Vision Research 176 (2020) 130–140

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

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

ARTICLE INFO

Keywords:
Saliency
Eye movements
Bottom-up processes
Top-down processes
Ecological validity

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 show 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.

1. 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, 1905; 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, Bonnar, & Gosselin, 2002), which flexibly adjusts on task constraints (*e.g.* Geangu et al., 2016; Jack, Blais, Scheepers, Schyns, & Caldara, 2009; Kanan, Bseiso, Ray, Hsiao, & Cottrell, 2015), information quantity (Caldara, Zhou, & Miellet, 2010; Miellet, He, Zhou, Lao, & Caldara, 2012; Papinutto, Lao, Ramon, Caldara, & Miellet, 2017; for a review see Caldara, 2017) and quality (Miellet, Caldara, & Schyns, 2011), culture (Blais, Jack, Scheepers, Fiset, & Caldara, 2008; Kelly et al., 2011; Miellet, Vizioli, He, Zhou, & Caldara, 2013; Rodger, Kelly, Blais, & Caldara, 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, Ramon, Lao, & Caldara, 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.* Malcolm & Henderson, 2009; Rao, Zelinsky, Hayhoe, & Ballard, 2002; Zelinsky, 2008). In ecological spatial

* Corresponding author at: Bd. de Pérolles 90, 1700 Fribourg, Switzerland.
E-mail address: michael.papinutto@unifr.ch (M. Papinutto).

https://doi.org/10.1016/j.visres.2020.08.001
Received 6 February 2020; Received in revised form 3 August 2020; Accepted 5 August 2020

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, Iyer, Itti, & Koch, 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, Elazary, & Itti, 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, & Kingstone, 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, Lin, Zaun, & Kanwisher, 2016) and both groups exhibited similar central bias (Foulsham 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 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, Guyader, Ionescu, & Caplier, 2012). Other aspects might also induce differences, such as head movements in real-world conditions ('t Hart et al., 2009; Pelz, Hayhoe, & Loeber, 2001), technical issues due to parallax errors in mobile eye-tracking (Evans, Jacobs, Tarduno, & Pelz, 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 Fig. 1 – Lappi, 2015, 2016). 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, McKinney, Chajka, & Pelz, 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 footsteps by looking two steps ahead during real-world walking (Hollands & Marple-Horvat, 2001; Marigold & Patla, 2007; Matthis, Yates, & Hayhoe, 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 were 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, Judd, Oliva, Torralba, & Durand, 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.

2. Methods

2.1. 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

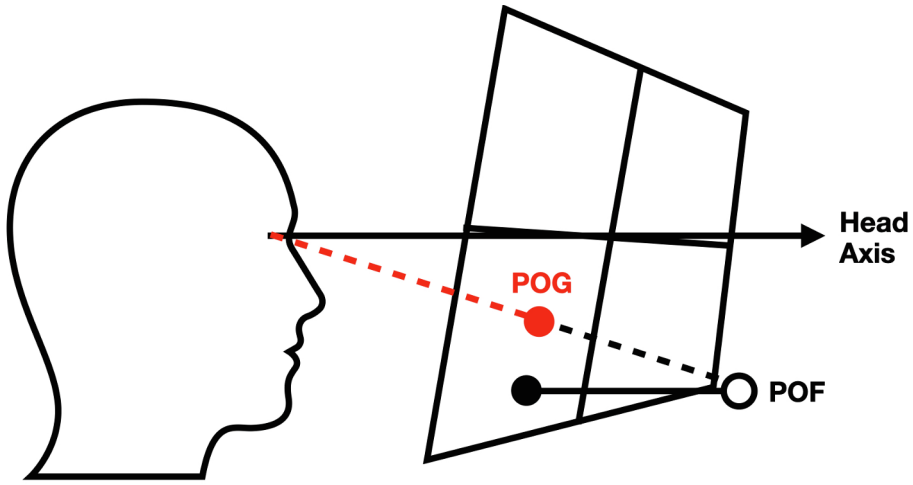

Fig. 1. Schematic representation of differences in reference frame in mobile eye-tracking. The black arrow represents the orientation of the glasses and is crossing the virtual plane. On this plane, there are two dots: the red one indicates the point of gaze (POG) and the black one is the projection of the point of fixation (POF, empty point). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
Adapted from Lappi (2015)

(28) = 0.14, $p > .05$; $BF_{10} = 0.36$. 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 experiment. All participants gave oral informed consent and the protocol was approved by the ethical committees of the Department of Psychology of Fribourg University, Switzerland.

2.2. Route

The route consisted of a walking path of about 200 m 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 Fig. 2). Participants were tested during the working hours leading to high probabilities of encountering people inside the building.

2.3. 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 × 1080-pixel resolution), subtending 44.80° horizontally and 33.60° vertically of visual angle. All videos were presented at a distance of 50 cm which is the minimal distance allowed to record eye movements with the SR Research Desktop-Mount EyeLink 2 K 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.

2.4. 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 0.5°. 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 2 K eye-tracker (with a chin and forehead rest), which has an average gaze position error of about 0.5° and a spatial resolution of 0.01°. 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, Brainard, & Pelli, 2007; Pelli, 1997) and EyeLink Toolbox extensions (Cornelissen, Peters, & Palmer, 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

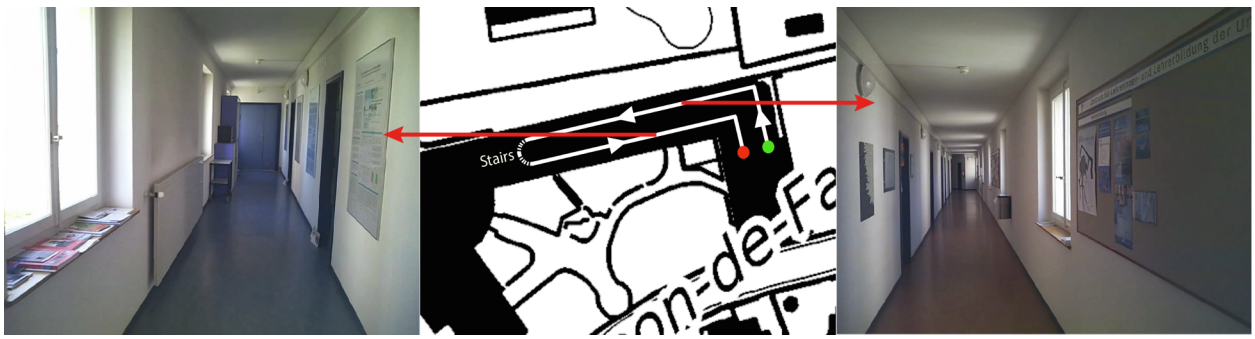

Fig. 2. Representation of the walking path that Walkers took. Left and right pictures show capture frame from the environment at two different stages of the path.

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.

2.5. 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 2 K 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 min and the laboratory condition lasted for about 50 min on average.

2.6. Data analysis

Raw eye movements were used as well as preprocessed data. Sampling rates were matched across conditions. Raw eye movements data were 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 (< 20 ms) 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, Judd, Oliva, Torralba, & Durand, 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, Martinetz, Gegenfurtner, & Barth, 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, Garcia-Diaz, Fdez-Vidal and Pardo (2017) whose previous static algorithm (Garcia-Diaz, Fdez-Vidal, Pardo, & Dosil, 2012) obtained a good evaluation in the MIT benchmarks in Judd, Durand and Torralba (2012) and in Borji, Sihite, and Itti (2013). Furthermore, the AWS algorithm was evaluated to have a low correlation with central bias (Nuthmann, Einhäuser, & Schütz, 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, Dorr, & Barth, 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 B). 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

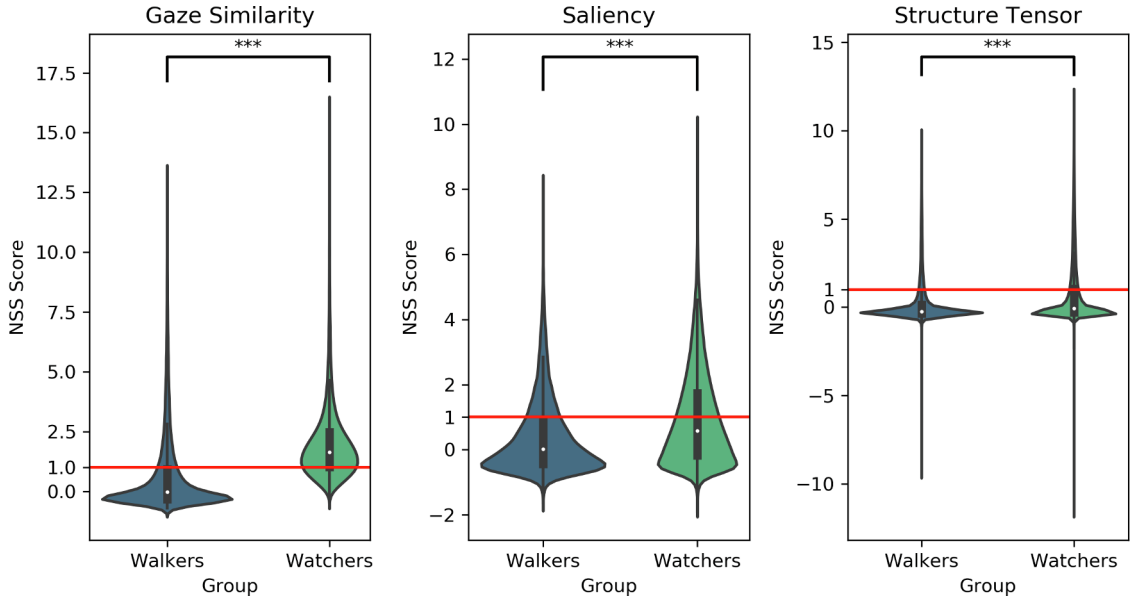

Fig. 3. Violin plots of the Watchers and the Walkers NSS scores computed on either (from left to right) gaze maps using a leave-one-out procedure on the watchers, the saliency maps or the motion maps. The red line indicates significance threshold for NSS Scores. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Watchers to land on the object of interest. Finally, the resulting differences were statistically evaluated by performing corrected t-tests against zero.

3. Results

3.1. 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 Fig. 3).

3.2. 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 Fig. 4).

3.3. 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 Fig. 4).

3.4. 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 Fig. 5, but see also Fig. 7 and section 3.6.).

The eye movement direction distribution indicated a greater central bias for Watchers, even when both distributions were fitted towards a

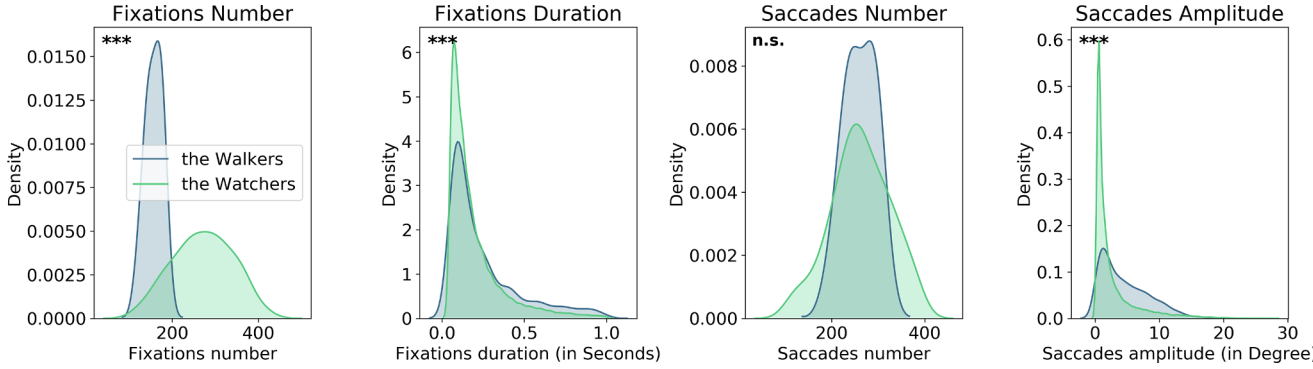

Fig. 4. Kernel density of fixation number and duration as well as saccade number, amplitude and duration for the Walkers (in blue) and the Watchers (in green). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

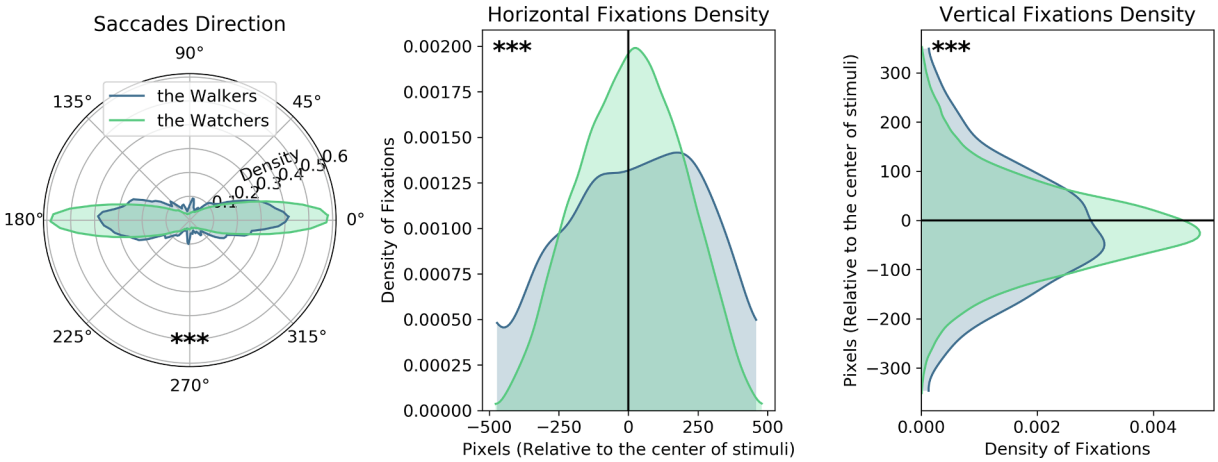

Fig. 5. Kernel density of saccades direction and Eye movements distribution for the vertical and horizontal direction for both the Walkers and the Watchers.

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 Fig. 5).

3.5. 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 Fig. 3). 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 Fig. 3).

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 Fig. 6). 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.

3.6. 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

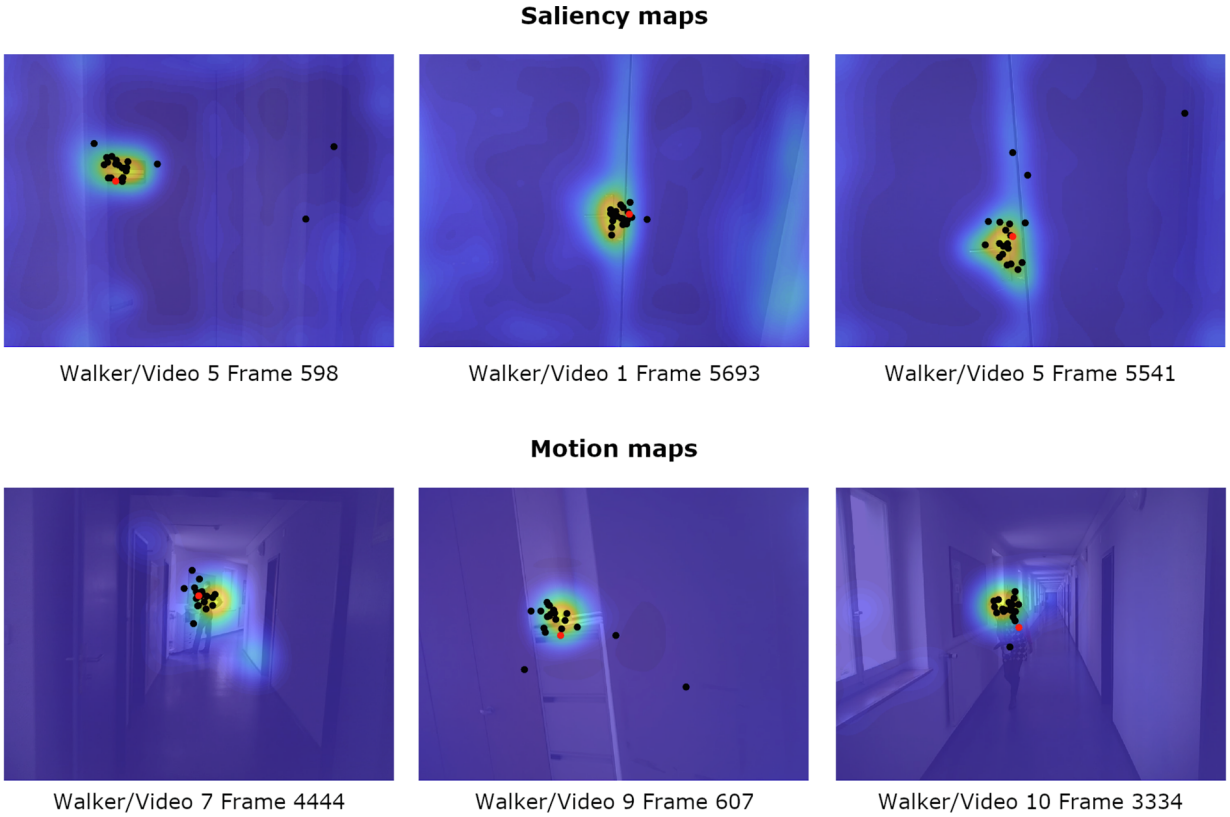

Fig. 6. Samples of frames with the highest NSS Score computed on saliency maps and motion maps for both the Walkers (red dot) and the Watchers (black dots). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

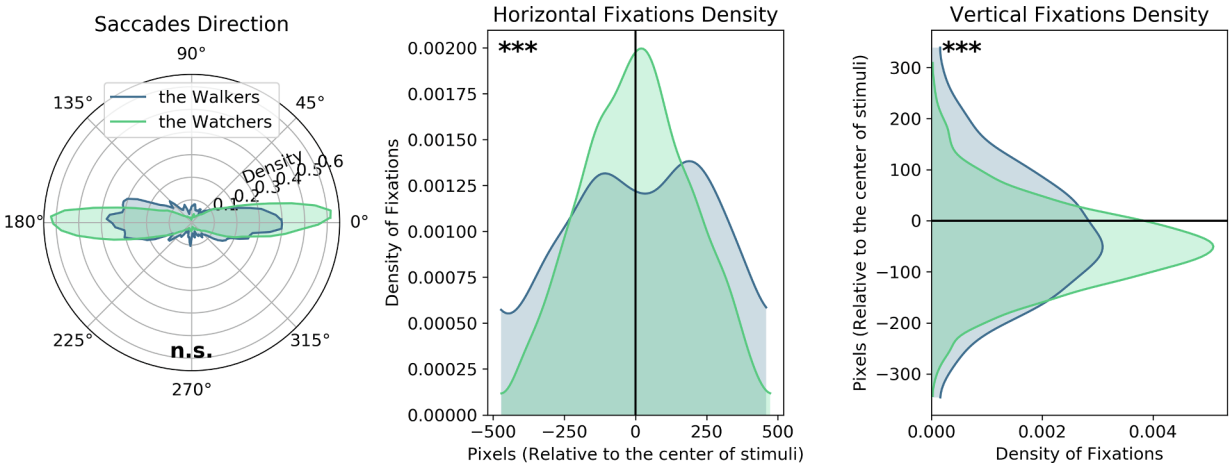

Fig. 7. Kernel density of saccades direction and Eye movements distribution for the vertical and horizontal direction for the first video watched by the Watchers.

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 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 Fig. 7).

4. 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 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 A). We also found robust differences in eye movements characteristics between the Walkers and the Watchers. On the contrary to Foulsham 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, & Stark, 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 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, Marple-Horvat, Henkes and Rowan (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 pre-processed 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.,* 't Hart et al., 2009; Foulsham et al., 2011). 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 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, Brockmole, Castelhano, and Mack (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., 2000). As such, the Watchers have more time and, resources to allocate their attention toward 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 show 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 (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, Miellet, Pernet, Sokhn, & Caldara, 2017) with a VR component *iMap4D* (Ticcinelli, de Lissa, Lalanne, Miellet, & Caldara, 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. 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 toward 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 show 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.

CRediT authorship contribution statement

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

Acknowledgements

This work was supported by the Swiss National Science Foundation grants (IZLJZ1_171065/1 and 316030_144998) awarded to R.C. The authors thank Víctor Leborán Álvarez for providing us with the MatLab code of the AWS-D saliency used in the present study.

Appendix A

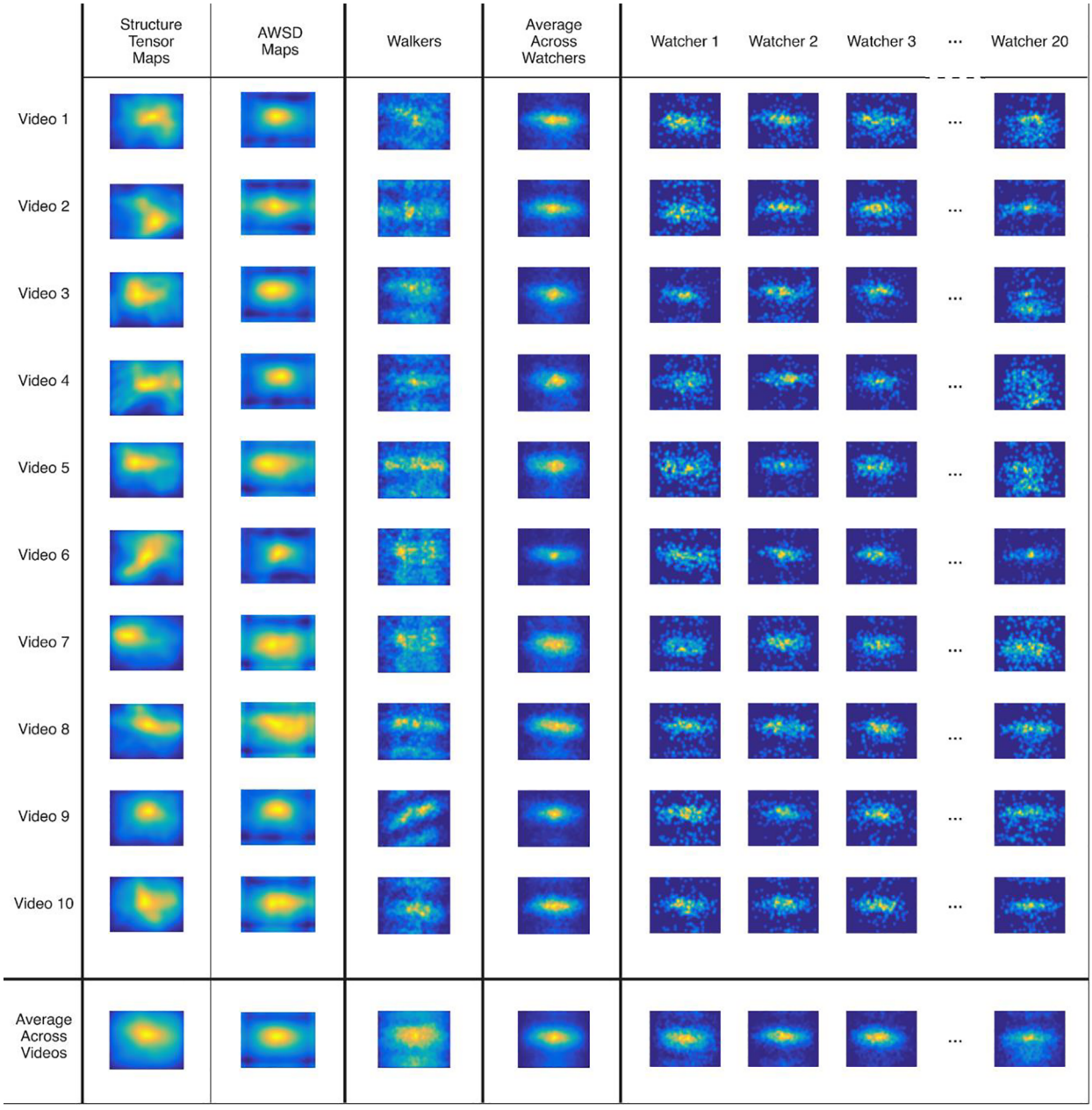

Fig. A. Saliency maps, motion maps and eye movements heat maps for all videos.

Appendix B

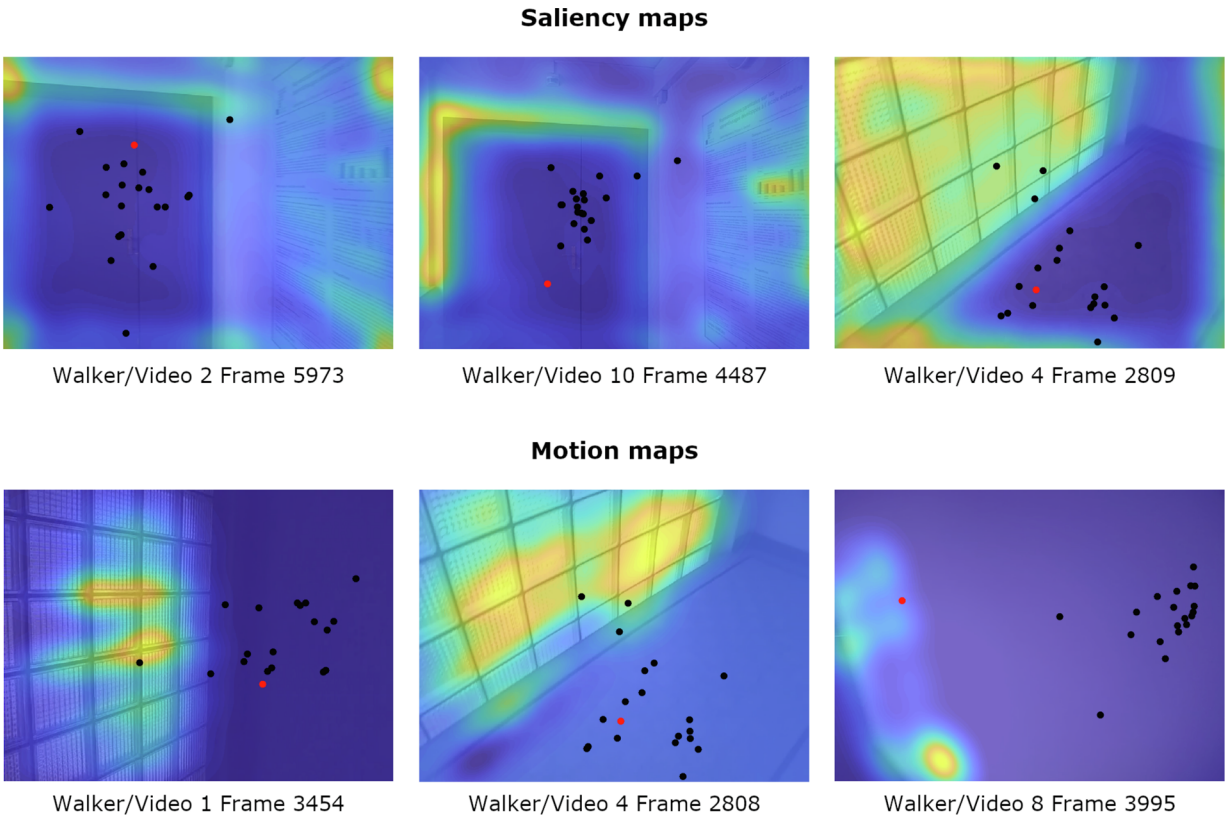

Fig. B. Samples of frames with the lowest NSS Score computed on saliency maps and motion maps for both the Walkers (red dots) and the Watchers (black dots).

References

'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.

Bahill, A. T., Adler, D., & Stark, L. (1975). Most naturally occurring human saccades have magnitudes of 15 degrees or less. Retrieved from *Investigative Ophthalmology, 14*(June), 468–469. http://www.ncbi.nlm.nih.gov/pubmed/1132942.

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 of perception in art.* University of Chicago Press.

Bylinskii, Z., Judd, T., Oliva, A., Torralba, A., & Durand, F. (2016). What do different evaluation metrics tell us about saliency models? ArXiv, 1–23. Retrieved from 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. 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. 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. X. R., Pardo, X. M. 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), R663–R664. 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. In Eye Movements (pp. 537–562). Elsevier. 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. Retrieved from *Vision Research, 40*(10–12), 1489–1506. http://www.ncbi.nlm.nih.gov/pubmed/10788654.
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. In Ralph Radach Jukka Hyona Heiner Deubel (Ed.), The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research (pp. 531–553). https://doi.org/10.1016/B978-044451020-4/50031-1.
Judd, C. H. (1905). 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. *Mit-Csail-Tr-2012, 1*, 1–7.
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 ECVP Abstract Supplement, 36(ECVP Abstract Supplement). https://doi.org/10.1068/v070821.
Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. Retrieved from *Human Neurobiology, 4*(4), 219–227. http://www.ncbi.nlm.nih.gov/pubmed/3836989.
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 and Biobehavioral Reviews*. 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.
Malcolm, 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. 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.e5. 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. Retrieved from *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.* New York, NY: Springer141–175.
Riche, N., & Mancas, M. (2016b). *Bottom-up saliency models for videos: A practical review.* New York, NY: Springer177–190.
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. https://doi.org/10.1523/JNEUROSCI.2968-18.2019.
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-movements and the aesthetics of visual form. Philos. Stud.
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.* St. Pete Beach, Florida: In VSS.
Vig, E., Dorr, M., & Barth, E. (2009). Efficient visual coding and the predictability of eye movements on natural movies (Vol. 22). Retrieved from https://pdfs.semanticscholar.org/dfaa/4accae796a941852f89af964993c104656b3.pdf.
Voorhies, R. C., Elazary, L., & Itti, L. (2012). Neuromorphic Bayesian surprise for far-range event detection. In 2012 IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance (pp. 1–6). IEEE. https://doi.org/10.1109/AVSS.2012.49.
Yarbus, A. L. (1967). Eye movements and vision. New York (Plenum Pre). New York: Springer US. https://doi.org/10.1007/978-1-4899-5379-7.
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), 601–621.
Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review, 115*(4), 787–835. https://doi.org/10.1037/a0013118.