores were then extracted according to eye movement locations, and the median was computed in order to have a single NSS score for each frame, each video and each participant. Furthermore, global eye movement maps were computed for each participant.

Saliency maps and motion maps were computed for each POV video and each frame to analyze their content and where eyes were attracted. The saliency maps provided information about salient regions in the scenery (i.e. region with high contrast, luminosity, and edges) whereas the motion maps described movement in the 3-dimensional space (i.e. change horizontally, vertically and over time). In this experiment, we selected the Dynamic Adaptive Whitening Saliency algorithm (AWS-D) from Leboran et al. (2017) whose previous static algorithm (Garcia-Diaz et al., 2012) obtained a good evaluation in the MIT benchmarks in T. Judd et al. (2012) and in Borji, Sihite, et al. (2013). Furthermore, the AWS algorithm was evaluated to have a low correlation with central bias (Nuthmann et al., 2017). AWS-D provides a dynamic approach to extract bottom-up saliency from video, considering temporality and performing better than most of the algorithms treating video saliency. Indeed, many algorithms are static and compute saliency from an image only, leading to bad saliency estimation in video. We used the MATLAB implementation of the AWS-D that the authors provided us (http://persoal.citius.usc.es/xose.vidal/research/aws/AWSmodel.html). Regarding the motion maps, they were obtained by computing the differential maps in horizontal axis, vertical axis and time (3D partial derivative). Following this, the value of each pixel was found by computing the K-invariant of the structure tensor (Vig et al., 2009).

To assess at what point participants relied on saliency and motion, NSS scores were computed for each participant (Watchers and Walkers). We used raw eye movements to gather maps value at eye positions and compute a median value for each frame, video, and viewer (For further information see Dorr et al., 2010).

All the above analyses were performed at the group level using either Kolmogorov-Smirnov tests to assess differences in eye movements characteristics distributions or Welch's t-tests to assess differences in NSS scores between the Watchers and the Walkers. To take into account differences in speed and walking paths across the Walkers, we estimated the statistical contribution of each Walkers' video on the group statistical effects. We carried out the same statistical analyses for each video independently, while correcting for multiple comparisons. In case of an impact we assessed if it could be attributed to the walking pace by computing a t-test on the duration of videos showing an effect and those that did not. Moreover, we controlled for the impact of Watchers habituation after seeing the 10 videos, by conducting the same analysis

on the first video seen by the Watchers.

To evaluate if and on which object the gaze of the Walkers and the Watchers converged, we used a data-driven peak detection of the saliency and motion median NSS scores of both the Walkers and the Watchers data. This analysis provided us with the frames of each of the Walkers' videos in which similarity usage of either saliency or motion was at the highest. We then ensured that such peak scores were present in both groups by visual inspection. Frames with a high NSS score for saliency and motion in both groups indicated convergence in gaze. The content of the frame in the video was then analysed and interpreted by visual inspection. Please note, we also provided the lowest peak values detection (see Figure 5.11.1). We also evaluated if one group fixated at those stimuli of interest before the other. To this aim, the frames of convergence were extracted from the previous analysis. Then, the video sequence preceding a frame of convergence was replayed to identify when each group landed their fixations on the object of interest. We then computed the time difference between both the Walkers and the Watchers to land on the object of interest. Finally, the resulting differences were statistically evaluated by performing corrected t-tests against zero.

5.4 Results

5.4.1 General Gaze Location

The NSS scores reveal that the Watchers significantly matched other Watchers (M= 1.99, SD= 1.57, median NSS score above 1) but not the Walkers (M= 0.69, SD= 1.68, median NSS score below 0 indicating anti-correspondence). Moreover, the difference in the Walkers and the Watchers NSS score was significant, $t(28)$ = 143.42, $p < .001$ (see Figure 5.4.1).

Figure 5.4.1

NSS score differences in gaze similarity, saliency and structure tensor between the Walkers and the Watchers
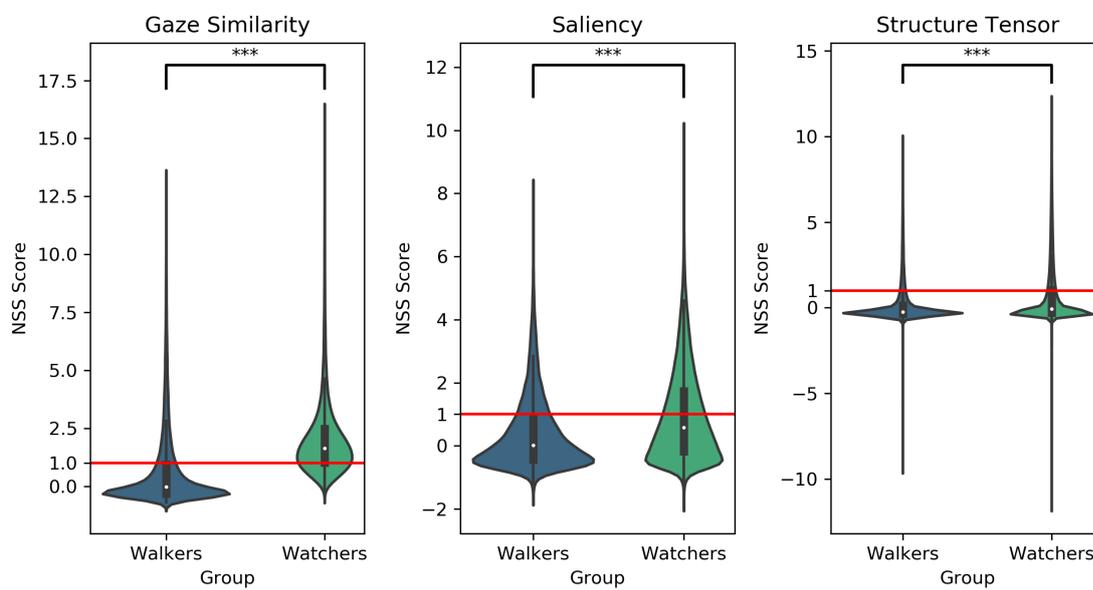
Note. Violin plots of the Watchers and the Walkers NSS scores computed on either (from left to right) gaze maps using a leave-one-out procedure on the watchers, the saliency maps or the motion maps. The red line indicates significance threshold for NSS Scores.

5.4.2   Fixations Characteristics

The Watchers differed significantly in fixation duration distribution from the Walkers who had overall lower fixation durations. Indeed, a Kolmogorov-Smirnov test indicated that the fixation duration distributions of the Watchers did not follow those of the Walkers' fixation duration distribution, $D = 0.11$, $p < .001$. Regarding the number of fixations, the Walkers made fewer fixations than the Watchers, $D = 0.87$, $p < .001$ (see Figure 5.4.2).

Figure 5.4.2

Difference in event characteristics between the Walkers and the Watchers
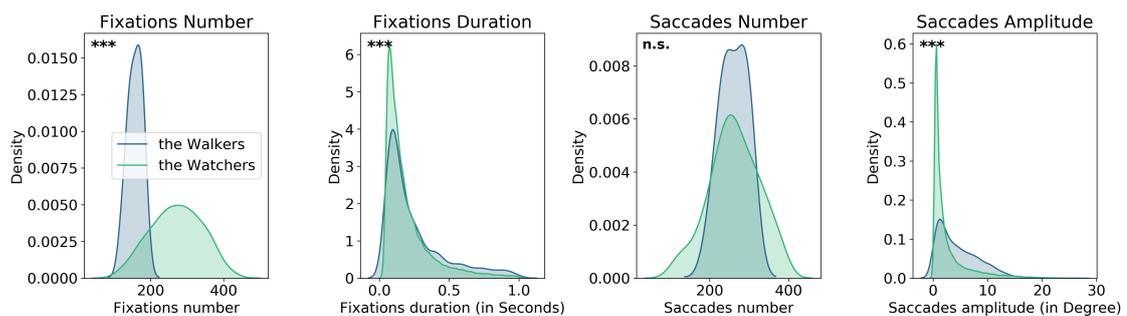
Note. Kernel density of fixation number and duration as well as saccade number, amplitude and duration for the Walkers (in blue) and the Watchers (in green).

5.4.3   Saccades Characteristics

The distribution of saccade amplitude for the Watchers did not follow those of the Walkers, $D = 0.32$, $p < .001$, as the Watchers deployed longer saccades to explore the visual scene. However, the Watchers and the Walkers did not show a difference in the number of saccades (see Figure 5.4.2).

5.4.4   Saccades Direction and Fixations Distribution

The saccade direction distribution shared a similar pattern between the Watchers and the Walkers, but the Watchers tended to direct their saccades more horizontally than the Walkers who directed their saccades more vertically; these differences were significant, $D = 0.05$, $p < .001$ (See Figure 5.4.3, but see also Figure 5.4.5 and section 5.4.6).

Figure 5.4.3

Gaze orientation differences between the Walkers and the Watchers
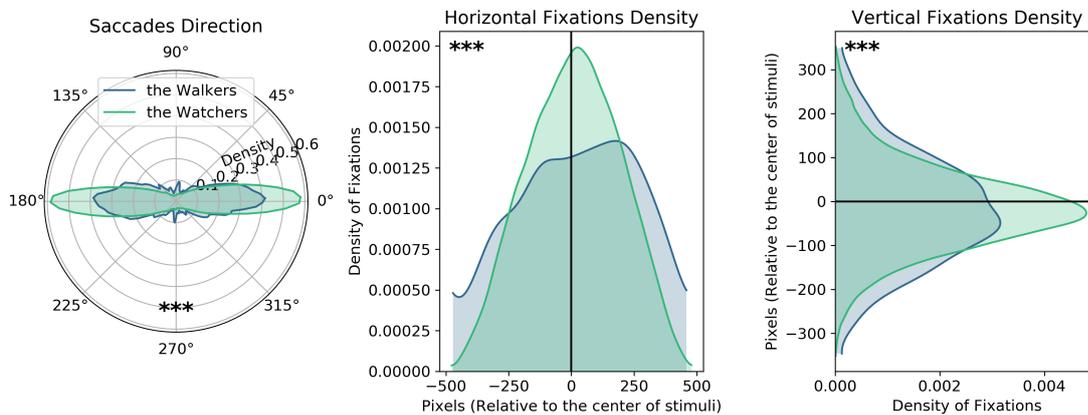
Note. Kernel density of saccades direction and Eye movements distribution for the vertical and horizontal direction for both the Walkers and the Watchers.

The eye movement direction distribution indicated a greater central bias for Watchers, even when both distributions were fitted towards a normalized screen size. Distributions differed significantly both vertically $(D = 0.10, p < .001)$ and horizontally $(D = 0.08, p < .001)$ between the Watchers and the Walkers (see Figure 5.4.3).

5.4.5   Saliency and Motion

The Watchers relied significantly more on saliency ($M = 0.94$, $SD = 1.44$) than the Walkers ($M = 0.39$, $SD = 1.15$), $t(28) = 117.78$, $p < .001$ (see Figure 5.4.1). Both groups showed correspondence with saliency (median score above 0) but to a low extent (median score below 1). A similar result was observed for motion, with the Watchers relying significantly more on motion ($M = 0.74$, SD $= 1.73$) than the Walkers ($M = 0.19$, $SD = 1.15$), $t(28) = 115.41$, $p < .001$. However, the median score of both groups was below 0, indicating that both groups did not, on average, rely on motion as a NSS score below 0 indicates anti-correspondence (see Figure 5.4.1).

Looking to highest NSS values reveals that social, written, and actionable stimuli (*e.g.,* door knock of the door that will be open) make both the Walkers and the Watchers to synchronize their fixation locations and rely on saliency and motion to a great extent (see Figure 5.4.4). The differences in time between groups to reach those objects of interest did not significantly differed from zero. This was the case for social, actionable and written stimuli.

5.4.6   Robustness of Group Results across the Walkers' Videos and the First Videos Watched by the Watchers

Overall, our group results hold. Indeed, a large majority of videos, if not all, revealed similar differences between the Walkers and the Watchers. This was the case for fixation number, vertical distribution of fixations, saccade number, saccade amplitudes as well as for saliency and motion maps. However, for some metrics, the effects were weaker: for the fixation durations (half of the videos showed a significant effect), the horizontal distributions of fixations (6 out of 10 videos showed an effect) and saccade orientation (only 2 videos showed an effect). Furthermore, we additionally evaluated if the differences in results were due to different walking speeds. Only the horizontal fixation distribution revealed significant differences as a function of the walking speed (i.e., video duration in seconds): differences in horizontal fixation distribution between the Walkers' and the Watchers were only found for the Walkers' videos with

Figure 5.4.4

Examples of video frames in which both groups eye movement locations converged

Saliency maps

| Walker/Video 5 Frame 598 | Walker/Video 1 Frame 5693 | Walker/Video 5 Frame 5541 |

Motion maps

| Walker/Video 7 Frame 4444 | Walker/Video 9 Frame 607 | Walker/Video 10 Frame 3334 |
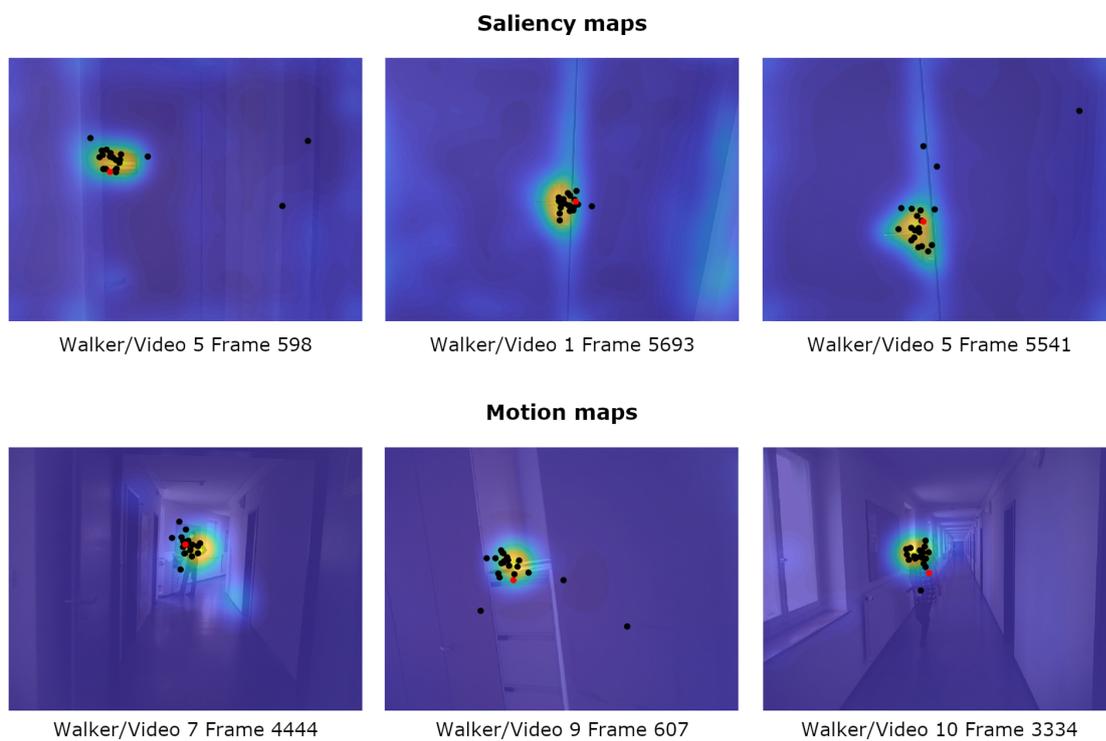
Note. Samples of frames with the highest NSS Score computed on saliency maps and motion maps for both the Walkers (red dot) and the Watchers (black dots).

a slow walking speed (*M*=237.53, *SD*=9.34); no differences were found for the fast walking speed (*M*= 205.37, *SD*= 22.31), $t(9)$= 3.14, *p*= .016.

Additionally, the group results were replicated when using only the first video watched by the Watchers. The only exception was for saccades orientation that was not found significant (see Figure 5.4.5).

Figure 5.4.5

Gaze orientation differences between the Walkers and the Watchers for the first video watched by the Watchers only
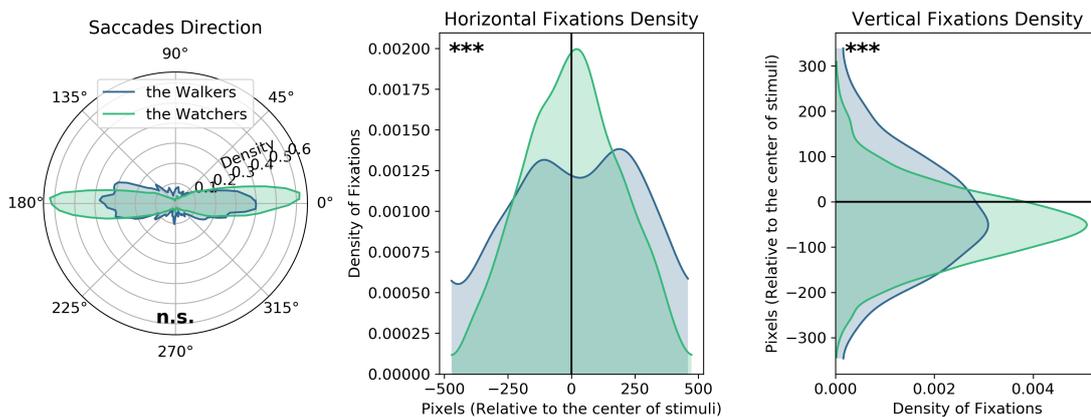
Note. Kernel density of saccades direction and Eye movements distribution for the vertical and horizontal direction for the first video watched by the Watchers.

5.5   DISCUSSION

This study aimed to determine the extent to which eye movements, deployed during visual scene processing in laboratory, generalize to real-world gaze behaviour. To this end, we compared the eye movements obtained from a group of observers in the real-world (the Walkers) with those obtained from observers in laboratory settings (the Watchers) on the same visual

and auditory inputs acquired form the Walkers. As such, these conditions were set to be as close as possible in terms of inputs between the groups. The contribution of bottom-up processes was evaluated by using the sensitivity to saliency and motion. Top-down processes' contribution due to task was equated as much as possible for the two groups with an active task, as both of them performed the same active task while exploring the walking path. Our data-driven preprocessing method based on a common angular speed threshold (75th percentile) across conditions allowed us to categorize eye movements into events according to the inherent idiosyncratic differences of observers, as well as the differences elicited by both the technical specifications of the eye-trackers and real-world and laboratory settings. Moreover, we controlled for the robustness of our results by evaluating for the impact of each of the Walkers' video on the group effects. In case a statistical effect was significant only for a particular Walkers' video, we assessed whether those effects could be attributed to differences in walking pace across the Walkers. The Watcher viewed many videos obtained from the Walkers. We thus also ensured that our group effects did not rely on particular Walkers' videos or repeated exposure of videos (*i.e.*, habituation), as the eye movements of the Watchers, contrary of those of the Walkers, were based on more than a single observation of the walking path. To rule out this potential confound, we also present the results of the first Walkers' videos watched by the Watchers. This allowed us to ensure that the differences found between conditions are robust, and due to the setting per se, and not to other potential flaws, such as differences induced by the methodology used.

Our data showed a significant different global gaze location between the Watchers and the Walkers. As in Foulsham, Walker, et al. (2011) and 't Hart et al. (2009), the Watchers exhibited a more focal central bias than the Walkers. This finding relates to the tendency of audio-visual material to present interesting content in the centre (see *e.g.* Dorr et al., 2010; Foulsham & Underwood, 2008; Tatler, 2007). In our study, the centrality of salient and moving content can also be observed in the general saliency and motion maps (see Figure 5.10.1). We also found robust differences in eye movements characteristics between the Walkers and the Watchers. On the contrary to Foulsham, Walker, et al. (2011), we found the Walkers to produce significantly less but longer fixations than the Watchers. This difference might arise from the Walkers engaging actively to navigate properly (*e.g.* they could have watched for their next step longer than the Watchers). Additionally, their saccades amplitudes were larger than the Watchers. As an

explanation, this can be due to free head movements, in the unconstrained settings of the real world as opposed to laboratory settings (Bahill, Adler, et al., 1975; Stahl, 1999). Those results were robust across the Walkers' videos and with only the first watched videos by the Watchers.

Other differences arising from navigation in the real world, *per se*, were found in the direction of saccades and in the vertical and horizontal distribution of fixations. The saccade directions differed significantly across group effects. Indeed, the Walkers tended to direct part of their saccades towards the bottom and the top part of the scene. At the group level, the Watchers instead oriented their gaze more horizontally. However, this result was not robust across the Walkers' videos nor with only the first watched videos by the Watchers. The group effect on saccade orientations was rather weak, the tendency of the Walkers to direct their gaze significantly more downward was also found in the vertical fixation distribution. Importantly, this result was robust across only the first videos watched by the Watchers and across the Walkers' videos. The rationale behind this oculomotor behaviour lies in the Walkers looking at their next footstep location as found in previous studies on locomotion (Hollands & Marple-Horvat, 2001; Marigold & Patla, 2007; Matthis et al., 2018; Patla & Vickers, 2003; but, see Foulsham, Walker, et al., 2011). The fixation distributions also significantly and robustly differed horizontally between both groups. Interestingly, this difference might have been rooted in speed differences across Walkers. While the fixation distribution of the Watchers remained constant, the Walkers exhibited a wider central bias when walking slowly, as compared to a faster walking pace. The slower the walking pace, the wider the central bias. This finding might relate with Hollands et al. (1995), who found that saccades amplitudes were related to strides. However, the Walkers' pace did not impact on the Watchers horizontal fixation distribution.

The eye movements characteristics also differed across the Walkers and the Wathcers. Indeed, despite using data-driven approach to preprocessed eye movements and controlling for both the auditory signals and the top-down task effects across conditions, we found a higher number of divergent oculomotor events between the real world and the laboratory conditions than previous studies (*i.e.* Foulsham, Walker, et al., 2011; 't Hart et al., 2009). Crucially, and on the contrary to previous studies, these differences were robustly replicated across both the first videos watched by the Watchers as well as across the Walkers' videos. Although these differences in eye movements could be attributed to differences across experimental settings, such as

the use of monophonic auditory signals in the laboratory, we do not think that the use of stereo signals and other factors would be sufficient to abolish such differences. We thus genuinely believe that differences will always persists between the ecological acquisition of eye movements in the wild and those artificially acquired in constrained laboratory settings.

To evaluate the contribution of bottom-up processes, we used a saliency algorithm developed especially for video contents. To the best of our knowledge, such saliency algorithm has never been used in this the present framework. Rather, previous studies used static saliency algorithms ('t Hart et al., 2009) or did not evaluated saliency (Foulsham, Walker, et al., 2011). Our analysis revealed that the Watchers' eye movements matched significantly – and robustly across the Walkers' videos and across only the first video watched by the Watchers – more with the saliency maps than those of the Walkers. This result differed from 't Hart et al. (2009) findings, that saliency models were a weak predictor of both, the real-world and the laboratory conditions, as in our study the saliency reasonably predicted the Watchers' eye movements. Instead, our findings are consistent with Henderson et al. (2007) who concluded that in the real world visual saliency is a less effective predictor of eye movements. The discrepancy in those results can be imputed to the differences in the algorithms used (static vs dynamic) but nevertheless highlights the necessity to apply a saliency algorithm specifically dedicated to stimuli types.

In addition to the dynamic saliency algorithm, we complemented our results with motion maps using the computation of the K-invariant of the structure tensor, which was never used in this context. This technique, despite not being completely independent from our dynamic saliency algorithm, provides a deeper focus on motion, which is one of the factors playing a key role in attracting eye movement fixations (see *e.g.*, Yantis & Jonides, 1984). Importantly, with this approach, the motion resulting from walking is held out (as being constant across the whole frame), whereas motion in the world should pop out. As such, the motion maps essentially measure the movements resulting from world agent. As such, the motion maps essentially measure the movement resulting from world agent. Similarly to saliency, we found the Watchers to significantly and robustly match more the motion maps prediction than the Walkers. However, the motion maps were weak predictors of both the eye movements in the real world and in the laboratory. This indicates that solely motion originating from objects in the scenery does not suffice to predict eye movements, both in the laboratory and in the wild.

The differences in the prediction of the Walkers' and the Watchers' eye movements by saliency and motion maps shed light on the differences in information use, as well as the available attentional resources across both conditions. Indeed, the laboratory settings might have required lower amount of attentional resources, given the absence of real-world constraints and the involvement of implicit or explicit anticipatory strategies typical of natural walking (Hayhoe et al., 2012; Hillstrom & Yantis, 1994; Pelz et al., 2001). As such, the Watchers have more time and, resources to allocate their attention towards salient and moving areas than the Walkers, who had to predict expected and unexpected events to navigate properly, as well as to predict their walking path. This continuous walking planification results in a series of subtasks such as collision avoidance (Lappi, 2015).

It is important to point out that our data also shows similarities between both groups. Surprisingly, despite the great number of differences in eye movements characteristics, both the Walkers and the Watchers performed a similar number of saccades. This absence of effect was found across groups, across the Walkers' videos and with only the first video watched by the Watchers. Moreover and in line with Peterson, Lin, Zaun and Kanwisher (Peterson et al., 2016), the Watchers and the Walkers behaved similarly when looking at social stimuli. Interestingly, they both relied on saliency and motion to the same extent when social, written, or actionable stimuli were present in the scenery. Additionally, both groups landed their fixations on those stimuli at about the same time. On these specific stimuli, they had similar fixation patterns, echoing the findings of Peterson et al. (2016). This shows the potency of biological social relevant stimuli in attracting attention, leading to the conclusion that top-down processes overrule bottom-up processes when social mechanisms are involved during scene processing.

Altogether, our findings show that there are robust differences in saccade and fixation patterns between the Walkers and the Watchers, when performing active vision during ecological spatial navigation, with the exception of the processing of social relevant inputs. These differences cannot be attributed to the influence of sound, or task constraints, as those factors were controlled. Rather, these differences should be attributed to the load on top-down processes, due to other subprocesses to effectively carry on a task in the wild. Thus, persistent differences with the real world should be expected when studying eye movements in laboratories. These differences with the real world should be imputed to the laboratory setting, *per se*, and appear

not to be easily amended. As a consequence, results obtained in the laboratory do not fully generalize to the real world, for ecological spatial navigation, with bottom-up processes playing a different role in both conditions. Findings obtained in laboratories should be interpreted with caution, as they cannot fully account for the top-down and bottom-up modulations that human beings use while navigating in real settings.

To already minimize such differences, future studies should keep experiments in the laboratory as close as possible to real-world experiments, using naturalistic stimuli and including sound when possible. Moreover, future studies should try to develop proper modelling of eye movements in the wild, allowing to further characterize and control for the differences across both conditions. Hopefully, the advent of virtual reality (VR) technologies might shortly allow the laboratory settings to be less restrictive, by including head motion (Jacob & Karn, 2003). VR settings will also allow the assessment of the same reference frame, i.e., the point of fixation instead of the point of gaze provided by eye-tracking glasses. Moreover, the development of tools allowing data-driven analyses of eye-movements such as *iMap4* (Lao et al., 2017) with a VR component *iMap4D* (Ticcinelli et al., 2019) could help in this feat. Indeed, similar to the evolution of eye-tracking technologies, VR technologies are likely to become more affordable and user-friendly. As such, further research is necessary to investigate whether eye movements obtained in VR settings would more closely match real-world conditions. If this is the case, the VR approach will become a method of choice to investigate the functional role of eye movements in human vision.

5.6 Conclusion

The present study investigated whether visual sampling strategies generalize across laboratory (the Watchers) and real-world (the Walkers) settings, during scene processing. Our data revealed differences in saccade and fixation patterns between the Watchers and the Walkers. The Watchers directed more of their attention towards salient and moving areas than the Walkers, except when written, social or actionable stimuli were in the scenery. This differences across observers were abolished when social relevant agents were in the scenery. Overall, our data shows

that results obtained in laboratories do not fully generalize to real-world settings, at least for ecological spatial navigation. This issue might be solved in the future, thanks to the virtual reality eye movement tracking technology with higher degree of freedom, than usual eye-tracking technologies used in laboratory settings. Altogether, the findings of our study suggest caution when interpreting eye movement findings in visual scene processing, obtained uniquely in laboratory settings.

5.7 Credit Authorship Contribution Statement

M. Papinutto: Methodology, Software, Formal analysis, Investigation, Writting—original draft, Visualization. **J. Lao:** Methodology, Writting—review & editing. **D. Lalanne:** Supervision. **R. Caldara:** Conceptualization, Resources, Writting—review & editing, Funding acquisition.

5.8 Acknowledgements

5.9 REFERENCES

Bahill, A. T., Adler, D., & Stark, L. (1975). Most naturally occurring human saccades have magnitudes of 15 degrees or less. *Investigative Ophthalmology & Visual Science, 14*(6), 468–469.

Blais, C., Jack, R. E., Scheepers, C., Fiset, D., & Caldara, R. (2008). Culture shapes how we look at faces. *PLoS ONE, 3*(8), e3022. https://doi.org/10.1371/journal.pone.0003022

Borji, A., & Itti, L. (2013). State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 35*(1), 185–207. https://doi.org/10.1109/TPAMI.2012.89

Borji, A., Sihite, D. N., & Itti, L. (2013). Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study. *IEEE Transactions on Image Processing, 22*(1), 55–69. https://doi.org/10.1109/TIP.2012.2210727

Buswell, G. T. (1935). *How people look at pictures: A study of the psychology and perception in art.* Univ. Chicago Press.

Bylinskii, Z., Judd, T., Oliva, A., Torralba, A., & Durand, F. (2016). What do different evaluation metrics tell us about saliency models? *ArXiv preprint*, 1–23. http://arxiv.org/abs/1604.03605

Caldara, R. (2017). Culture reveals a flexible system for face processing. *Current Directions in Psychological Science, 26*(3), 249–255. https://doi.org/10.1177/0963721417710036

Caldara, R., Zhou, X., & Miellet, S. (2010). Putting culture under the "spotlight" reveals universal information use for face recognition. *PLoS ONE, 5*(3), e9708. https://doi.org/10.1371/journal.pone.0009708.

Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The eyelink toolbox: Eye tracking with matlab and the psychophysics toolbox. *Behavior Research Methods, Instruments, & Computers, 34*(4), 613–617. https://doi.org/10.3758/BF03195489

Coutrot, A., & Guyader, N. (2014). How saliency, faces, and sound influence gaze in dynamic social scenes. *Journal of Vision*, *14*(8), 5–5. https://doi.org/10.1167/14.8.5

Coutrot, A., Guyader, N., Ionescu, G., & Caplier, A. (2012). Influence of soundtrack on eye movements during video exploration. *Journal of Eye Movement Research*, *5*(4), 1–10. https://doi.org/10.16910/jemr.5.4.2

Dorr, M., Martinetz, T., Gegenfurtner, K. R., & Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision*, *10*(10), 28–28. https://doi.org/10.1167/10.10.28

Evans, K. M., Jacobs, R. A., Tarduno, J. A., & Pelz, J. B. (2012). Collecting and analyzing eye-tracking data in outdoor environments. *Journal of Eye Movement Research*, *5*(2), 1–19. https://doi.org/10.16910/JEMR.5.2.6

Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision*, *8*(2), 6. https://doi.org/10.1167/8.2.6

Foulsham, T., Walker, E., & Kingstone, A. (2011). The where, what and when of gaze allocation in the lab and the natural environment. *Vision Research*, *51*(17), 1920–1931. https://doi.org/10.1016/j.visres.2011.07.002

Garcia-Diaz, A. A., Fdez-Vidal, X. R., Pardo, X. M., & Dosil, R. (2012). Saliency from hierarchical adaptation through decorrelation and variance normalization. *Image and Vision Computing*, *30*(1), 51–64. https://doi.org/10.1016/j.imavis.2011.11.007

Geangu, E., Ichikawa, H., Lao, J., Kanazawa, S., Yamaguchi, M. K., Caldara, R., & Turati, C. (2016). Culture shapes 7-month-olds' perceptual strategies in discriminating facial expressions of emotion. *Current Biology*, *26*(14), 663–664. https://doi.org/10.1016/j.cub.2016.05.072

Gosselin, F., & Schyns, P. G. (2001). Bubbles: A technique to reveal the use of information in recognition tasks. *Vision Research*, *41*(17), 2261–2271. https://doi.org/10.1016/S0042-6989(01)00097-9

Hayhoe, M., McKinney, T., Chajka, K., & Pelz, J. B. (2012). Predictive eye movements in natural vision. *Experimental Brain Research*, *217*(1), 125–136. https://doi.org/10.1007/s00221-011-2979-2

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes, 537–562. https://doi.org/10.1016/B978-008044980-7/50027-6

Hillstrom, A. P., & Yantis, S. (1994). Visual motion and attentional capture. *Perception & Psychophysics*, *55*(4), 399–411. https://doi.org/10.3758/BF03205298

Hollands, M. A., & Marple-Horvat, D. E. (2001). Coordination of eye and leg movements during visually guided stepping. *Journal of Motor Behavior*, *33*(2), 205–216. https://doi.org/10.1080/00222890109603151

Hollands, M. A., Marple-Horvat, D. E., Henkes, S., & Rowan, A. K. (1995). Human eye movements during visually guided stepping. *Journal of Motor Behavior*, *27*(2). https://doi.org/10.1080/00222895.1995.9941707

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision research*, *40*(10-12), 1489–1506.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2*(3), 194–203. https://doi.org/10.1038/35058500

Jack, R. E., Blais, C., Scheepers, C., Schyns, P. G., & Caldara, R. (2009). Cultural confusions show that facial expressions are not universal. *Current Biology*, *19*. https://doi.org/10.1016/j.cub.2009.07.051

Jacob, R. J. K., & Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *The mind's eye* (pp. 573–605). Elsevier.

Judd, C. H. (1905a). Movement and consciousness. *The Psychological Monograph*, *7*, 199–226.

Judd, T., Durand, F., & Torralba, A. (2012). A benchmark of computational models of saliency to predict human fixations.

Kanan, C., Bseiso, D. N. F., Ray, N. A., Hsiao, J. H., & Cottrell, G. W. (2015). Humans have idiosyncratic and task-specific scanpaths for judging faces. *Vision Research*, *108*, 67–76. https://doi.org/10.1016/j.visres.2015.01.013.

Kelly, D. J., Liu, S., Rodger, H., Miellet, S., Ge, L., & Caldara, R. (2011). Developing cultural differences in face processing. *Developmental Science*, *14*(5), 1176–1184. https://doi.org/10.1111/j.1467-7687.2011.01067.x.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in psychtoolbox-3? *Perception*, *36*. https://doi.org/10.1068/v070821.

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

Lao, J., Miellet, S., Pernet, C., Sokhn, N., & Caldara, R. (2017). Imap4: An open source toolbox for the statistical fixation mapping of eye movement data with linear mixed modeling. *Behavior Research Methods*, *49*(12), 559–575. https://doi.org/10.3758/s13428-016-0737-x

Lappi, O. (2015). Eye tracking in the wild : The good , the bad and the ugly. *Journal of Eye Movement Research*, *8*(5), 1–21. https://doi.org/10.16910/JEMR.8.5.1

Lappi, O. (2016). Eye movements in the wild: Oculomotor control, gaze behavior & frames of reference. *Neuroscience & Biobehavioral Reviews*, *69*, 49–68. https://doi.org/10.1016/j.neubiorev.2016.06.006

Leboran, V., Garcia-DIaz, A., Fdez-Vidal, X. R., & Pardo, X. M. (2017). Dynamic whitening saliency. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(5), 893–907. https://doi.org/10.1109/TPAMI.2016.2567391

Malcom, G. L., & Henderson, J. M. (2009). The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements. *Journal of Vision*, *9*(11), 8–8. https://doi.org/10.1167/9.11.8

Marigold, D. S., & Patla, A. E. (2007). Gaze fixation patterns for negotiating complex ground terrain. *Neuroscience*, *144*(1), 302–313. https://doi.org/10.1016/j.neuroscience.2006.09.006

Matthis, J. S., Yates, J. L., & Hayhoe, M. M. (2018). Gaze and the control of foot placement when walking in natural terrain. *Current Biology, 28*(8), 1224–1233. https://doi.org/10.1016/j.cub.2018.03.008

Miellet, S., Caldara, R., & Schyns, P. G. (2011). Local jekyll and global hyde: The dual identity of face identification. *Psychological Science, 22*(12), 1518–1526. https://doi.org/10.1177/0956797611424290.

Miellet, S., He, L., Zhou, X., Lao, J., & Caldara, R. (2012). When east meets west: Gaze-contingent blindspots abolish cultural diversity in eye movements for faces. *Journal of Eye Movement Research, 5*(2). https://doi.org/10.16910/JEMR.5.2.5

Miellet, S., Vizioli, L., He, L., Zhou, X., & Caldara, R. (2013). Mapping face recognition information use across cultures. *Frontiers in Psychology, 4*(February), 34. https://doi.org/10.3389/fpsyg.2013.00034

Nuthmann, A., Einhäuser, W., & Schütz, I. (2017). How well can saliency models predict fixation selection in scenes beyond central bias? a new approach to model evaluation using generalized linear mixed models. *Frontiers in Human Neuroscience, 11*, 491. https://doi.org/10.3389/fnhum.2017.00491

Papinutto, M., Lao, J., Ramon, M., Caldara, R., & Miellet, S. (2017). The facespan-the perceptual span for face recognition. *Journal of Vision, 17*(5). https://doi.org/10.1167/17.5.16

Patla, A. E., & Vickers, J. N. (2003). How far ahead do we look when required to step on specific locations in the travel path during locomotion? *Experimental Brain Research, 148*(1), 133–138. https://doi.org/10.1007/s00221-002-1246-y

Pelli, D. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437–442. https://doi.org/10.1163/156856897X00366.

Pelz, J. B., Canosa, R. L., Kucharczyk, D., Babcock, J. S., Silver, A., & Konno, D. (2000). Portable eyetracking: A study of natural eye movements. *Proceedings of the SPIE, 3959*, 566–583.

Pelz, J. B., Hayhoe, M., & Loeber, R. (2001). The coordination of eye, head, and hand movements in a natural task. *Experimental Brain Research, 139*(3), 266–277. https://doi.org/10.1007/s002210100745

Peters, R. J., Iyer, A., Itti, L., & Koch, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision Research*, *45*(18), 2397–2416. https://doi.org/10.1016/j.visres.2005.03.019

Peterson, M. F., Lin, J., Zaun, I., & Kanwisher, N. (2016). Individual differences in face-looking behavior generalize from the lab to the world. *Journal of Vision*, *16*(7), 12. https://doi.org/10.1167/16.7.12

Rao, R. P. N., Zelinsky, G. J., Hayhoe, M. M., & Ballard, D. H. (2002). Eye movements in iconic visual search. *Vision Research*, *42*(11), 1447–1463. https://doi.org/10.1016/S0042-6989(02)00040-8

Riche, N., & Mancas, M. (2016a). *Bottom-up saliency models for still images: A practical review*. Springer. https://doi.org/10.1007/978-1-4939-3435-5_9

Riche, N., & Mancas, M. (2016b). Bottom-up saliency models for videos: A practical review. *From human attention to computational attention* (pp. 177–190). Springer.

Rodger, H., Kelly, D. J., Blais, C., & Caldara, R. (2010). Inverting faces does not abolish cultural diversity in eye movements. *Perception*, *39*(11), 1491–1503. https://doi.org/10.1068/p6750

Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show me the features! understanding recognition from the use of visual information. *Psychological Science*, *13*(5), 402–409. https://doi.org/10.1111/1467-9280.00472.

Stacchi, L., Ramon, M., Lao, J., & Caldara, R. (2019). Neural representations of faces are tuned to eye movements. *Journal of Neuroscience*, *39*(21), 4113–4123.

Stahl, J. S. (1999). Amplitude of human head movements associated with horizontal saccades. *Experimental Brain Research*, *126*(1), 41–54. https://doi.org/10.1007/s002210050715

Stratton, G. M. (1902). Eye-movement and the aesthetics of visual form. *Philosophische Studien*, *20*, 336–359.

't Hart, B. M., Vockeroth, J., Schumann, F., Bartl, K., Schneider, E., König, P., & Einhäuser, W. (2009). Gaze allocation in natural stimuli: Comparing free exploration to head-fixed

viewing conditions. *Visual Cognition*, *17*(6–7), 1132–1158. https://doi.org/10.1080/13506280902812304

Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, *7*(14), 4. https://doi.org/10.1167/7.14.4

Ticcinelli, V., De Lissa, P., Lalanne, D., Miellet, S., & Caldara, R. (2019). Imap4d: An open source toolbox for statistical fixation mapping of eye-tracking data in virtual reality. *Journal of Vision*, *19*(10), 127c. https://doi.org/10.1167/19.10.127c

Vig, E., Dorr, M., & Barth, E. (2009). Efficient visual coding and the predictability of eye movements on natural movies. *Spatial Vision*, *22*(5).

Voorhies, R. C., Elazary, L., & Itti, L. (2012). Neuromorphic bayesian surprise for far-range event detection. *2012 IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance*, 1–6.

Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human perception and performance*, *10*(5), 601.

Yarbus, A. L. (1967). *Eye movements and vision*. Plenum Press.

Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological review*, *115*(4), 787.

5.10   APPENDIX A

Figure 5.10.1

Saliency, motion and eye movements heat maps for all videos and watchers
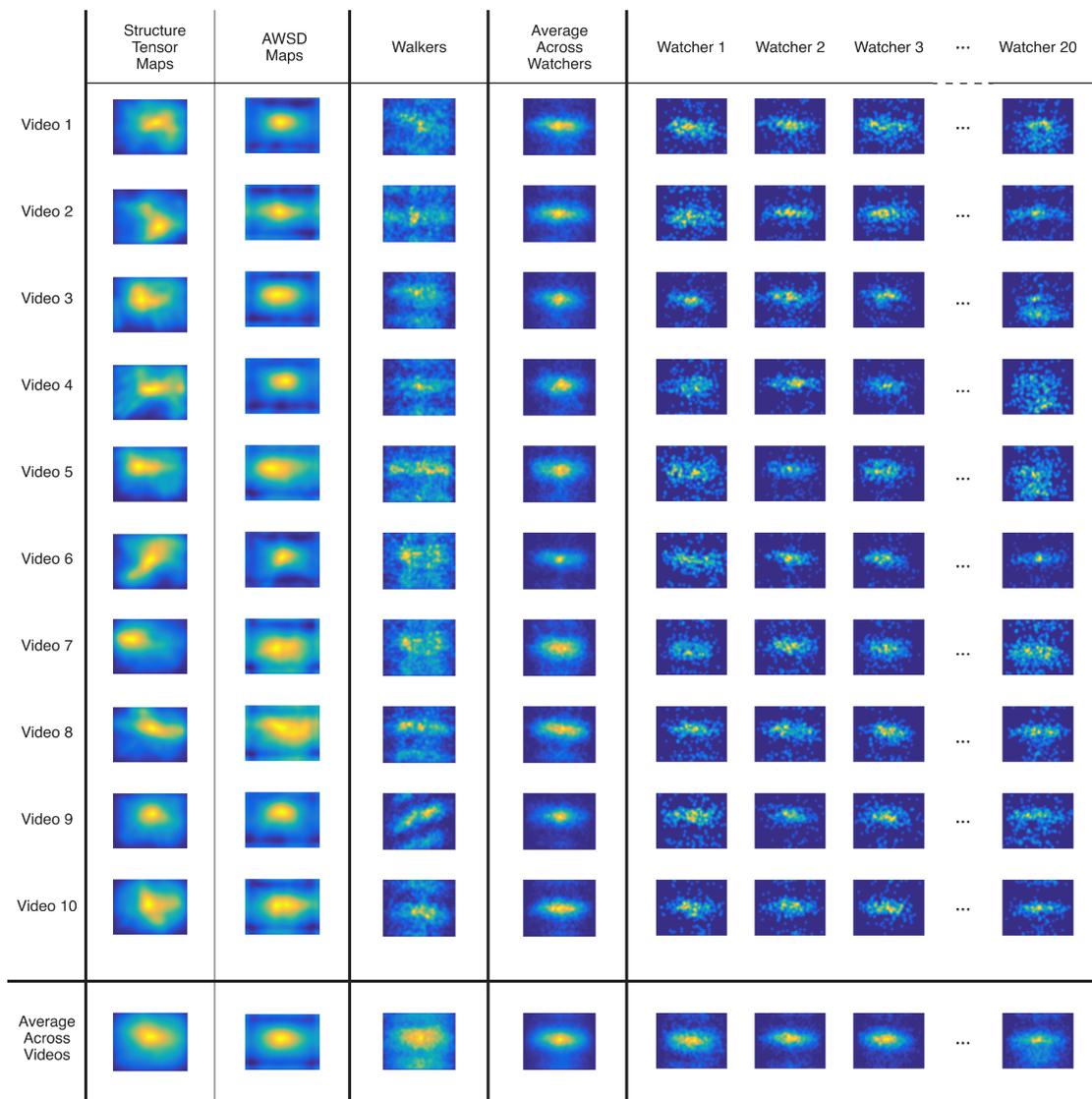
Note. Saliency maps, motion maps and eye movements heat maps for all videos.

5.11 APPENDIX B

Figure 5.11.1

Examples of frames in which both groups eye movement locations diverged

Saliency maps

Walker/Video 2 Frame 5973	Walker/Video 10 Frame 4487	Walker/Video 4 Frame 2809

Motion maps

Walker/Video 1 Frame 3454	Walker/Video 4 Frame 2808	Walker/Video 8 Frame 3995
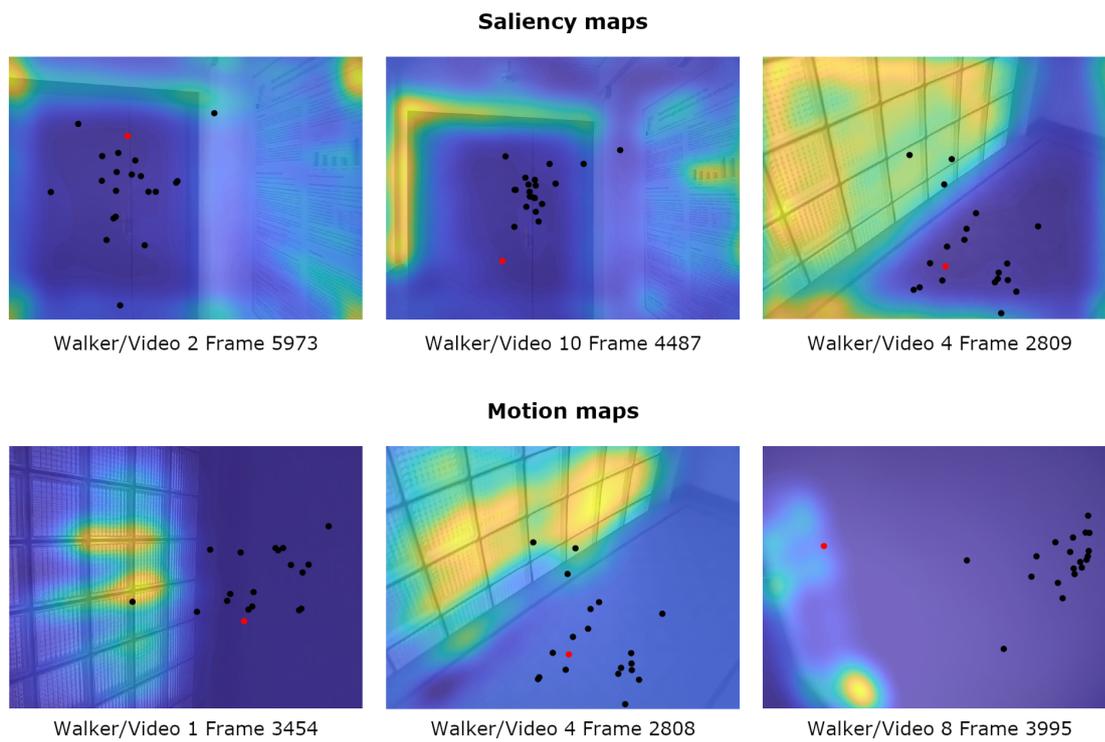
Note. Samples of frames with the lowest NSS Score computed on saliency maps and motion maps for both the Walkers (red dots) and the Watchers (black dots).

Chi va piano va sano e va lontano. Chi va forte va incontro alla morte.

Italian proverb

6

General Discussion

6.1 Key Findings

This thesis aimed to a) determine the minimal amount of information used for face recognition, b) uncover the modulation of information intake by hearing loss, and c) evaluate the ecological validity of eye movements studies performed in the laboratory. To this end, we carried out three experiments to fulfil the three goals and answer the following three questions:

1. What is the minimal quantity of information required at each fixation to achieve normal face recognition? (Contribution 1)

2. Does non-hearing elicit a different intake of both qualitative and quantitative facial information? (Contribution 2)

3. Does information sampled in the laboratory match the information sampled in the wild? (Contribution 3)

This section briefly discusses the key findings related to those questions.

6.1.1 Contribution 1

Regarding the first contribution of this thesis, we used the gaze-contingent moving-window paradigm during an old-new face recognition task. Eleven window sizes (from 9° to 19°, increasing in 1° increments) were used to compare the performance and eye movement behaviours between natural vision and each window size. The main idea is that the smaller aperture, allowing for normal performance and oculomotor behaviour, encompasses the minimal quantity of information used by the brain to recognise faces. Our results revealed that an aperture of 17° allowed participants to achieve both normal eye movement behaviour and performance. Within this aperture, we evaluated the information preserved from our gaze-contingent manipulation

and used for face recognition to achieve a visual angle of approximately 7°. In other words, overall, the facespan spans a 7° visual angle or roughly 45% of the face, allowing the eyes to sample the whole face with few fixations (Papinutto et al., 2017).

6.1.2   Contribution 2

As part of the second contribution, we evaluated the difference in information sampled and used across hearing non-signer and non-hearing signer participants during delayed-matching tasks with upright and inverted faces. We also employed a gaze-contingent moving-window paradigm, but this time, the aperture enlarged with the duration of the fixation (*i.e.*, the longer the fixation, the larger the aperture). Our results revealed that while both groups fixated similar facial features in the upright and natural viewing conditions, non-hearing participants were less affected by the FIE. Interestingly, non-hearing participants maintained the same sampling strategies on inverted face stimuli contrary to hearing participants. Additionally, the expanding spotlight paradigm revealed that non-hearing participants needed more peripheral facial information, likely indicating a larger facespan than their hearing counterparts. Altogether, these findings indicate that hearing loss impacts face identification by requiring a broader sampling of facial information but provides a more robust sample of information for stimuli manipulation.

6.1.3   Contribution 3

The two above-mentioned studies, together with recommendations for using naturalistic stimuli, raised the broader question of the ecological validity of the eye movements collected in laboratory studies. Therefore, we recorded the eye movements of participants in the wild (the Walkers) during a navigation task and those of participants in the laboratory (the Watchers) viewing the videos recorded by the Walkers. Our results revealed differences in most eye movement characteristics. Interestingly, the Walkers tended to look at their steps, whereas despite watching this movement towards the bottom, the Watchers landed their fixations and directed

their saccades more horizontally. Nevertheless, similarities in eye movement behaviours were found when social, actionable or readable stimuli were present in the scenery. Overall, our results highlight persistent differences between the laboratory and real-world settings due to the inherent differences. As such, eye movements gathered in a confined laboratory should be interpreted with caution (Papinutto et al., 2020).

6.2   Discussion

In this thesis, two studies focused on facial information use, and one explored the ecological validity of eye-tracking studies in laboratory settings. The first two studies are discussed together in Section 6.2.1, while the last one is discussed separately in Section 6.2.2. Section 6.3 presents the limitation of the aforementioned studies and future studies to expand on the results presented in this thesis are proposed in Section 6.4.

6.2.1   The Minimal Quantity and Modulation of Information Used for Face Processing

For the first contribution of this thesis, we used a gaze-contingent paradigm with fixed window size, and for the second contribution, a window size that increased with the duration of the fixations. Notably, the border of the windows was gradually merged with the average face mask. However, other studies using the gaze-contingent moving window also utilised a hard border window with background colour outside of the window. Crucially, our version of the paradigm ensures that oculomotor behaviour is minimally impacted. However, it does not provide a straightforward evaluation of the quantity of information preserved. Indeed, the quantity assessed in the windows possibly depends on the method used. While this potential difference in the evaluation of information preserved is not to be neglected, ensuring the preservation of oculomotor behaviour is even more critical, as behavioural performance increases once typical fixation patterns are deployed (Henderson et al., 2005).

The quantity of information used at each fixation—the perceptual span—depends on the quantity and the type of information to be processed. Indeed, the perceptual span has been evaluated for reading (McConkie & Rayner, 1975), scene search (Nuthmann, 2013) and faces (Papinutto et al., 2017). Interestingly, the resulting size of the perceptual span is larger for continuous information in face and scene stimuli than for discrete information found in reading. Importantly, the perceptual span for a scene search (8°) is slightly larger than for faces (7°). Such differences are likely due to the amount of information to process, which is greater in scene searches, and smaller and discrete in reading.

It is also important to note that the perceptual spans are probably not perfectly round but rather elliptical. This was first found in reading with three to four letters to the left and 14–15 to the right of a fixation (McConkie & Rayner, 1975). While the perceptual span might be more difficult to assess with continuous stimuli, it is likely also asymmetric for faces. Indeed, the elongated part follows the reading direction (from left to right in occidental countries). As for faces, the sequence of fixations usually goes from the left eye to the right eye and then the mouth; the facespan likely also follows such a shape. Moreover, as Miellet et al. (2009) assessed in reading using the parafoveal magnification paradigm, the perceptual span would probably not be affected by visual acuity.

Another modulation of perceptual spans is the ability of participants to conduct the specified task. This modulation could result in either a larger or smaller span than the one in normal participants. The greater the perceptual span, the better the task will usually be conducted as a larger amount of information can be processed simultaneously. This result was confirmed in reading where the perceptual span has been extensively studied. Indeed, studies in reading have found a larger span for fast readers and a smaller span for children (Rayner et al., 2010). Regarding the perceptual span for faces, I assume that SRs would elicit a larger facespan than normal participants as they were found to fixate a central part of the face to perform face recognition (see, *e.g.*, Ramon et al., 2019). Regarding other impairments such as prosopagnosia, I would presume a smaller facespan, as individuals with such impairments elicit stochastic sampling with most fixations focused on a specific feature of the face such as the mouth (Lee et al., 2019; Schwarzer et al., 2007; Stephan & Caine, 2009). More globally, I assume that the sampling strategy drives the facespan. With a sampling strategy centred on the nose, such as the

one observed in EAs, the facespan might be wider to encompass information from both the eyes and mouth.

Furthermore, in this thesis, we evaluated the information intake in the particular case of non-hearing signers. Interestingly, non-hearing signers exhibited a higher information intake than their hearing counterparts. Moreover, face inversion had a relatively small impact on the information used by non-hearing participants, and solely hearing participants changed their oculomotor strategy by exhibiting a greater use of mouth information. Crucially, this difference in information sampling did not result in different degrees of accuracy. Finally, in reading, hearing loss led to a larger perceptual span for reading (Bélanger et al., 2018; Bélanger et al., 2012). Accordingly, in our study, we demonstrated that non-hearing participants exhibited a larger information intake from faces. Their facespan is thus likely larger than that of hearing participants.

Finally, despite all of the modulation that the perceptual span might be confronted with, on average, it allows for the intake of all information from a face with few fixations. Importantly, with such a facespan, a single fixation at the centre does not allow the eyes to sample all features. Participants are also likely to adapt their information intake—their facespan—according to the sampling strategy deployed and the fixation location (*i.e.*, the feature fixated). This finding and the convergence towards a flexible facial information sampling system weaken the belief in a unique representation for face processing, such as the holistic processing thought to intervene in many paradigms. Rather, in my opinion, faces are likely to be processed in a coarse to fine way, with rough processing intervening at the beginning to perceive the gist of the face. This early process would allow for the programming of subsequent eye movement sampling according to an idiosyncratic preferred strategy and according to the facial representation template stored in memory. A disruption of the oculomotor strategy because of either stimuli alteration or disorders would lead to a more difficult or impaired matching, with the stored representation leading to lower face recognition performance. On top of that, the size of information intake also seems to play a crucial role in face recognition performance, with a higher intake sometimes exhibiting higher performances or more robust processing.

6.2.2   Ecological Validity of Information Sampled in Laboratory Studies

For the third contribution of this thesis, we evaluated the ecological validity of eye tracking in laboratory settings. Importantly, this study aimed to account for all differences emerging from the methodologies and paradigm used. To this end, the same task, including auditory stimulation, was used for both tasks in the wild and in the laboratory setting, thereby minimising the top-down influence in both conditions. Nevertheless, a non-negligible number of differences in eye movement behaviour persisted. Most of these differences arose from the setting itself and related to the sub-task involved. In our study, walking elicited most of the differences, such as participants watching their next step. In my opinion, such differences also emerged from tasks other than walking, such as those involving hand–gaze coordination.

Importantly, however, our study revealed a synchronisation in eye movements when actionable, readable or social stimuli were present in the scenery. In line with this finding, it was recently shown that faces in the wild and in the laboratory elicited the same pattern of fixations indicating that the same visual information was sampled in both conditions. Crucially, such findings suggest that results found in laboratory settings might convey reliable information, at least for specific processes. However, those results were found according to averaged eye movement characteristics or fixation maps. To go further, scan paths could be compared (*e.g.,* using ScanMatch: Cristino et al., 2010; or MultiMatch: Jarodzka et al., 2010) to ensure that the similarity between the real world and the laboratory continues throughout the temporal sequence of eye movements.

In conclusion, our findings regarding the facespan and the modulation of information used in the special case of non-hearing signers might also apply to face processing in the wild. Moreover, the recent hypothesis of an idiosyncratic strategy for face processing might be based in the real world. However, this should be confirmed with further studies.

6.3 LIMITATIONS

6.3.1 LIMITATIONS OF THE FIRST EXPERIMENTAL CONTRIBUTION

The facespan was evaluated in a between-subjects design. However, such a methodology does not account for the variability among participants: one participant could have had a different facespan than another. We did not choose to use this method, as our study was exploratory, and we would have been required to evaluate a high number of different window sizes, which would have meant a long and exhausting study for participants. However, to complement this study, another is currently in progress to evaluate the facespan individually. An additional limitation arising from our methodology is that participants learnt faces with the gaze-contingent moving-window, meaning that the facespan is both the facespan to encode and to recognise faces. There are two rationales behind this choice. First, in classical psychological experiments, the context congruence between learning and recall has been found to be crucial (see, *e.g.*, Godden & Baddeley, 1975). Second, previous studies have postulated the hypothesis of a scan path replay, under which the scan path employed during learning would be replayed identically to proceed to recognition (Noton & Stark, 1971). Accordingly, few differences in fixation patterns have been found between learning and recognition (see, *e.g.*, Blais et al., 2008; Caldara et al., 2010; Hsiao & Cottrell, 2008). However, some differences in fixation patterns remain between the two conditions, with learning exhibiting a more analytical scan path and recognition a more holistic one (Arizpe et al., 2019). Thus, evaluating these two aspects separately could be interesting to ensure similarity between the facespan to learn and the facespan to recall faces. Finally, the perceptual span is only one of three concepts characterising the limitations of the human visual system. The other two are the visual attention span and the visual span. The former is defined as the maximal number of distinct stimuli that can be processed simultaneously in a fixation while observing a multi-element configuration. Regarding the vision span, it represents the distance on either side of a fixation within which a stimulus of a given size can be recognised (Frey & Bosse, 2018).

6.3.2   Limitations of the Second Experimental Contribution

One of the limitations of the second contribution is that we evaluated only hearing non-signers. However, our results could have arisen from the non-hearing participants' signing ability, which might have increased the perceptual span of non-hearing signers. Indeed, non-hearing signers must consider both facial expression and hand motions, which require a greater ability to perceive extra-foveal information. Another limitation is that the gaze-contingent paradigm used—the expanding spotlight—does not accurately reflect the size of information intake. Indeed, the main assumption of the expanding spotlight paradigm is that fixation on a location will be maintened until observers extract a sufficient amount of information from this location to solve the task at hand (Miellet et al., 2013). However, although less likely, it could be that the duration of the fixation and thus the processing time matter more than the amount of information taken in. Indeed, in face processing, the coarse-to-fine processing (Van Rijsbergen & Schyns, 2009; Winston et al., 2003) and the duration of the fixation increasing over time (see, *e.g.*, Hsiao & Cottrell, 2008) support the idea of fast processing at the beginning and slower and more complex processing at the end. In this case, the extent of the window would not reflect the amount of information required at the fixation location but rather the processing time required.

6.3.3   Limitations of the Third Experimental Contribution

In evaluating the ecological validity of eye movements in laboratory studies during navigation, we conveyed auditory input in a monophonic fashion using speakers. However, monophonic auditory input does not convey any information about the spatial distribution of the sound. The differences between the real-world and the laboratory settings found in our study could consequently have been amplified by the lack of spatial localisation of the sound. Spatial localisation of the sound has indeed been found to attract eye movements (Coutrot & Guyader, 2014; Coutrot et al., 2012). The other limitation of this last contribution is that both conditions (the real-world and the laboratory setting) were a between-subjects factor. As such, differences could arise between observers, whereas a within-subjects design could have accounted for group differences. However, as previously mentioned, a within-group design raises other

challenges, such as habituation to the scenery and the recognition of participants' own walk-ing pace in the laboratory setting. For those reasons, I think that a between-subjects paradigm might be the best to evaluate the ecological validity of eye movement studies, although many participants would be required for it to be sound.

6.4 FUTURE PERSPECTIVES

Is the facespan idiosyncratic, and does it decrease with information intake?

As described in Section 1.3 and discussed in the previous section, face-related research has deduced that eye movements are probably idiosyncratic and flexible, with the first fixation determining subsequent fixations. It is likely that for scene processing, faces are processed in a coarse-to-fine fashion(Van Rijsbergen & Schyns, 2009; Winston et al., 2003). More specifi-cally, the first fixation location might gather coarse information about a the face stimulus, and then further fixations gather more fine-grained facial information. As such, the perceptual span for the first fixation is likely to be larger than the facespan of the subsequent fixations and might decrease throughout the scan path. Evaluating such a strong claim would require a new gaze-contingent paradigm to reduce available information throughout the trial. Similarly to the ex-panding spotlight created by Miellet et al. (2013), in this gaze-contingent paradigm, which could be named the constricting spotlight paradigm, the spotlight decreases in size linearly with time while its size would remain stable during fixations. Two parameters can be manipulated in such a gaze-contingent paradigm: the window's initial size and the speed of constriction. These parameters could either be discretised into several conditions or be evaluated through the use of staircase paradigms, such as the Quest paradigm (Watson & Pelli, 1983) to assess which pa-rameters best fit participants' idiosyncratic face processing. Importantly, two baselines should be considered: first, natural viewing, which is regularly used in gaze-contingent paradigms, and second a window with a fixed size throughout the experiment, similar to the one used in the facespan experiment. Ideal parameters, as well as individual facespans, could subsequently be related to eye movement sampling strategies—for example, using HMMEM (Jarodzka et al.,

2010) or scan path methods (Cristino et al., 2010)—thereby uncovering the relation between sampling strategies and information use throughout the facial recognition process. According to the claim above, I would expect a decrease in information intake, (*i.e.,* a smaller window required to achieve a normal behaviour) in each trial. According to the discussion on the modulation of the facespan in Section 6.2.1, I would expect a fixation strategy eliciting more central fixations to exhibit a larger information intake and a smaller decrease along the scan path, whereas fixation patterns focusing on features would probably elicit a smaller information intake with a marked decrease in information intake along the scan path. Crucially, such an experiment would profoundly deepen our understanding of the time course of face processing in terms of the size of information intake. Additionally, this could provide further clues about the currently debated representations of faces.

Are there any brain indices for reaching the minimal quantity of information used at each fixation?

As part of the first contribution, we shed light on the minimal amount of information required at each fixation to produce normal oculomotor behaviour and behavioural performance. While the use of a retinal filter uncovers the information that is fed to the brain, studying the brain's reaction to such a restriction of information intake could fill the gap between information use and brain processing. Moreover, one of my colleagues recently revealed that neural responses to faces are tuned to fixation locations. Accordingly, brain response is likely modulated by both the qualitative and quantitative nature of the information available at the fixation location. In this regard, an electroencephalogram (EEG) appears to be a well-defined technique to assess brain response to the modulation of the quantity of information available at the fixation location. Regular event-related potentials (ERPs) could be utilised alongside gaze-contingent paradigms, specifically, showing faces for about 300 ms at several locations in natural viewing and with a spotlight of different sizes. While being controlled, such a paradigm would require the presentation of a tremendous number of stimuli. Moreover, as EEG response seems to be finely tuned to idiosyncratic eye movements, arbitrarily selecting a unique fixation location might hinder participant specificity and the time course of eye movements. The eye fixation-related potentials (EFRPs) constitutes another, but more ecological, paradigm that ad-

dresses brain response to gaze-contingent moving-window paradigm. Indeed, this paradigm would allow researchers to analyse brain response during free viewing in both natural vision and with several gaze-contingent moving-window sizes. However, although it is used in reading (for recent reviews, see Degno & Liversedge, 2020; Degno et al., 2021), in scene search (see, *e.g.*, Brouwer et al., 2013; Devillez et al., 2015; Kaunitz et al., 2014), and in few experiments strictly related to face processing (excluding face search studies, De Lissa et al., 2019; Mares et al., 2016), performing an EEG during free viewing is challenging. Indeed, EEG signal during fixation is affected by the preceding saccades' lengths and directions and by the duration of the fixation (Nikolaev et al., 2016). Such impacts can be controlled by employing linear regression or a generalised additive mixed model to account for the non-linearity of overlapping EEG signals. However, the dependency of eye movements on preceding ones (see, Tatler & Vincent, 2008; Unema et al., 2005) render the use of the linear method to correct for overlapping EEG signals during free viewing flawed (Miller & Chapman, 2001; Wildt et al., 1978). A more efficient and simpler method to counteract the impact of preceding eye movements is to match the fixation duration and preceding saccades' lengths and directions across conditions using a cut-off on a distance metric, such as the Mahalanobis distance. Finally, two additional impacts should be controlled for. The first is the sequence of fixations, as the EFRP amplitude might decrease for each trial, with the first fixation eliciting a greater amplitude than subsequent fixations (see, *e.g.*, Rämä & Baccino, 2010, for an example of object identification). Moreover, the luminance should be accounted for as the fixation location might also impact EFRPs. While also requiring many trials, such a paradigm would enable more ecological face processing than standard ERP experiments. Importantly, to my knowledge, such an experiment would be the first to combine EFRPs with a gaze-contingent paradigm to investigate face recognition during free-viewing. The lack of research of this kind is likely due to the technical difficulty of collecting data in co-registration of EEG and eye-tracking signals. Additionally, evaluating EFRPs requires a consequent control for possible flaws due to preceding and current eye movements, which can explain why there are so few EFRP experiments. Although difficult to set up, conduct, and analyse properly, such an experiment would allow researchers to fill crucial gaps between information use and brain processing. Additionally, it has not only fundamental research implications but also methodological implications. Indeed, the impact of the gaze-contingent moving window could be further investigated by evaluating presaccadic activity between natural vision and the gaze-contingent moving-window.

Does signing increase the quantity of information used during face processing?

As noted in the discussion regarding the third contribution, when studying non-hearing participants, the extent to which their signing ability contributes to the improvement of non-hearing parafoveal processing remains to be clarified (for a review, see Bavelier et al., 2006). More specifically, visual processing in non-hearing signers is a chicken-and-egg problem; that is, it remains to be determined whether their signing ability increases their parafoveal processing, which in turn adapts and makes their sampling strategy more robust. Conversely, it could be that hearing loss increases parafoveal processing, resulting in a more robust strategy without being impacted by one's signing ability. Interestingly, few studies shed some light on this question but do not address the information intake. Indeed, when looking at a signing person, children who are native or experienced signers tend to look higher on the face than non-signing or beginning signers (Bosworth & Stone, 2021; Emmorey et al., 2009). Accordingly, signers likely process a face differently than non-signers. Experimentally assessing such an issue would require the inclusion of a group of hearing signers and a comparison of their results with both non-hearing signers and hearing non-signers. Although finding non-hearing non-signers might be unlikely to obtain a full picture, including hearing signers would nevertheless allow researchers to unpack the influence of signing on the sampling strategy and the information intake during face identification. Altogether, such an experiment would shed light on the contribution of multi-sensory integration for face processing.

Do new technologies reduce or discard differences between laboratory and real-world studies?

For the last contribution of this thesis, we investigated the ecological validity of laboratory studies. Such studies seem valid for several particular forms of processing. However, more natural tasks are likely to be biased by the laboratory setting. Fortunately, in the past year, advances in technologies such as VR or AR enabled one to study eye movements in a controlled environment while preserving the third dimension and head movements. On the one hand, AR will match eye movements from the real world as it reproduces real-world scenery and can

modify it at will. However, it requires real-world stimuli, which might not be ideal for vision research. On the other hand, VR provides a fully customisable environment, resulting in more closely and more contextualised stimuli from the real world than usual screens. However, while the differences between the real-world and laboratory settings become clearer with the increasing number of studies, the differences in such new technologies remain unknown. Interestingly, VR could reduce or even discard the differences observed in the real world. Evaluating the consistency between VR and real-world experiments and possibly demonstrating the ecological validity of VR would allow us to highlight the importance and encourage the use of those new technologies to assess eye movements in ecological settings with a strong rationale.

6.5   Conclusion

In conclusion, we uncovered the minimal amount of information used for face processing and the modulation of information intake in non-hearing signers. In summary, in face recognition, at least 45% of the information used at each fixation is required for individuals to achieve normal face recognition performance and natural eye movement behaviour. Importantly, normal eye movements behaviour is required for an increase in performance. Moreover, the minimal quantity of information used at each fixation allows for the sampling of the whole face in a few fixations. Interestingly, information use could be modulated by other sensory deficits and possibly other constraints, such as neuropsychological disorders. In this thesis, this was the case for hearing loss exhibiting a higher information intake and resulting in a more robust sampling strategy towards stimuli alteration. Moreover, we also demonstrated that intrinsic environmental differences persist between laboratory and real-world experiments. Crucially, the experiments conducted in this thesis were performed in a confined laboratory. It consequently remains unclear whether information use studied in the laboratory holds in real-world settings. Nevertheless, our study on the ecological validation of eye movement studies and other recent studies indicate that information intake is hard-wired for the special case of faces and does not seem to be impacted by external constraints. Additionally, the promising advent of portable eye-tracking devices, VR and AR will soon enable researchers to address those issues empirically. Indeed, manipulating the 3D information conveyed by such devices will allow researchers to

present participants with more ecologically valid and contextualised stimulation. In conclusion, I hope this brief overview of the vast field of vision science has convinced readers that eye movements are fruitful cues to access underlying cognitive representations and brain processes with a promising future alongside technological advances, both empirically and in everyday life.

References

ARTICLES

Ahmadlou, M., Zweifel, L. S., & Heimel, J. A. (2018). Functional modulation of prim. visual cortex by the superior colliculus in the mouse. *Nature communications*, *9*(1), 1–13.

Albrecht, M. A., Foster, J. K., Joosten, A., Falkmer, M., Tang, J., Leung, D., Ordqvist, A., & Falkmer, T. (2014). Visual search strategies during facial recognition in children with asd. *Research in Autism Spectrum Disorders*, *8*(5), 559–569.

Althoff, R. R., & Cohen, N. J. (1999). Eye-movement-based memory effect: A reprocessing effect in face perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(4), 997.

Anderson, N. C., Anderson, F., Kingstone, A., & Bischof, W. F. (2015). A comparison of scan-path comparison methods. *Behavior research methods*, *47*(4), 1377–1392.

Angelaki, D. E. (2004). Eyes on target: What neurons must do for the vestibuloocular reflex during linear motion. *Journal of neurophysiology*, *92*(1), 20–35.

Antes, J. R. (1974). The time course of picture viewing. *Journal of experimental psychology*, *103*(1), 62.

Anzures, G., Quinn, P. C., Pascalis, O., Slater, A. M., & Lee, K. (2013). Development of own-race biases. *Visual cognition*, *21*(9-10), 1165–1182.

Anzures, G., Quinn, P. C., Pascalis, O., Slater, A. M., Tanaka, J. W., & Lee, K. (2013). Developmental origins of the other-race effect. *Current directions in psychological science*, *22*(3), 173–178.

Arcaro, M. J., Schade, P. F., Vincent, J. L., Ponce, C. R., & Livingstone, M. S. (2017). Seeing faces is necessary for face-domain formation. *Nature neuroscience*, *20*(10), 1404.

Arizpe, J., Kravitz, D. J., Yovel, G., & Baker, C. I. (2012). Start position strongly influences fixation patterns during face processing: Difficulties with eye movements as a measure of information use. *PloS one*, *7*(2), e31106.

Arizpe, J., Noles, D. L., Tsao, J. W., & Chan, A. W.-Y. (2019). Eye movement dynamics differ between encoding and recognition of faces. *Vision*, *3*(1), 9.

Arizpe, J., Walsh, V., Yovel, G., & Baker, C. I. (2017). The categories, frequencies, and stability of idiosyncratic eye-movement patterns to faces. *Vision research*, *141*, 191–203.

Arnold, P., & Murray, C. (1998). Memory for faces and objects by deaf and hearing signers and hearing nonsigners. *Journal of psycholinguistic research*, *27*(4), 481–497.

Bahill, A. T., Adler, D., & Stark, L. (1975). Most naturally occurring human saccades have magnitudes of 15 degrees or less. *Investigative Ophthalmology & Visual Science*, *14*(6), 468–469.

Bahill, A. T., Clark, M. R., & Stark, L. (1975). The main sequence, a tool for studying human eye movements. *Mathematical Biosciences*, *24*(3-4), 191–204.

Ballard, D. H. (1991). Animate vision. *Artificial intelligence*, *48*(1), 57–86.

Ballard, D. H., Hayhoe, M., Li, F., & Whitehead, S. D. (1992). Hand-eye coordination during sequential tasks. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *337*(1281), 331–339.

Barnes, G. R., & Smith, R. (1981). The effects on visual discrimination of image movement across the stationary retina. *Aviation, space, and environmental medicine*.

Bate, S., Haslam, C., Jansari, A., & Hodgson, T. L. (2009). Covert face recognition relies on affective valence in congenital prosopagnosia. *Cognitive Neuropsychology, 26*(4), 391–411.

Bavelier, D., Brozinsky, C., Tomann, A., Mitchell, T., Neville, H., & Liu, G. (2001). Impact of early deafness and early exposure to sign language on the cerebral organization for motion processing. *Journal of Neuroscience, 21*(22), 8931–8942.

Bavelier, D., Dye, M. W., & Hauser, P. C. (2006). Do deaf individuals see better? *Trends in cognitive sciences, 10*(11), 512–518.

Bélanger, N. N., Lee, M., & Schotter, E. R. (2018). Young skilled deaf readers have an enhanced perceptual span in reading. *Quarterly Journal of Experimental Psychology, 71*(1), 291–301.

Bélanger, N. N., Slattery, T. J., Mayberry, R. I., & Rayner, K. (2012). Skilled deaf readers have an enhanced perceptual span in reading. *Psychological science, 23*(7), 816–823.

Bender, M., & Shanzer, S. (1983). History of optokinetic nystagmus. *Neuro-ophthalmology, 3*(2), 73–88.

Bennetts, R. J., Mole, J., & Bate, S. (2017). Super-recognition in development: A case study of an adolescent with extraordinary face recognition skills. *Cognitive neuropsychology, 34*(6), 357–376.

Bestelmeyer, P. E. G., Tatler, B. W., Phillips, L. H., Fraser, G., Benson, P. J., & Clair, D. S. (2006). Global visual scanning abnormalities in schizophrenia and bipolar disorder. *Schizophrenia Research, 87*(1-3), 212–222.

Bettger, J. G., Emmorey, K., McCullough, S. H., & Bellugi, U. (1997). Enhanced facial discrimination: Effects of experience with american sign language. *Journal of Deaf Studies and Deaf Education*, 223–233.

Biederman, I. (1972). Perceiving real-world scenes. *Science, 177*(4043), 77–80.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive psychology, 14*(2), 143–177.

Billino, J., van Belle, G., Rossion, B., & Schwarzer, G. (2018). The nature of individual face recognition in preschool children: Insights from a gaze-contingent paradigm. *Cognitive Development, 47*, 168–180.

Blais, C., Jack, R. E., Scheepers, C., Fiset, D., & Caldara, R. (2008). Culture shapes how we look at faces. *PLoS ONE, 3*(8), e3022. https://doi.org/10.1371/journal.pone.0003022

Blank, H., Anwander, A., & Von Kriegstein, K. (2011). Direct structural connections between voice-and face-recognition areas. *Journal of Neuroscience, 31*(36), 12906–12915.

Bobak, A. K., Parris, B. A., Gregory, N. J., Bennetts, R. J., & Bate, S. (2017). Eye-movement strategies in developmental prosopagnosia and "super" face recognition. *Quarterly journal of experimental psychology, 70*(2), 201–217.

Bombari, D., Mast, F. W., & Lobmaier, J. S. (2009). Featural, configural, and holistic face-processing strategies evoke different scan patterns. *Perception, 38*(10), 1508–1521.

Borji, A., & Itti, L. (2013). State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 35*(1), 185–207. https://doi.org/10.1109/TPAMI.2012.89

Borji, A., Sihite, D. N., & Itti, L. (2013). Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study. *IEEE Transactions on Image Processing, 22*(1), 55–69. https://doi.org/10.1109/TIP.2012.2210727

Borji, A., Tavakoli, H. R., Sihite, D. N., & Itti, L. (2013). Analysis of scores, datasets, and models in visual saliency prediction. *Proceedings of the IEEE international conference on computer vision*, 921–928.

Bortolon, C., Capdevielle, D., & Raffard, S. (2015). Face recognition in schizophrenia disorder: A comprehensive review of behavioral, neuroimaging and neurophysiological studies. *Neuroscience & Biobehavioral Reviews, 53*, 79–107.

Bosworth, R. G., & Dobkins, K. R. (2002). Visual field asymmetries for motion processing in deaf and hearing signers. *Brain and cognition, 49*(1), 170–181.

Bosworth, R. G., & Stone, A. (2021). Rapid development of perceptual gaze control in hearing native signing infants and children. *Developmental Science*, e13086.

Brouwer, A.-M., Reuderink, B., Vincent, J., van Gerven, M. A., & van Erp, J. B. (2013). Distinguishing between target and nontarget fixations in a visual search task using fixation-related potentials. *Journal of vision*, *13*(3), 17–17.

Buckley, D., Codina, C., Bhardwaj, P., & Pascalis, O. (2010). Action video game players and deaf observers have larger goldmann visual fields. *Vision research*, *50*(5), 548–556.

Butler, S., Gilchrist, I. D., Burt, D. M., Perrett, D. I., Jones, E., & Harvey, M. (2005). Are the perceptual biases found in chimeric face processing reflected in eye-movement patterns? *Neuropsychologia*, *43*(1), 52–59.

Bylinskii, Z., Judd, T., Oliva, A., Torralba, A., & Durand, F. (2016). What do different evaluation metrics tell us about saliency models? *ArXiv preprint*, 1–23. http://arxiv.org/abs/1604.03605

Caldara, R. (2017). Culture reveals a flexible system for face processing. *Current Directions in Psychological Science*, *26*(3), 249–255. https://doi.org/10.1177/0963721417710036

Caldara, R., & Miellet, S. (2011). Imap: A novel method for statistical fixation mapping of eye movement data. *Behavior Research Methods*, *43*(3), 864–878. https://doi.org/10.3758/s13428-011-0092-x.

Caldara, R., Zhou, X., & Miellet, S. (2010). Putting culture under the "spotlight" reveals universal information use for face recognition. *PLoS ONE*, *5*(3), e9708. https://doi.org/10.1371/journal.pone.0009708.

Cattell, J. M. (1900). On relations of time and space in vision. *Psychological Review*, *7*(4), 325.

Chajka, K., Hayhoe, M., Sullivan, B., Pelz, J., Mennie, N., & Droll, J. (2006). Predictive eye movements in squash. *Journal of Vision*, *6*(6), 481–481.

Chakraborty, A., & Chakrabarti, B. (2018). Looking at my own face: Visual processing strategies in self–other face recognition. *Frontiers in psychology*, *9*, 121.

Chan, J. P. K., & Ryan, J. D. (2012). Holistic representations of internal and external face features are used to support recognition. *Frontiers in psychology*, 3, 87.

Chauvin, A., Worsley, K. J., Schyns, P. G., Arguin, M., & Gosselin, F. (2005). Accurate statistical tests for smooth classification images. *Journal of Vision*, 5(9), 659–667. https://doi.org/10.1167/5.9.1.

Chua, H. F., Boland, J. E., & Nisbett, R. E. (2005). Cultural variation in eye movements during scene perception. *Proceedings of the National Academy of Sciences*, 102(35), 12629–12633.

Chuk, T., Chan, A. B., & Hsiao, J. H. (2014). Understanding eye movements in face recognition using hidden markov models. *Journal of vision*, 14(11), 8–8.

Chuk, T., Chan, A. B., Shimojo, S., & Hsiao, J. H. (2019). Eye movement analysis with switching hidden markov models. *Behavior research methods*, 1–18.

Codina, C., Buckley, D., Port, M., & Pascalis, O. (2011). Deaf and hearing children: A comparison of peripheral vision development. *Developmental science*, 14(4), 725–737.

Cognolato, M., Atzori, M., & Müller, H. (2018). Head-mounted eye gaze tracking devices: An overview of modern devices and recent advances. *Journal of rehabilitation and assistive technologies engineering*, 5, 2055668318773991.

Corbetta, M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Ollinger, J. M., Drury, H. A., Linenweber, M. R., Petersen, S. E., Raichle, M. E., Van Essen, D. C., & Shulman, G. (1998). A common network of functional areas for attention and eye movements. *Neuron*, 21(4), 761–773.

Corbetta, M., & Shulman, G. L. (1998). Human cortical mechanisms of visual attention during orienting and search. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 353(1373), 1353–1362.

Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The eyelink toolbox: Eye tracking with matlab and the psychophysics toolbox. *Behavior Research Methods, Instruments, & Computers*, 34(4), 613–617. https://doi.org/10.3758/BF03195489

Coutrot, A., & Guyader, N. (2014). How saliency, faces, and sound influence gaze in dynamic social scenes. *Journal of Vision*, *14*(8), 5–5. https://doi.org/10.1167/14.8.5

Coutrot, A., Guyader, N., Ionescu, G., & Caplier, A. (2012). Influence of soundtrack on eye movements during video exploration. *Journal of Eye Movement Research*, *5*(4), 1–10. https://doi.org/10.16910/jemr.5.4.2

Cristino, F., Mathôt, S., Theeuwes, J., & Gilchrist, I. D. (2010). Scanmatch: A novel method for comparing fixation sequences. *Behavior research methods*, *42*(3), 692–700.

Dalton, K. M., Nacewicz, B. M., Johnstone, T., Schaefer, H. S., Gernsbacher, M. A., Goldsmith, H. H., Alexander, A. L., & Davidson, R. J. (2005). Gaze fixation and the neural circuitry of face processing in autism. *Nature neuroscience*, *8*(4), 519–526.

Davis, J., McKone, E., Zirnsak, M., Moore, T., O'Kearney, R., Apthorp, D., & Palermo, R. (2017). Social and attention-to-detail subclusters of autistic traits differentially predict looking at eyes and face identity recognition ability. *British Journal of Psychology*, *108*(1), 191–219.

De Filippo, C. L., & Lansing, C. R. (2006). Eye fixations of deaf and hearing observers in simultaneous communication perception. *Ear and hearing*, *27*(4), 331–352.

De Graef, P., Christiaens, D., & d'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological research*, *52*(4), 317–329.

De Heering, A., Rossion, B., & Pascalis, O. (2012). Early deafness increases the face inversion effect but does not modulate the composite face effect. *Frontiers in psychology*, *3*, 124.

De Heering, A., Rossion, B., Turati, C., & Simion, F. (2008). Holistic face processing can be independent of gaze behaviour: Evidence from the composite face illusion. *Journal of Neuropsychology*, *2*(1), 183–195.

De Lissa, P., McArthur, G., Hawelka, S., Palermo, R., Mahajan, Y., Degno, F., & Hutzler, F. (2019). Peripheral preview abolishes n170 face-sensitivity at fixation: Using fixation-related potentials to investigate dynamic face processing. *Visual Cognition*, *27*(9-10), 740–759.

Degno, F., & Liversedge, S. P. (2020). Eye movements and fixation-related potentials in reading: A review. *Vision*, *4*(1), 11.

Degno, F., Loberg, O., & Liversedge, S. P. (2021). Co-registration of eye movements and fixation—related potentials in natural reading: Practical issues of experimental design and data analysis. *Collabra: Psychology*, *7*(1).

Delabarre, E. B. (1898). A method of recording eye-movements. *The American Journal of Psychology*, *9*(4), 572–574.

Den Buurman, R., Roersema, T., & Gerrissen, J. F. (1981). Eye movements and the perceptual span in reading. *Reading Research Quarterly*, 227–235.

Deubel, H., OŔegan, K., & Radach, R. (2000). Attention, information processing and eye movement control. *Reading as a perceptual process*, 355–374.

Devillez, H., Guyader, N., & Guérin-Dugué, A. (2015). An eye fixation–related potentials analysis of the p300 potential for fixations onto a target object when exploring natural scenes. *Journal of vision*, *15*(13), 20–20.

Dodge, R. (1916). Visual motor functions. *Psychological Bulletin*, *13*(11), 421.

Dodge, R., & Cline, T. S. (1901). The angle velocity of eye movements. *Psychological Review*, *8*(2), 145.

Dorr, M., Martinetz, T., Gegenfurtner, K. R., & Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision*, *10*(10), 28–28. https://doi.org/10.1167/10.10.28

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of experimental psychology: General*, *113*(4), 501.

Dye, M. W., Hauser, P. C., & Bavelier, D. (2009). Is visual selective attention in deaf individuals enhanced or deficient? the case of the useful field of view. *PloS one*, *4*(5), e5640.

Dye, M. W., Seymour, J. L., & Hauser, P. C. (2016). Response bias reveals enhanced attention to inferior visual field in signers of american sign language. *Experimental brain research*, *234*(4), 1067–1076.

Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). Faces—a database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior Research Methods, 42*(1), 351–362. https://doi.org/10.3758/BRM.42.1.351.

Eckstein, M. P. (2011). Visual search: A retrospective. *Journal of vision, 11*(5), 14–14.

Ehrlichman, H., & Micic, D. (2012). Why do people move their eyes when they think? *Current Directions in Psychological Science, 21*(2), 96–100.

EinhÄ, W., Rutishauser, U., & Koch, C. (2008). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of vision, 8*(2), 2–2.

Ekstrom, L. B., Roelfsema, P. R., Arsenault, J. T., Bonmassar, G., & Vanduffel, W. (2008). Bottom-up dependent gating of frontal signals in early visual cortex. *Science, 321*(5887), 414–417.

Emmorey, K., Thompson, R., & Colvin, R. (2009). Eye gaze during comprehension of american sign language by native and beginning signers. *Journal of Deaf Studies and Deaf Education, 14*(2), 237–243.

Eriksen, C. W., & Colegate, R. L. (1971). Selective attention and serial processing in briefly presented visual displays. *Perception & Psychophysics, 10*(5), 321–326.

Eriksen, C. W., & Hoffman, J. E. (1972). Temporal and spatial characteristics of selective encoding from visual displays. *Perception & psychophysics, 12*(2), 201–204.

Eriksen, C. W., & James, J. D. S. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & psychophysics, 40*(4), 225–240.

Evans, K. M., Jacobs, R. A., Tarduno, J. A., & Pelz, J. B. (2012). Collecting and analyzing eye-tracking data in outdoor environments. *Journal of Eye Movement Research, 5*(2), 1–19. https://doi.org/10.16910/JEMR.5.2.6

Evans, K. M., Rotello, C. M., Li, X., & Rayner, K. (2009). Scene perception and memory revealed by eye movements and receiver-operating characteristic analyses: Does a cultural difference truly exist? *The Quarterly Journal of Experimental Psychology, 62*(2), 276–285.

Findlay, J. M. (1997). Saccade target selection during visual search. *Vision research*, *37*(5), 617–631.

Fine, I., Finney, E. M., Boynton, G. M., & Dobkins, K. R. (2005). Comparing the effects of auditory deprivation and sign language within the auditory and visual cortex. *Journal of cognitive neuroscience*, *17*(10), 1621–1637.

Fletcher-Watson, S., Findlay, J. M., Leekam, S. R., & Benson, V. (2008). Rapid detection of person information in a naturalistic scene. *Perception*, *37*(4), 571–583.

Foulsham, T., Gray, A., Nasiopoulos, E., & Kingstone, A. (2013). Leftward biases in picture scanning and line bisection: A gaze-contingent window study. *Vision research*, *78*, 14–25.

Foulsham, T., Teszka, R., & Kingstone, A. (2011). Saccade control in natural images is shaped by the information visible at fixation: Evidence from asymmetric gaze-contingent windows. *Attention, Perception, & Psychophysics*, *73*(1), 266–283.

Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision*, *8*(2), 6. https://doi.org/10.1167/8.2.6

Foulsham, T., & Underwood, G. (2011). If visual saliency predicts search, then why? evidence from normal and gaze-contingent search tasks in natural scenes. *Cognitive Computation*, *3*(1), 48–63.

Foulsham, T., Walker, E., & Kingstone, A. (2011). The where, what and when of gaze allocation in the lab and the natural environment. *Vision Research*, *51*(17), 1920–1931. https://doi.org/10.1016/j.visres.2011.07.002

Fox, E. (2002). Processing emotional facial expressions: The role of anxiety and awareness. *Cognitive, Affective, & Behavioral Neuroscience*, *2*(1), 52–63.

Frey, A., & Bosse, M.-L. (2018). Perceptual span, visual span, and visual attention span: Three potential ways to quantify limits on visual processing during reading. *Visual Cognition*, *26*(6), 412–429.

Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of experimental psychology: General*, *108*(3), 316.

Garcia-Diaz, A. A., Fdez-Vidal, X. R., Pardo, X. M., & Dosil, R. (2012). Saliency from hierarchical adaptation through decorrelation and variance normalization. *Image and Vision Computing*, *30*(1), 51–64. https://doi.org/10.1016/j.imavis.2011.11.007

Geangu, E., Ichikawa, H., Lao, J., Kanazawa, S., Yamaguchi, M. K., Caldara, R., & Turati, C. (2016). Culture shapes 7-month-olds' perceptual strategies in discriminating facial expressions of emotion. *Current Biology*, *26*(14), 663–664. https://doi.org/10.1016/j.cub.2016.05.072

Gibson, J. J. (1966). The senses considered as perceptual systems.

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of psychology*, *66*(3), 325–331.

Gosselin, F., & Schyns, P. G. (2001). Bubbles: A technique to reveal the use of information in recognition tasks. *Vision Research*, *41*(17), 2261–2271. https://doi.org/10.1016/S0042-6989(01)00097-9

Green, M. J., Williams, L., & Davidson, D. (2003). Visual scanpaths to threat-related faces in deluded schizophrenia. *Psychiatry research*, *119*(3), 271–285.

Grimes, J. (1996). On the failure to detect changes in scenes across saccades.

Groner, R., Walder, F., & Groner, M. (1984). Looking at faces: Local and global aspects of scanpaths. *Advances in psychology*, *22*, 523–533.

Guo, K. (2012). Holistic gaze strategy to categorize facial expression of varying intensities. *PLoS One*, *7*(8), e42585.

Hansen, D. W., & Ji, Q. (2009). In the eye of the beholder: A survey of models for eyes and gaze. *IEEE transactions on pattern analysis and machine intelligence*, *32*(3), 478–500.

Hartridge, H., & Thomson, L. C. (1948). Methods of investigating eye movements. *The British journal of ophthalmology*, *32*(9), 581.

Hayhoe, M., & Ballard, D. H. (2005). Eye movements in natural behavior. *Trends in cognitive sciences*, *9*(4), 188–194.

Hayhoe, M., McKinney, T., Chajka, K., & Pelz, J. B. (2012). Predictive eye movements in natural vision. *Experimental Brain Research*, *217*(1), 125–136. https://doi.org/10.1007/s00221-011-2979-2

He, H., Xu, B., & Tanaka, J. (2016). Investigating the face inversion effect in a deaf population using the dimensions tasks. *Visual Cognition*, *24*(3), 201–211.

He, Y., Scholz, J. M., Gage, R., Kallie, C. S., Liu, T., & Legge, G. E. (2015). Comparing the visual spans for faces and letters. *Journal of Vision*, *15*(8), 1–13. https://doi.org/10.1167/15.8.7.

Hedley, D., Brewer, N., & Young, R. (2015). The effect of inversion on face recognition in adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *45*(5), 1368–1379.

Hegdé, J. (2008). Time course of visual perception: Coarse-to-fine processing and beyond. *Progress in neurobiology*, *84*(4), 405–439.

Henderson, J. M. (1992). Object identification in context: The visual processing of natural scenes. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *46*(3), 319.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in cognitive sciences*, *7*(11), 498–504.

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes, 537–562. https://doi.org/10.1016/B978-008044980-7/50027-6

Henderson, J. M., & Ferreira, F. (2004). Scene perception for psycholinguists.

Henderson, J. M., Ferreira, F., Rayner, K., & Pollatsek, A. (1990). Effects of foveal processing difficulty on the perceptual span in reading: Implications for attention and eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(3), 41–429. https://doi.org/10.1037/0278-7393.16.3.417.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual review of psychology, 50*(1), 243–271.

Henderson, J. M., Pollatsek, A., & Rayner, K. (1989). Covert visual attention and extrafoveal information use during object identification. *Perception & psychophysics, 45*(3), 196–208.

Henderson, J. M., Weeks Jr, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of experimental psychology: Human perception and performance, 25*(1), 210.

Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. *Memory & Cognition, 33*(1), 98–106. https://doi.org/10.3758/BF03195300.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and brain sciences, 33*(2-3), 61–83.

Hering, E. (1879). Der raumsinn und die bewegungen des auges. *Handbuch der Physiologie, Vogel, 3*, 434–601.

Hills, P. J. (2018). Children process the self face using configural and featural encoding: Evidence from eye tracking. *Cognitive Development, 48*, 82–93.

Hills, P. J., Cooper, R. E., & Pake, J. M. (2013). First fixations in face processing: The more diagnostic they are the smaller the face-inversion effect. *Acta Psychologica, 142*(2), 211–219.

Hills, P. J., Marquardt, Z., Young, I., & Goodenough, I. (2017). Explaining sad people's memory advantage for faces. *Frontiers in psychology, 8*, 207.

Hillstrom, A. P., & Yantis, S. (1994). Visual motion and attentional capture. *Perception & Psychophysics, 55*(4), 399–411. https://doi.org/10.3758/BF03205298

Hoffman, M. D., Gelman, A. et al. (2014). The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res., 15*(1), 1593–1623.

Hollands, M. A., & Marple-Horvat, D. E. (2001). Coordination of eye and leg movements during visually guided stepping. *Journal of Motor Behavior*, *33*(2), 205–216. https://doi.org/10.1080/00222890109603151

Hollands, M. A., Marple-Horvat, D. E., Henkes, S., & Rowan, A. K. (1995). Human eye movements during visually guided stepping. *Journal of Motor Behavior*, *27*(2). https://doi.org/10.1080/00222895.1995.9941707

Hollingworth, A. (2006). Scene and position specificity in visual memory for objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(1), 58.

Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(1), 113.

Hsiao, J. H., & Cottrell, G. (2008). Two fixations suffice in face recognition. *Psychological Science*, *19*(10), 998–1006. https://doi.org/10.1111/j.1467-9280.2008.02191.x.

Hsiao, J. H., Lan, H., Zheng, Y., & Chan, A. B. (2021). Eye movement analysis with hidden markov models (emhmm) with co-clustering. *Behavior Research Methods*, 1–14.

Hsiao, J. H., & Liu, T. T. (2012). The optimal viewing position in face recognition. *Journal of Vision*, *12*(2), 22–22.

Hu, C., Wang, Q., Fu, G., Quinn, P. C., & Lee, K. (2014). Both children and adults scan faces of own and other races differently. *Vision research*, *102*, 1–10.

Huey, E. B. (1898). Preliminary experiments in the physiology and psychology of reading. *The American Journal of Psychology*, *9*(4), 575–586.

Huey, E. B. (1900). On the psychology and physiology of reading. i. *The American Journal of Psychology*, *11*(3), 283–302.

Inhoff, A. W., & Liu, W. (1998). The perceptual span and oculomotor activity during the reading of chinese sentences. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(1), 20–34. https://doi.org/10.1037/0096-1523.24.1.20

Intraub, H. (1980). Presentation rate and the representation of briefly glimpsed pictures in memory. *Journal of Experimental Psychology: Human Learning and Memory, 6*(1), 1.

Intraub, H. (1981). Rapid conceptual identification of sequentially presented pictures. *Journal of Experimental Psychology: Human Perception and Performance, 7*(3), 604.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision research, 40*(10-12), 1489–1506.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience, 2*(3), 194–203. https://doi.org/10.1038/35058500

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence, 20*(11), 1254–1259.

Jack, R. E., Blais, C., Scheepers, C., Schyns, P. G., & Caldara, R. (2009). Cultural confusions show that facial expressions are not universal. *Current Biology, 19*. https://doi.org/10.1016/j.cub.2009.07.051

Jarodzka, H., Holmqvist, K., & Nyström, M. (2010). A vector-based, multidimensional scanpath similarity measure. *Proceedings of the 2010 symposium on eye-tracking research & applications*, 211–218.

Javal, E. (1878). Essai sur la physiologie de la lecture. *Annales d'Ocilistique, 80*, 97–117.

Javal, E. (1879). Conditions de la lecture facile. *Comptes rendus de la Societie de Biologie*, 8.

Jeffery, L., & Rhodes, G. (2011). Insights into the development of face recognition mechanisms revealed by face aftereffects. *British Journal of Psychology, 102*(4), 799–815.

Judd, C. H. (1905a). Movement and consciousness. *The Psychological Monograph, 7*, 199–226.

Judd, C. H. (1905b). The muller-lyer illusion. *The Psychological Review: Monograph Supplements*.

Judd, T., Durand, F., & Torralba, A. (2012). A benchmark of computational models of saliency to predict human fixations.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological review*, *87*(4), 329.

Kanan, C., Bseiso, D. N. F., Ray, N. A., Hsiao, J. H., & Cottrell, G. W. (2015). Humans have idiosyncratic and task-specific scanpaths for judging faces. *Vision Research*, *108*, 67–76. https://doi.org/10.1016/j.visres.2015.01.013.

Kaunitz, L. N., Kamienkowski, J. E., Varatharajah, A., Sigman, M., Quiroga, R. Q., & Ison, M. J. (2014). Looking for a face in the crowd: Fixation-related potentials in an eye-movement visual search task. *NeuroImage*, *89*, 297–305.

Kelly, D. J., Jack, R. E., Miellet, S., De Luca, E., Foreman, K., & Caldara, R. (2011). Social experience does not abolish cultural diversity in eye movements. *Frontiers in psychology*, *2*, 95.

Kelly, D. J., Liu, S., Rodger, H., Miellet, S., Ge, L., & Caldara, R. (2011). Developing cultural differences in face processing. *Developmental Science*, *14*(5), 1176–1184. https://doi.org/10.1111/j.1467-7687.2011.01067.x.

Klein, R., & Farrell, M. (1989). Search performance without eye movements. *Perception & Psychophysics*, *46*(5), 476–482.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in psychtoolbox-3? *Perception*, *36*. https://doi.org/10.1068/v070821.

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

Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: A review of audiovisual studies. *Acta psychologica*, *134*(3), 372–384.

Kümmerer, M., Wallis, T., & Bethge, M. (2017). Deepgaze ii: Predicting fixations from deep features over time and tasks. *Journal of Vision*, *17*(10), 1147–1147.

Kwon, M., & Legge, G. E. (2012). Spatial-frequency requirements for reading revisited. *Vision Research*, *62*, 139–147. https://doi.org/10.1016/j.visres.2012.03.025.

Kwon, M., Liu, R., & Chien, L. (2016). Compensation for blur requires increase in field of view and viewing time. *PLOS ONE, 11*(9), e0162711. https://doi.org/10.1371/journal.pone.0162711.

Lamare, M. (1892). Des mouvements des yeux dans la lecture. *Bulletins et mémoires de la Société française d'Ophthalmologie, 10*, 354–364.

Lampinen, J. M., Copeland, S. M., & Neuschatz, J. S. (2001). Recollections of things schematic: Room schemas revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(5), 1211.

Land, M. F. (1992). Predictable eye-head coordination during driving. *Nature, 359*(6393), 318–320.

Land, M. F., & Furneaux, S. (1997). The knowledge base of the oculomotor system. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 352*(1358), 1231–1239.

Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision research, 41*(25-26), 3559–3565.

Land, M. F., & Horwood, J. (1995). Which parts of the road guide steering? *Nature, 377*(6547), 339–340.

Land, M. F., & Lee, D. N. (1994). Where we look when we steer. *Nature, 369*(6483), 742–744.

Land, M. F., & McLeod, P. (2000). From eye movements to actions: How batsmen hit the ball. *Nature neuroscience, 3*(12), 1340–1345.

Land, M. F., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception, 28*(11), 1311–1328.

Lao, J., Miellet, S., Pernet, C., Sokhn, N., & Caldara, R. (2017). Imap4: An open source toolbox for the statistical fixation mapping of eye movement data with linear mixed modeling. *Behavior Research Methods, 49*(12), 559–575. https://doi.org/10.3758/s13428-016-0737-x

Lappi, O. (2015). Eye tracking in the wild : The good , the bad and the ugly. *Journal of Eye Movement Research, 8*(5), 1–21. https://doi.org/10.16910/JEMR.8.5.1

Lappi, O. (2016). Eye movements in the wild: Oculomotor control, gaze behavior & frames of reference. *Neuroscience & Biobehavioral Reviews, 69*, 49–68. https://doi.org/10.1016/j. neubiorev.2016.06.006

Lê, S., Raufaste, E., Roussel, S., Puel, M., & Démonet, J.-F. (2003). Implicit face perception in a patient with visual agnosia? evidence from behavioural and eye-tracking analyses. *Neuropsychologia, 41*(6), 702–712.

Le Meur, O., & Baccino, T. (2013). Methods for comparing scanpaths and saliency maps: Strengths and weaknesses. *Behavior research methods, 45*(1), 251–266.

Leboran, V., Garcia-DIaz, A., Fdez-Vidal, X. R., & Pardo, X. M. (2017). Dynamic whitening saliency. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(5), 893–907. https://doi.org/10.1109/TPAMI.2016.2567391

Lee, D.-H., Corrow, S. L., Pancaroglu, R., & Barton, J. J. (2019). The scanpaths of subjects with developmental prosopagnosia during a face memory task. *Brain sciences, 9*(8), 188.

Legge, G. E., Mansfield, J. S., & Chung, S. T. (2001). Psychophysics of reading. xx. linking letter recognition to reading speed in central and peripheral vision. *Vision Research, 41*(6), 725–743.

Letourneau, S. M., & Mitchell, T. V. (2011). Gaze patterns during identity and emotion judgments in hearing adults and deaf users of american sign language. *Perception, 40*(5), 563–575.

Liu, S., Quinn, P. C., Xiao, N. G., Wu, Z., Liu, G., & Lee, K. (2018). Relations between scanning and recognition of own- and other-race faces in 6- and 9-month-old infants. *PsyCh Journal, 7*(2), 92–102. https://doi.org/https://doi.org/10.1002/pchj.211

Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human perception and performance, 4*(4), 565.

Lore, W. H., & Song, S. (1991). Central and peripheral visual processing in hearing and non-hearing individuals. *Bulletin of the Psychonomic Society*, *29*(5), 437–440.

Mackworth, J. F., & Mackworth, N. H. (1958). Eye fixations recorded on changing visual scenes by the television eye-marker. *JOSA*, *48*(7), 439–445.

Mackworth, N. H., & Morandi, A. J. (1967). The gaze selects informative details within pictures. *Perception & psychophysics*, *2*(11), 547–552.

Mackworth, N. H., & Thomas, E. L. (1962). Head-mounted eye-marker camera. *JOSA*, *52*(6), 713–716.

Malaspina, M., Albonico, A., Lao, J., Caldara, R., & Daini, R. (2018). Mapping self-face recognition strategies in congenital prosopagnosia. *Neuropsychology*, *32*(2), 123.

Malcolm, G. L., Lanyon, L. J., Fugard, A. J., & Barton, J. J. (2008). Scan patterns during the processing of facial expression versus identity: An exploration of task-driven and stimulus-driven effects. *Journal of Vision*, *8*(8), 2–2.

Malcolm, G. L., Nuthmann, A., & Schyns, P. G. (2014). Beyond gist: Strategic and incremental information accumulation for scene categorization. *Psychological science*, *25*(5), 1087–1097.

Malcom, G. L., & Henderson, J. M. (2009). The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements. *Journal of Vision*, *9*(11), 8–8. https://doi.org/10.1167/9.11.8

Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive psychology*, *9*(1), 111–151.

Manor, B. R., Gordon, E., Williams, L., Rennie, C. J., Bahramali, H., Latimer, C. R., Barry, R. J., & Meares, R. A. (1999). Eye movements reflect impaired face processing in patients with schizophrenia. *Biological psychiatry*, *46*(7), 963–969.

Mares, I., Smith, M. L., Johnson, M. H., & Senju, A. (2016). Direct gaze facilitates rapid orienting to faces: Evidence from express saccades and saccadic potentials. *Biological psychology*, *121*, 84–90.

Marigold, D. S., & Patla, A. E. (2007). Gaze fixation patterns for negotiating complex ground terrain. *Neuroscience, 144*(1), 302–313. https://doi.org/10.1016/j.neuroscience.2006.09.006

Martinez-Conde, S., Macknik, S. L., Troncoso, X. G., & Dyar, T. A. (2006). Microsaccades counteract visual fading during fixation. *Neuron, 49*(2), 297–305.

Masuda, T., & Nisbett, R. E. (2001). Attending holistically versus analytically: Comparing the context sensitivity of japanese and americans. *Journal of personality and social psychology, 81*(5), 922.

Matthis, J. S., Yates, J. L., & Hayhoe, M. M. (2018). Gaze and the control of foot placement when walking in natural terrain. *Current Biology, 28*(8), 1224–1233. https://doi.org/10.1016/j.cub.2018.03.008

Maw, N. N., & Pomplun, M. (2004). Studying human face recognition with the gaze-contingent window technique. *Proceedings of the Annual Meeting of the Cognitive Science Society, 26*(26).

McClure, E. B. (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychological bulletin, 126*(3), 424.

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics, 17*(6), 578–586. https://doi.org/10.3758/BF03203972.

McConkie, G. W., & Rayner, K. (1976a). Asymmetry of the perceptual span in reading. *Bulletin of the psychonomic society, 8*(5), 365–368.

McConkie, G. W., & Rayner, K. (1976b). Identifying the span of the effective stimulus in reading: Literature review and theories of reading. *Theoretical models and processes of reading*, 137–162.

McCullough, S., & Emmorey, K. (1997). Face processing by deaf asl signers: Evidence for expertise in distinguishing local features. *Journal of Deaf Studies and Deaf Education*, 212–222.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746–748.

McKone, E., Robbins, R. et al. (2011). Are faces special. *Oxford handbook of face perception,* 149–176.

Mehoudar, E., Arizpe, J., Baker, C. I., & Yovel, G. (2014). Faces in the eye of the beholder: Unique and stable eye scanning patterns of individual observers. *Journal of Vision, 14*(7), 1–11. https://doi.org/10.1167/14.7.6.

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law, 7*(1), 3–35. https://doi.org/10.1037/1076-8971.7.1.3.

Melcher, D. (2005). Spatiotopic transfer of visual-form adaptation across saccadic eye movements. *Current biology, 15*(19), 1745–1748.

Melcher, D. (2007). Predictive remapping of visual features precedes saccadic eye movements. *Nature neuroscience, 10*(7), 903–907.

Melcher, D. (2009). Selective attention and the active remapping of object features in transsaccadic perception. *Vision research, 49*(10), 1249–1255.

Melcher, D., & Kowler, E. (1999). Shapes, surfaces and saccades. *Vision research, 39*(17), 2929–2946.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics, 21*(6), 1087–1092.

Miellet, S., Caldara, R., & Schyns, P. G. (2011). Local jekyll and global hyde: The dual identity of face identification. *Psychological Science, 22*(12), 1518–1526. https://doi.org/10.1177/0956797611424290.

Miellet, S., He, L., Zhou, X., Lao, J., & Caldara, R. (2012). When east meets west: Gaze-contingent blindspots abolish cultural diversity in eye movements for faces. *Journal of Eye Movement Research, 5*(2). https://doi.org/10.16910/JEMR.5.2.5

Miellet, S., Lao, J., & Caldara, R. (2014). An appropriate use of imap produces correct statistical results: A reply to mcmanus. *Perception*, *43*(5), 451–457. https://doi.org/10.1068/p7682.

Miellet, S., O'Donnell, P. J., & Sereno, S. C. (2009). Parafoveal magnification: Visual acuity does not modulate the perceptual span in reading. *Psychological Science*, *20*(6), 721–728. https://doi.org/10.1111/j.1467-9280.2009.02364.x.

Miellet, S., Vizioli, L., He, L., Zhou, X., & Caldara, R. (2013). Mapping face recognition information use across cultures. *Frontiers in Psychology*, *4*(February), 34. https://doi.org/10.3389/fpsyg.2013.00034

Miellet, S., Zhou, X., He, L., Rodger, H., & Caldara, R. (2010). Investigating cultural diversity for extrafoveal information use in visual scenes. *Journal of vision*, *10*(6), 21–21.

Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of abnormal psychology*, *110*(1), 40.

Moore, T., & Armstrong, K. M. (2003). Selective gating of visual signals by microstimulation of frontal cortex. *Nature*, *421*(6921), 370–373.

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

Morton, J., & J.son, M. (1991). The perception of facial structure in infancy.

Murphy, J., Gray, K. L. H., & Cook, R. (2017). The composite face illusion. *Psychonomic Bulletin & Review*, *24*(2), 245–261.

Murthy, A., Thompson, K. G., & Schall, J. D. (2001). Dynamic dissociation of visual selection from saccade programming in frontal eye field. *Journal of neurophysiology*, *86*(5), 2634–2637.

Näsänen, R., & Ojanpää, H. (2004). How many faces can be processed during a single eye fixation? *Perception*, *33*(1), 67–77. https://doi.org/10.1068/p3417

Neville, H. J., & Lawson, D. (1987). Attention to central and peripheral visual space in a movement detection task. iii. separate effects of auditory deprivation and acquisition of a visual language. *Brain research, 405*(2), 284–294.

Nikolaev, A. R., Pannasch, S., Ito, J., & Belopolsky, A. V. (2014). Eye movement-related brain activity during perceptual and cognitive processing. *Frontiers in systems neuroscience, 8*, 62.

Nikolaev, A. R., Meghanathan, R. N., & van Leeuwen, C. (2016). Combining eeg and eye movement recording in free viewing: Pitfalls and possibilities. *Brain and cognition, 107*, 55–83.

Nisbett, R. E., & Masuda, T. (2003). Culture and point of view. *Proceedings of the National Academy of Sciences, 100*(19), 11163–11170.

Nobre, A. C., Gitelman, D., Dias, E., & Mesulam, M. (2000). Covert visual spatial orienting and saccades: Overlapping neural systems. *Neuroimage, 11*(3), 210–216.

Noton, D., & Stark, L. (1971). Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision research, 11*(9), 929–IN8.

Nuthmann, A. (2013). On the visual span during object search in real-world scenes. *Visual Cognition, 21*(7), 803–837. https://doi.org/10.1080/13506285.2013.832449.

Nuthmann, A., Einhäuser, W., & Schütz, I. (2017). How well can saliency models predict fixation selection in scenes beyond central bias? a new approach to model evaluation using generalized linear mixed models. *Frontiers in Human Neuroscience, 11*, 491. https://doi.org/10.3389/fnhum.2017.00491

Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges? evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology, 34*(1), 72–107. https://doi.org/10.1006/cogp.1997.0667.

Oliva, A., & Schyns, P. G. (2017). Hybrid image illusions. *The Oxford Compendium of Visual Illusions*, 763–766.

Oliva, A., Torralba, A., & Schyns, P. G. (2006). Hybrid images. *ACM Transactions on Graphics (TOG), 25*(3), 527–532.

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and brain sciences, 24*(5), 939.

Orschansky, J. (1899). Eine methode die augenbewegungen direct zu untersuchen. *Zentralblatt für Physiologie, 12*, 785–785.

Papinutto, M., Lao, J., Lalanne, D., & Caldara, R. (2020). Watchers do not follow the eye movements of walkers. *Vision Research, 176*, 130–140.

Papinutto, M., Lao, J., Ramon, M., Caldara, R., & Miellet, S. (2017). The facespan-the perceptual span for face recognition. *Journal of Vision, 17*(5). https://doi.org/10.1167/17.5.16

Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J., & Hays, J. (2016). Webgazer: Scalable webcam eye tracking using user interactions. *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI)*, 3839–3845.

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

Patla, A. E., & Vickers, J. N. (2003). How far ahead do we look when required to step on specific locations in the travel path during locomotion? *Experimental Brain Research, 148*(1), 133–138. https://doi.org/10.1007/s00221-002-1246-y

Pelli, D. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437–442. https://doi.org/10.1163/156856897X00366.

Pelli, D., Tillman, K. A., Freeman, J., Su, M., Berger, T. D., & Majaj, N. J. (2007). Crowding and eccentricity determine reading rate. *Journal of Vision, 7*(2), 1–36. https://doi.org/10.1167/7.2.20.

Pelphrey, K. A., Sasson, N. J., Reznick, J. S., Paul, G., Goldman, B. D., & Piven, J. (2002). Visual scanning of faces in autism. *Journal of autism and developmental disorders, 32*(4), 249–261.

Pelz, J. B., & Canosa, R. (2001). Oculomotor behavior and perceptual strategies in complex tasks. *Vision research*, *41*(25-26), 3587–3596.

Pelz, J. B., Canosa, R. L., Kucharczyk, D., Babcock, J. S., Silver, A., & Konno, D. (2000). Portable eyetracking: A study of natural eye movements. *Proceedings of the SPIE, 3959*, 566–583.

Pelz, J. B., Hayhoe, M., & Loeber, R. (2001). The coordination of eye, head, and hand movements in a natural task. *Experimental Brain Research*, *139*(3), 266–277. https://doi.org/10.1007/s002210100745

Peters, R. J., Iyer, A., Itti, L., & Koch, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision Research*, *45*(18), 2397–2416. https://doi.org/10.1016/j.visres.2005.03.019

Peterson, M. F., & Eckstein, M. P. (2012). Looking just below the eyes is optimal across face recognition tasks. *Proceedings of the National Academy of Sciences*, *109*(48), 3314–3323. https://doi.org/10.1073/pnas.1214269109.

Peterson, M. F., & Eckstein, M. P. (2013). Individual differences in eye movements during face identification reflect observer-specific optimal points of fixation. *Psychological Science*, *24*(7), 1216–1225. https://doi.org/10.1177/0956797612471684.

Peterson, M. F., Lin, J., Zaun, I., & Kanwisher, N. (2016). Individual differences in face-looking behavior generalize from the lab to the world. *Journal of Vision*, *16*(7), 12. https://doi.org/10.1167/16.7.12

Phillips, M. L., & David, A. S. (1997). Viewing strategies for simple and chimeric faces: An investigation of perceptual bias in normals and schizophrenic patients using visual scan paths. *Brain and Cognition*, *35*(2), 225–238.

Phillips, M. L., & David, A. S. (1998). Abnormal visual scan paths: A psychophysiological marker of delusions in schizophrenia. *Schizophrenia Research*, *29*(3), 235–245.

Piepers, D., & Robbins, R. (2012). A review and clarification of the terms "holistic,""configural," and "relational" in the face perception literature. *Frontiers in psychology*, *3*, 559.

Pomplun, M. (2006). Saccadic selectivity in complex visual search displays. *Vision research, 46*(12), 1886–1900.

Porterfield, W. (1737). An essay concerning the motions of our eyes. part i. of their external motions. *Edinburgh medical essays and observations, 3*(1737), 160–263.

Posner, M. I. (1980). Orienting of attention. *Quarterly journal of experimental psychology, 32*(1), 3–25.

Posner, M. I. (1982). Cumulative development of attentional theory. *American Psychologist, 37*(2), 168.

Posner, M. I., Nissen, M. J., & Ogden, W. C. (1978). Attended and unattended processing modes: The role of set for spatial location. *Modes of perceiving and processing information, 137*(158), 2.

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of experimental psychology: General, 109*(2), 160.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of experimental psychology: human learning and memory, 2*(5), 509.

Proksch, J., & Bavelier, D. (2002). Changes in the spatial distribution of visual attention after early deafness. *Journal of cognitive neuroscience, 14*(5), 687–701.

Rämä, P., & Baccino, T. (2010). Eye fixation–related potentials (efrps) during object identification. *Visual Neuroscience, 27*(5-6), 187–192.

Ramon, M., Bobak, A. K., & White, D. (2019). Super-recognizers: From the lab to the world and back again. *British Journal of Psychology, 110*(3), 461–479.

Ramon, M., Miellet, S., Dzieciol, A. M., Konrad, B. N., Dresler, M., & Caldara, R. (2016). Super-memorizers are not super-recognizers. *PloS One, 11*(3), e0150972. https://doi.org/10.1371/journal.pone.0150972.

Ramon, M., Sokhn, N., Lao, J., & Caldara, R. (2018). Decisional space determines saccadic reaction times in healthy observers and acquired prosopagnosia. *Cognitive neuropsychology, 35*(5-6), 304–313.

Rao, R. P. N., Zelinsky, G. J., Hayhoe, M. M., & Ballard, D. H. (2002). Eye movements in iconic visual search. *Vision Research*, *42*(11), 1447–1463. https://doi.org/10.1016/S0042-6989(02)00040-8

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive psychology*, *7*(1), 65–81.

Rayner, K. (1983). Eye movements, perceptual span, and reading disability. *Annals of Dyslexia*, *33*(1), 163–173. https://doi.org/10.1007/BF02648003.

Rayner, K. (1986). Eye movements and the perceptual span in beginning and skilled readers. *Journal of Experimental Child Psychology*, *41*(2), 211–236. https://doi.org/10.1016/0022-0965(86)90037-8.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, *124*(3), 372.

Rayner, K., & Bertera, J. H. (1979). Reading without a fovea. *Science*, *206*(4417), 468–469.

Rayner, K., Castelhano, M. S., & Yang, J. (2009). Eye movements when looking at unusual/weird scenes: Are there cultural differences? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(1), 254.

Rayner, K., Slattery, T. J., & Belanger, N. N. (2010). Eye movements, the perceptual span, and reading speed. *Psychonomic Bulletin and Review*, *17*(6), 834–839. https://doi.org/10.3758/PBR.17.6.834.

Rayner, K., Well, A. D., & Pollatsek, A. (1980). Asymmetry of the effective visual field in reading. *Perception & Psychophysics*, *27*(6), 537–544.

Rensink, R. A. (2000). Seeing, sensing, and scrutinizing. *Vision research*, *40*(10-12), 1469–1487.

Richler, J. J., & Gauthier, I. (2014). A meta-analysis and review of holistic face processing. *Psychological Bulletin*, *140*(5), 1281.

Richoz, A.-R., Lao, J., Pascalis, O., & Caldara, R. (2018). Tracking the recognition of static and dynamic facial expressions of emotion across the life span. *Journal of vision*, *18*(9), 5–5.

Ristic, J., Mottron, L., Friesen, C. K., Larocci, G., Burack, J. A., & Kingstone, A. (2005). Eyes are special but not for everyone: The case of autism. *Cognitive Brain Research*, *24*(3), 715–718.

Rizzolatti, G., Riggio, L., Dascola, I., & Umiltá, C. (1987). Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention. *Neuropsychologia*, *25*(1), 31–40.

Rodger, H., Kelly, D. J., Blais, C., & Caldara, R. (2010). Inverting faces does not abolish cultural diversity in eye movements. *Perception*, *39*(11), 1491–1503. https://doi.org/10.1068/p6750

Rolfs, M. (2009). Microsaccades: Small steps on a long way. *Vision research*, *49*(20), 2415–2441.

Ruff, C. C., Blankenburg, F., Bjoertomt, O., Bestmann, S., Freeman, E., Haynes, J.-D., Rees, G., Josephs, O., Deichmann, R., & Driver, J. (2006). Concurrent tms-fmri and psychophysics reveal frontal influences on human retinotopic visual cortex. *Current Biology*, *16*(15), 1479–1488.

Saida, S., & Ikeda, M. (1979). Useful visual field size for pattern perception. *Perception & Psychophysics*, *25*(2), 119–125.

Schauder, K. B., Park, W. J., Tsank, Y., Eckstein, M. P., Tadin, D., & Bennetto, L. (2019). Initial eye gaze to faces and its functional consequence on face identification abilities in autism spectrum disorder. *Journal of neurodevelopmental disorders*, *11*(1), 1–20.

Schmalzl, L., Palermo, R., Green, M., Brunsdon, R., & Coltheart, M. (2008). Training of familiar face recognition and visual scan paths for faces in a child with congenital prosopagnosia. *Cognitive Neuropsychology*, *25*(5), 704–729.

Schwarzer, G., Huber, S., Grüter, M., Grüter, T., Groß, C., Hipfel, M., & Kennerknecht, I. (2007). Gaze behaviour in hereditary prosopagnosia. *Psychological research*, *71*(5), 583–590.

Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show me the features! understanding recognition from the use of visual information. *Psychological Science*, *13*(5), 402–409. https://doi.org/10.1111/1467-9280.00472.

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time-and spatial-scale-dependent scene recognition. *Psychological science, 5*(4), 195–200.

Searle, A., & Rowe, F. J. (2016). Vergence neural pathways: A systematic narrative literature review. *Neuro-ophthalmology, 40*(5), 209–218.

Sethi, B. (1986). Vergence adaptation: A review. *Documenta Ophthalmologica, 63*(3), 247–263.

Shapiro, F. (1999). Eye movement desensitization and reprocessing (emdr) and the anxiety disorders: Clinical and research implications of an integrated psychotherapy treatment. *Journal of Anxiety disorders, 13*(1-2), 35–67.

Sheliga, B., Craighero, L., Riggio, L., & Rizzolatti, G. (1997). Effects of spatial attention on directional manual and ocular responses. *Experimental brain research, 114*(2), 339–351.

Spering, M., & Montagnini, A. (2011). Do we track what we see? common versus independent processing for motion perception and smooth pursuit eye movements: A review. *Vision research, 51*(8), 836–852.

Spezio, M. L., Adolphs, R., Hurley, R. S. E., & Piven, J. (2007). Abnormal use of facial information in high-functioning autism. *Journal of autism and developmental disorders, 37*(5), 929–939.

Stacchi, L., Ramon, M., Lao, J., & Caldara, R. (2019). Neural representations of faces are tuned to eye movements. *Journal of Neuroscience, 39*(21), 4113–4123.

Stahl, J. S. (1999). Amplitude of human head movements associated with horizontal saccades. *Experimental Brain Research, 126*(1), 41–54. https://doi.org/10.1007/s002210050715

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137–149. https://doi.org/10.3758/BF03207704.

Stephan, B. C. M., & Caine, D. (2009). Aberrant pattern of scanning in prosopagnosia reflects impaired face processing. *Brain and Cognition, 69*(2), 262–268.

Stivalet, P., Moreno, Y., Richard, J., Barraud, P.-A., & Raphel, C. (1998). Differences in visual search tasks between congenitally deaf and normally hearing adults. *Cognitive Brain Research*, *6*(3), 227–232.

Stoll, C., Palluel-Germain, R., Caldara, R., Lao, J., Dye, M. W., Aptel, F., & Pascalis, O. (2018). Face recognition is shaped by the use of sign language. *The Journal of Deaf Studies and Deaf Education*, *23*(1), 62–70.

Stratton, G. M. (1902). Eye-movement and the aesthetics of visual form. *Philosophische Studien*, *20*, 336–359.

Streit, M., Wölwer, W., & Gaebel, W. (1997). Facial-affect recognition and visual scanning behaviour in the course of schizophrenia. *Schizophrenia research*, *24*(3), 311–317.

Suzuki, S., & Cavanagh, P. (1995). Facial organization blocks access to low-level features: An object inferiority effect. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(4), 901.

't Hart, B. M., Vockeroth, J., Schumann, F., Bartl, K., Schneider, E., König, P., & Einhäuser, W. (2009). Gaze allocation in natural stimuli: Comparing free exploration to head-fixed viewing conditions. *Visual Cognition*, *17*(6–7), 1132–1158. https://doi.org/10.1080/13506280902812304

Tallon, M., Greenlee, M. W., Wagner, E., Rakoczy, K., & Frick, U. (2021). How do art skills influence visual search?–eye movements analyzed with hidden markov models. *Frontiers in psychology*, *12*, 65.

Tanaka, J. W., & Simonyi, D. (2016). The "parts and wholes" of face recognition: A review of the literature. *Quarterly Journal of Experimental Psychology*, *69*(10), 1876–1889.

Tanaka, J. W., & Sung, A. (2016). The "eye avoidance" hypothesis of autism face processing. *Journal of autism and developmental disorders*, *46*(5), 1538–1552.

Targino Da Costa, A. L. N., & Do, M. N. (2014). A retina-based perceptually lossless limit and a gaussian foveation scheme with loss control. *IEEE Journal of Selected Topics in Signal Processing*, *8*(3), 438–453.

Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, 7(14), 4. https://doi.org/10.1167/7.14.4

Tatler, B. W. (2009). Current understanding of eye guidance. *Visual Cognition*, 17(6-7), 777–789.

Tatler, B. W., Baddeley, R. J., & Gilchrist, L. D. (2005). Visual correlates of fixation selection: Effects of scale and time. *Vision research*, 45(5), 643–659.

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

Tatler, B. W., & Land, M. F. (2011). Vision and the representation of the surroundings in spatial memory. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1564), 596–610.

Tatler, B. W., & Vincent, B. T. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research*, 2(2).

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *nature*, 381(6582), 520–522.

Ticcinelli, V., De Lissa, P., Lalanne, D., Miellet, S., & Caldara, R. (2019). Imap4d: An open source toolbox for statistical fixation mapping of eye-tracking data in virtual reality. *Journal of Vision*, 19(10), 127c. https://doi.org/10.1167/19.10.127c

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological review*, 113(4), 766.

Underwood, G., Foulsham, T., & Humphrey, K. (2009). Saliency and scan patterns in the inspection of real-world scenes: Eye movements during encoding and recognition. *Visual Cognition*, 17(6-7), 812–834.

Underwood, N. R., & McConkie, G. W. (1985). Perceptual span for letter distinctions during reading. *Reading Research Quarterly*, 153–162.

Unema, P. J., Pannasch, S., Joos, M., & Velichkovsky, B. M. (2005). Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual cognition, 12*(3), 473–494.

Valiente-Gómez, A., Moreno-Alcázar, A., Treen, D., Cedrón, C., Colom, F., Perez, V., & Amann, B. L. (2017). Emdr beyond ptsd: A systematic literature review. *Frontiers in psychology, 8*, 1668.

Van Belle, G., Busigny, T., Lefèvre, P., Joubert, S., Felician, O., Gentile, F., & Rossion, B. (2011). Impairment of holistic face perception following right occipito-temporal damage in prosopagnosia: Converging evidence from gaze-contingency. *Neuropsychologia, 49*(11), 3145–3150.

Van Belle, G., De Graef, P., Verfaillie, K., Busigny, T., & Rossion, B. (2010). Whole not hole: Expert face recognition requires holistic perception. *Neuropsychologia, 48*(9), 2620–2629.

Van Belle, G., De Graef, P., Verfaillie, K., Rossion, B., & Lefevre, P. (2010). Face inversion impairs holistic perception: Evidence from gaze-contingent stimulation. *Journal of Vision, 10*(5), 1–13. https://doi.org/10.1167/10.5.10.

Van Rijsbergen, N. J., & Schyns, P. G. (2009). Dynamics of trimming the content of face representations for categorization in the brain. *PLoS Computational Biology, 5*(11), e1000561.

van Diepen, P. M. J., Ruelens, L., & d'Ydewalle, G. (1999). Brief foveal masking during scene perception. *Acta Psychologica, 101*(1), 91–103.

VanRullen, R., & Thorpe, S. J. (2001). The time course of visual processing: From early perception to decision-making. *Journal of cognitive neuroscience, 13*(4), 454–461.

Vettori, S., Van der Donck, S., Nys, J., Moors, P., Van Wesemael, T., Steyaert, J., Rossion, B., Dzhelyova, M., & Boets, B. (2020). Combined frequency-tagging eeg and eye-tracking measures provide no support for the "excess mouth/diminished eye attention" hypothesis in autism. *Molecular autism, 11*(1), 1–22.

Vig, E., Dorr, M., & Barth, E. (2009). Efficient visual coding and the predictability of eye movements on natural movies. *Spatial Vision, 22*(5).

Vishwanath, D., & Kowler, E. (2003). Localization of shapes: Eye movements and perception compared. *Vision Research, 43*(15), 1637–1653.

Vishwanath, D., & Kowler, E. (2004). Saccadic localization in the presence of cues to three-dimensional shape. *Journal of Vision, 4*(6), 4–4.

Viviani, P. (1990). Eye movements in visual search: Cognitive, perceptual and motor control aspects. *Reviews of oculomotor, 4*, 353–393.

Voorhies, R. C., Elazary, L., & Itti, L. (2012). Neuromorphic bayesian surprise for far-range event detection. *2012 IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance*, 1–6.

Vuilleumier, P. (2000). Faces call for attention: Evidence from patients with visual extinction. *Neuropsychologia, 38*(5), 693–700.

Vuilleumier, P., & Schwartz, S. (2001). Emotional facial expressions capture attention. *Neurology, 56*(2), 153–158.

Walker-Smith, G. J., Gale, A. G., & Findlay, J. M. (1977). Eye movement strategies involved in face perception. *Perception, 6*(3), 313–326.

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing, 13*(4), 600–612. https://doi.org/10.1109/TIP.2003.819861.

Watson, A. B., & Pelli, D. G. (1983). Quest: A bayesian adaptive psychometric method. *Perception & psychophysics, 33*(2), 113–120.

Westheimer, G., & McKee, S. P. (1975). Visual acuity in the presence of retinal-image motion. *JOSA, 65*(7), 847–850.

Wheeler, A., Anzures, G., Quinn, P. C., Pascalis, O., Omrin, D. S., & Lee, K. (2011). Caucasian infants scan own-and other-race faces differently. *PloS one, 6*(4), e18621.

Williams, C. C., & Henderson, J. M. (2007). The face inversion effect is not a consequence of aberrant eye movements. *Memory & Cognition, 35*(8), 1977–1985.

Williams, L. (1966). The effect of target specification on objects fixated during visual search. *Perception & Psychophysics, 1*(5), 315–318.

Williams, L., Loughland, C. M., Gordon, E., & Davidson, D. (1999). Visual scanpaths in schizophrenia: Is there a deficit in face recognition? *Schizophrenia research, 40*(3), 189–199.

Winston, J. S., Vuilleumier, P., & Dolan, R. J. (2003). Effects of low-spatial frequency components of fearful faces on fusiform cortex activity. *Current Biology, 13*(20), 1824–1829.

Xiao, N. G., Perrotta, S., Quinn, P. C., Wang, Z., Sun, Y.-H. P., & Lee, K. (2014). On the facilitative effects of face motion on face recognition and its development. *Frontiers in psychology, 5*, 633.

Xiao, N. G., Quinn, P. C., Liu, S., Ge, L., Pascalis, O., & Lee, K. (2015). Eye tracking reveals a crucial role for facial motion in recognition of faces by infants. *Developmental psychology, 51*(6), 744.

Xivry, J.-J. O., Ramon, M., Lefèvre, P., & Rossion, B. (2008). Reduced fixation on the upper area of personally familiar faces following acquired prosopagnosia. *Journal of Neuropsychology, 2*(1), 245–268. https://doi.org/10.1348/174866407X260199.

Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human perception and performance, 10*(5), 601.

Yin, R. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81*(1), 141.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception, 16*(6), 747–759.

Yu, D., Cheung, S.-H., Legge, G. E., & Chung, S. T. (2007). Effect of letter spacing on visual span and reading speed. *Journal of vision, 7*(2), 2–2.

Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological review, 115*(4), 787.

Zhao, M., Gersch, T. M., Schnitzer, B. S., Dosher, B. A., & Kowler, E. (2012). Eye movements and attention: The role of pre-saccadic shifts of attention in perception, memory and the control of saccades. *Vision research, 74*, 40–60.

Books

American Psychological Association. (2020). *Publication manual of the american psychological association (7th ed.)* American Psychological Association. https://doi.org/10.1037/0000165-000

Bang, S., Kim, D., & Choi, S. (2001). Asian face image database.

Bear, M., Connors, B., & Paradiso, M. (2007). *Neurosciences: à la découverte du cerveau* (A. Nieoullon, Trans.). Pradel Editions.

Buswell, G. T. (1935). *How people look at pictures: A study of the psychology and perception in art.* Univ. Chicago Press.

Capretto, T., Piho, C., Kumar, R., Westfall, J., Yarkoni, T., & Martin, O. A. (2020). Bambi: A simple interface for fitting bayesian linear models in python.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Erlbaum.

Erdmann, B., & Dodge, R. (1898). *Psychologische untersuchungen über das lesen auf experimenteller grundlage.* Niemeyer.

Fawcett, J. M., Risko, E. F., & Kingstone, A. (2015). *The handbook of attention.* MIT Press.

Felten, D. L., O'Banion, M. K., & Maida, M. E. (2015). *Netter's atlas of neuroscience.* Elsevier Health Sciences.

Findlay, J. M., & Gilchrist, L. D. (2003). *Active vision: The psychology of looking and seeing.* Oxford University Press.

Forster, E. S. (1963). *The works of aristotle. Problemata* (W. D. Ross & J. A. Smith, Trans.). Clarendon Press.

Galien, C. (1970). *Galen and the usefulness of the parts of the body. De l'usage des parties du corps humain* (M. T. May, Trans.). Cornell University Press.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. CRC press.

Hall, E. T. (1966). *The hidden dimension*. Doubleday. https://doi.org/10.1017/CBO9781107415324.004.

Henderson, J. M., & Hollingworth, A. (1998). Eye movements during scene viewing: An overview. *Eye guidance in reading and scene perception* (pp. 269–293). Elsevier.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.

Jacob, R. J. K., & Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *The mind's eye* (pp. 573–605). Elsevier.

Kümmerer, M., Wallis, T. S. A., & Bethge, M. (2018). Saliency benchmarking made easy: Separating models, maps and metrics. In V. Ferrari, M. Hebert, C. Sminchisescu, & Y. Weiss (Eds.), *Computer vision – ECCV 2018* (pp. 798–814). Springer International Publishing.

Land, M. F., & Nilsson, D.-E. (2012). *Animal eyes*. Oxford University Press.

Legge, G. E. (2006). *Psychophysics of reading in normal and low vision*. CRC Press.

Liddell, S. (2003). *Grammar, gesture, and meaning in american sign language*. Cambridge.

Liversedge, S., Gilchrist, I., & Everling, S. (2011). *The oxford handbook of eye movements*. Oxford University Press.

Lundqvist, D., Flykt, A., & Öhman, A. (1998). The karolinska directed emotional faces (kdef).

Milner, D., & Goodale, M. (2006). *The visual brain in action* (Vol. 27). OUP Oxford.

Porterfield, W. (1759). *A treatise on the eye: The manner and phænomena of vision...* (Vol. 1). A. Miller at London, for G. Hamilton; J. Balfour at Edinburgh.

Purves, D., Augustine, G., Fitzpatrick, D., Hall, W., LaMantia, A., & White, L. (2011). *Neurosciences* (J. Coquery, N. Tajeddine, & P. Gailly, Trans.). De Boeck supérieur.

Qu, Q.-X., Zhang, L., Chao, W.-Y., & Duffy, V. (2017). User experience design based on eye-tracking technology: A case study on smartphone apps. *Advances in applied digital human modeling and simulation* (pp. 303–315). Springer.

Riche, N., & Mancas, M. (2016a). *Bottom-up saliency models for still images: A practical review.* Springer. https://doi.org/10.1007/978-1-4939-3435-5_9

Riche, N., & Mancas, M. (2016b). Bottom-up saliency models for videos: A practical review. *From human attention to computational attention* (pp. 177–190). Springer.

Rizzolatti, G., Riggio, L., & Sheliga, B. M. (1994). Space and selective attention. *Attention and performance xv: Conscious and nonconscious information processing.* The MIT Press. https://doi.org/10.7551/mitpress/1478.003.0016

Spector, R. H. (1990). Visual fields. *Clinical methods: The history, physical, and laboratory examinations.* Walker, H. K.; Hall W. D.; Butterworths J. W. H.

Stein, B. E. (2012). *The new handbook of multisensory processing.* Mit Press.

Thewissen, J. G., & Nummela, S. (2008). *Sensory evolution on the threshold: Adaptations in secondarily aquatic vertebrates.* Univ of California Press.

Van Gompel, R. P. G. (2007). *Eye movements: A window on mind and brain.* Elsevier.

Wildt, A. R., Wildt, A. A., Ahtola, O., & Ahtola, O. T. (1978). *Analysis of covariance* (Vol. 12). Sage.

Woodworth, R. S. (1938). *Experimental psychology.* Holt.

Yarbus, A. L. (1967). *Eye movements and vision.* Plenum Press.

Michael Papinutto

Assistant doctorant

32 ans
Suisse/Italien
078 715 81 30
michaelpapinutto@live.fr
michaelpapinutto.ch

Compétences

Langages de programmation et logiciels statistique
(Connaissances en fonction de la taille)

PYTHON MATLAB R C JAVA SPSS JASP SQL

Statistiques
(Connaissances en fonction de la taille)

FREQUENTISTES INFERENTIELLES BAYESIENNES SERIES TEMPORELLES DESCRIPTIVES CONTINGENCES CAUSALES

Logiciels de mise en forme
Microsoft Office | Apple iWork | LATEX

Langues (A1|A2|B1|B2|C1|C2)

Français ● ● ● ● ● ●
Anglais ● ● ● ● ● ○
Allemand ● ● ● ○ ● ○
Italien ● ● ● ○ ○ ○

Intérêts et Activités

Squash | Marche | Moto
Jeux Vidéos | Domotique

Far better an approximate answer to the right question than an exact answer to the wrong question

(John Tukey, 1962)

FORMATIONS

2016 - ... **Bachelor en Informatique** | Obtention prévue: 09/21 📍 **Université de Fribourg, Fribourg**

2015 - ... **Doctorat en Psychologie** | Obtention prévue: 09/21 - **H**uman Centered **I**nteraction **S**cience and **T**echnology Institute (Human-IST Institute, Prof. Lalanne) & **I**(Eye) and **B**rain **M**apping **Lab**oratory (iBMLab, Prof. Caldara) 📍 **Université de Fribourg**

2021 **Winter School in Data Analytics and Machine Learning** | 📍 **Université de Fribourg**

2010 - 2015 **Bachelor et Master en Psychologie** | Magna et Insigni Cum Laude 📍 **Université de Fribourg**

2010 **First Certificate in English** | 📍 **Wall Street English, Neuchâtel**

2005 - 2009 **Lycée en Biologie/Chimie** | 📍 **Lycée Denis de Rougemont, Neuchâtel**

EXPÉRIENCES PROFESSIONNELLES

05.2021 - ... **Assistant doctorant | 85%, and 100% from 09.2021** iBMLab 📍 **Université de Fribourg**
- Conduite du projet de doctorat et d'autres projets
- Programmation d'expérimentation
- Aide à la Supervision d'étudiant-e-s de Master

01.2021 - 04.2021 **Collaborateur scientifique | 60%** Projet OFEN: Lucideles, Dr. Nembrini 📍 **Université de Fribourg**
- Gestion de la méthodologie expérimentale
- Mise en forme et analyse de données
- Production de communication scientifique

09.2020 - 08.2021 **Collaborateur scientifique | 15%** Fin de contrat: 31/08/21 - Équipe de Psychologie Clinique et de la Santé 📍 **Université de Fribourg**
- Programmation d'expérimentation
- Mise en forme de données
- Gestion d'un serveur d'expérimentation

2015 - 2020 **Assistant doctorant | 80%** Human-IST Institute & iBMLab 📍 **Université de Fribourg**
- Conduite du projet de doctorat et d'autres projets
- Supervision d'étudiant-e-s de Bachelor (10 étudiant-e-s)
- Chargé du cours magistral *Travaux pratiques de recherche en Psychologie* (96 étudiant-e-s)

2013 - 2015 **Assistant étudiant | 100%** iBMLab 📍 **Université de Fribourg**
- Gestion des Laboratoires
- Assistance dans la recherche
- Supervision d'étudiant-e-s de Bachelor (22 étudiant-e-s)

Publications List

Peer Reviewed Journal Articles

Dunn, J. D., de Lima Varela, V. P., Nicholls, V. I., **Papinutto**, **M.,** White, D., & Miellet, S. (2021). Visual information sampling in super-recognizers. https://doi.org/10.31234/osf.io/z2k4a

Papinutto, **M.,** Lalanne, D., & Nembrini, J. (In preparation-a). Behavioural reaction lag to decreasing light: Impacts of task difficulty and physiological indices potentials.

Papinutto, **M.,** Lalanne, D., & Nembrini, J. (In preparation-b). The circumplexes to evaluate light: Assesement of luminaires design, light quality and emotion brought by luminaires through the model of affects.

***Papinutto**, **M.,** Lao, J., Dye, M., Pascalis, O., Lalanne, D., & Caldara, C. (Submitted). Non-hearing have an amplified facial information use. *Journal of Vision*.

***Papinutto**, **M.,** Lao, J., Lalanne, D., & Caldara, R. (2020). Watchers do not follow the eye movements of walkers. *Vision Research, 176*, 130–140.

***Papinutto**, **M.,** Lao, J., Ramon, M., Caldara, R., & Miellet, S. (2017). The facespan—the perceptual span for face recognition. *Journal of Vision, 17*(5), 16–16.

Papinutto, **M.**, Nembrini, N., & Lalanne, D. (2020). "working in the dark?" investigation of physiological and psychological indices and prediction of back-lit screen users' reactions to light dimming. *Building and Environment, 186,* 107356.

Richoz, A.-R., **Papinutto**, **M.**, Ticcinelli, V., Schaller, P., Lao, J., & Caldara, R. (In preparation). Setting the record straight: Dynamic but not static facial expressions are better recognised.

* refers to a contribution presented in this thesis

CONFERENCE PROCEEDINGS ARTICLES

Alavi, H. S., Verma, H., **Papinutto**, **M.**, & Lalanne, D. (2017). Comfort: A coordinate of user experience in interactive built environments. *IFIP conference on human-computer interaction,* 247–257.

Basurto, C., Boghetti, R., Colombo, M., **Papinutto**, **M.**, Nembrini, J., & Kämpf, J. H. (Accepted). Implementation of machine learning techniques for the quasi real-time blind and electric lighting optimization in a controlled experimental facility. *Journal of Physics: Conference Series.*

Nembrini, J., Kampf, J. H., **Papinutto**, **M.**, & Lalanne, D. (2019). A smart luminaire in an office environment: Impact on light distribution, user interactions and comfort. *Journal of Physics: Conference Series, 1343*(1), 012164.

Papinutto, **M.**, Colombo, M., M.and Golsouzidou, Reutter, K., Lalanne, D., & Nembrini, J. (Accepted). Towards the integration of personal task-lighting in an optimized balance between electric lighting and daylighting: A user-centred study of emotion, visual comfort, interaction and form-factor of task lights. *Journal of Physics: Conference Series.*

Zhong, S., **Papinutto**, **M.**, Rosset, L., Lalanne, D., & Alavi, H. (Submitted). Binaural audio in hybrid meetings: Effects on speaker identification, comprehension, and user experience. *Proceedings of the ACM on Human Computer Interaction.*

Talks

Papinutto, **M.**, Caldara, R., Lao, J., & Miellet, S. (2014). The facespan: Isolating the perceptual span for face recognition. *Talk presented at The Fribourg Day of Cognition. Fribourg, Switzerland.*

Papinutto, **M.**, Caldara, R., Lao, J., & Miellet, S. (2015). The facespan: Isolating the perceptual span for face recognition. *Presented at The European Conference on Eye Movements (ECEM15). Vienna, Austria, 8*(4), 126.

Papinutto, **M.**, Colombo, M., Golsouzidou, M., Reutter, K., Lalanne, D., & Nembrini, J. (2021). Towards the integration of personal task-lighting in an optimized balance between electric lighting and daylighting: A user-centred study of emotion, visual comfort, interaction and form-factor of task lights. *Presented at CISBAT 2021. Lausanne, Switzerland.*

Invited Talks

Papinutto, **M.** (2016). The facespan: Isolating the perceptual span for face recognition. *Talk presented at the Colloquium of the psychology department, University of Bournemouth, United Kingdom.*

Papinutto, **M.** (2018). Light in human-building interaction: Reaction time to uncomfortable light and its impact on productivity. *Talk presented at the Colloquium of the psychology department, University of Neuchâtel, Switzerland.*

Posters

Papinutto, **M.**, Lao, J., Lalanne, D., & Caldara, R. (2017). Watchers do not follow the eye movements of walker. *Presented at the 15th SSP/SGP conference, Lausanne, Switzerland.*

Papinutto, **M.,** Lao, J., Lalanne, D., & Caldara, R. (2018). Watchers do not follow the eye movements of walker. *Poster presented at the Vision Science Society conference (VSS18). Florida, United States of America, 18*(10), 244.

Déclaration sur l'honneur

Par ma signature, j'atteste avoir rédigé personnellement cette thèse sans aide extérieure non autorisée, n'avoir utilisé que les sources et moyens autorisés, et mentionné comme telles les citations et paraphrases. Cette thèse n'a pas déjà été présentée devant une autre faculté.

J'ai pris connaissance que le comportement scientifique délictueux selon les directives de l'Université de Fribourg* est sanctionné.

Nom(s) : Papinutto   Prénom(s) : Michael

Lieu et date : Fribourg, le 08.09.21

Signature manuscrite :

Références bases légales
- Règlement du 11 décembre 2014 pour l'obtention du doctorat à la Faculté des lettres de l'Université de Fribourg
- Règlement du 18 décembre 1990 (Etat le 22 janvier 2004) de doctorat de la Faculté des lettres de l'Université de Fribourg

* Directives du 13 mai 2008 concernant la procédure de prononcé des sanctions disciplinaires selon l'art. 101 des Statuts du 31 mars 2000 de l'Université de Fribourg dans les cas de violation des règles de l'intégrité scientifique lors de la rédaction de travaux pendant la durée de la formation

Décanat de la Faculté des lettres
et des sciences humaines   T +41 26 300 75 01   decanat-lettres@unifr.ch