VISUAL COGNITION
https://doi.org/10.1080/13506285.2018.1561567

Routledge
Taylor & Francis Group

Waldo reveals cultural differences in return fixations

Patrick Lüthold[a]*, Junpeng Lao[a]*, Lingnan He[b], Xinyue Zhou[c] and Roberto Caldara[a]

[a]Eye and Brain Mapping Laboratory (iBMLab), Department of Psychology, University of Fribourg, Fribourg, Switzerland; [b]School of Communication and Design, Sun Yat-Sen University, Guangzhou, People's Republic of China; [c]School of Management, Zhejiang University, Zhejiang, People's Republic of China

ABSTRACT

Humans routinely perform visual search towards targets to adapt to the environment. These sequences of ballistic eye movements are shaped by a combination of top–down and bottom–up factors. Recent research documented that human observers display cultural-specific fixation patterns in a range of visual processing tasks. In particular, eye movement strategies extracting information from faces clearly differs between Western Caucasian (WC) and East Asian (EA) observers. However, whether such cultural differences are also present for visual scene processing remains debated. To this aim, we recorded the eye movements of WC and EA observers while they were solving visual search problems parametrically varying in difficulty: *Where's Waldo*. Both groups had a comparable familiarity with the Waldo books reaching a comparable level of accuracy in target detection. Both cultural groups also showed a comparable *temporal* effect on inhibition of return, with longer fixation durations when saccades were performed to a return location compared to other locations. Westerners, however, located Waldo faster than Easterners. Interestingly, this modulation of speed was likely related to differences occurring on the low-level mechanisms of *spatial* inhibition of return, with EA observers returning more often to previously visited locations than the WC observers. This suboptimal eye movement strategy in the Easterners might be engendered by their cultural perceptual bias consisting in a greater use of extra-foveal information. Overall, our data point towards the existence of a subtle, but significant difference in the processing of visual scenes across observers from different cultures during active visual search.

ARTICLE HISTORY
Received 27 February 2018
Accepted 12 December 2018

KEYWORDS
Visual search; eye movements; inhibition of return; culture

Visual search is a critical perceptual task routinely performed by humans to select objects and information of interest in the environment. Eye movements play a crucial role in achieving this visual challenge, by continuously selecting targets among distractors with a series of fixations and saccades (Liversedge & Findlay, 2000). Importantly, these scan paths are neither randomly distributed nor completely deterministic, which comlexifies their understanding. In fact, fixation patterns are highly stable *within* the same observer exploring a particular visual scene (Andrews & Coppola, 1999), but vary greatly *between* observers and as a function of diverse task constraints (Henderson, 2003).

In the past 50 years, many theories have been proposed to model and predict eye movement scan paths during visual scene processing. One of the most prominent approaches consists of quantifying the visual saliency of images by using a class of "bottom–up" computer vision models (Itti & Koch, 2001). For example, the seminal work by Itti, Koch, and Niebur (1998) extracted saliency maps based on low-level properties of the images (i.e., spatial frequency, edge density, and local contrast) to predict free-viewing eye movements. Nowadays, there are more than 50 saliency-inspired fixation prediction models, which also include the broader categories of visual attention models (Bylinskii et al., 2015; Kümmerer, Wallis, & Bethge, 2015). With the recent advance in deep learning with multi-layer neural networks, performance in modelling and predicting eye movement patterns on a benchmark dataset with saliency-inspired models continues to improve (e.g., Kümmerer, Wallis, & Bethge, 2016). For more details, see the MIT Saliency Benchmark website http://saliency.mit.edu. However, human observers display distinctive scan paths even on an identical image as a function of the task at

---

CONTACT Roberto Caldara roberto.caldara@unifr.ch Eye and Brain Mapping Laboratory (iBMLab), Department of Psychology, University of Fribourg, Faucigny 2, 1700 Fribourg, Switzerland

*Joint first authors.

Supplemental data for this article can be accessed at https://doi.org/10.1080/13506285.2018.1561567

hand, an effect already highlighted in the very early eye tracking studies (Buswell, 1935; Yarbus, 1967). Attention models have also been integrated as a set of predictors in those models (e.g., Torralba, Oliva, Castelhano, & Henderson, 2003) to account for this "top-down" information (i.e., task effects, scene-schema knowledge, and other factors). Importantly, even with the consideration of both low-level bottom–up visual information and high-level top-down information, predicting individual eye movement scan paths remains a scientific challenge (e.g., Greene, Liu, & Wolfe, 2012; Haji-Abolhassani & Clark, 2014), due to the variability present across observers.

A compelling example of human variability in eye movements is face processing. While early studies (Yarbus, 1967) suggested the presence of a common triangular scan path during the viewing of faces, increasing evidence has highlighted the existence of robust idiosyncratic scan paths for faces (Chuk, Chan, & Hsiao, 2014; Mehoudar, Arizpe, Baker, & Yovel, 2014). Kanan, Bseiso, Ray, Hsiao, and Cottrell (2015) showed that observers do not exhibit universal scan path patterns while performing a series of face processing tasks. These observations also resonate with the perceptual cultural differences reported over the last decade. Cross-cultural studies comparing East Asians (EA) and Western Caucasians (WC) observers have highlighted distinct eye movement strategies during face identification (Blais, Jack, Scheepers, Fiset, & Caldara, 2008; Caldara, 2015; Calder & Young, 2005; Kelly et al., 2011; Kelly, Miellet, & Caldara, 2010; Miellet, He, Zhou, Lao, & Caldara, 2012; Rodger, Kelly, Blais, & Caldara, 2010) and the decoding of facial expressions of emotion (Geangu et al., 2016; Jack, Blais, Scheepers, Schyns, & Caldara, 2009). Notably, Western observers fixate more *local* face features (i.e., the eye and the mouth) during face recognition, whereas Eastern observers deploy more *global* fixation to the center of the face while obtaining a comparable level of accuracy. More importantly, these culturally distinctive scan paths are also related to a cultural tuning towards distinct spatial frequency information, as demonstrated by eye movement studies using a gaze-contingent technique combined with retinal filter models (Miellet, Vizioli, He, Zhou, & Caldara, 2013) and psychophysical experiments (Tardif et al., 2016). Western observers use more high spatial frequency information sampling, whereas Eastern observers rely more on low-spatial frequencies to process faces (for a review see, Caldara, 2017).

While such clear cultural differences in visual sampling strategies across observers from different cultures during face processing are well-established, it still remains debated whether such modulations extend to scene perception. Early studies have reported a similar local/global fixation bias between WCs and EAs during the viewing of scenes (Chua, Boland, & Nisbett, 2005; Goh, Tan, & Park, 2009; Masuda, Akase, Radford, & Wang, 2008). For example, Chua et al. (2005) showed that Americans spent a greater proportion of viewing time on focal objects relative to the background than Chinese participants, whereas Chinese observers made more fixations towards the background than Americans. Similarly, Goh et al. (2009) also found that WC observers fixated more on the focal object in a picture, whereas EAs constantly shifted their gaze between the focal object and the background. However, other studies reported little or no difference between the two cultures (Evans, Rotello, Li, & Rayner, 2009; Miellet, Zhou, He, Rodger, & Caldara, 2010; Rayner, Castelhano, & Yang, 2009; Rayner, Li, Williams, Cave, & Well, 2007). For example, a replication of the study by Chua et al. (2005) did not yield to any difference between the two cultural groups (Evans et al., 2009). Miellet et al. (2010) also did not find any difference in eye movement strategy between WCs and EAs in a visual search task involving the search of an animal in natural scenes with a gaze-contingent technique. Using centered parametric *Blindspots* of various sizes, they showed that WCs and EAs are equally impacted by the (lack of) foveal information in the search task, thus questioning the hypothesis that EAs rely more on peripheral information during (active) visual search.

A possible explanation for such inconsistencies might arise from the extremely rich visual information contained in natural scenes which could undermine the possibility of revealing subtle scan path differences across WC and EA observers. In fact, when less complex stimuli are used during visual search (Cramer, Dusko, & Rensink, 2016; Petrova, Wentura, & Fu, 2013) or visual categorization (Boduroglu, Shah, & Nisbett, 2009; Lao, Vizioli, & Caldara, 2013) cultural differences do emerge. Cramer et al. (2016) showed that Easterners raised in an EA environment have no significant asymmetry in a visual search task involving the search of long versus short lines, whereas EA immigrants present this asymmetry. In addition, Boduroglu et al. (2009) showed EA observers were better than WCs in a colour detection task with simple geometry shapes that required efficient parafoveal vision. Similarly, Petrova et al. (2013) found that task-irrelevant distractors had a greater influence on EAs than WCs during a simple saccade task. Nevertheless, whether such cultural perceptual differences in visual search would extend to more complex visual scenes remains to be clarified.

Importantly, previous studies comparing eye movement scanpaths of WCs and EAs observers using high-level stimuli during visual search have overlooked spatial saccade selection. Human observers certainly do not select the next fixation location randomly (e.g., Friston, Adams, Perrinet, & Breakspear, 2012) and given the efficiency and preferences in processing visual information, the gaze foraging mechanisms might also be different across cultures. If the cultural tuning towards different spatial frequency information between the two cultures holds also in scene perception, we should expect to detect these (subtle) differences in the *spatial* distribution of the foraging saccade. One of the possible markers of spatio-temporal sequential scanning strategies is *inhibition of return* (IOR, Posner & Cohen, 1984). IOR relates to the difficulty of reaction (delay in response) to a previously attended location. This effect was first observed in the Posner paradigm (Posner, 1980; Posner & Cohen, 1984), an experimental procedure where participants, after viewing a probable cue (valid, invalid), have to respond to a location next to the central fixation-cross that had been previously fixated. The effect of IOR has been extensively investigated in experimental psychology (for a review, see Klein, 2001) with the hypothesis that IOR helps the attentional system to reorganize information and plan forward movements during scene inspection. This bias in saccade direction is also shown in more general visual search tasks (Klein & MacInnes, 1999). One of these tasks was inspired by the famous children's book *Where's Waldo* (Handford, 1987), where participants were asked to fixate a suddenly appearing probe during the search of Waldo. The probe was presented either at the previously located position or at one of five possible positions deflected by 60° around a circle orbiting at the radius of the previous saccade length. Saccade latencies were slowest when the probe was on the previously fixated location, supporting the idea of a foraging facilitator effect as a result of IOR, which was highlighted by a forward bias in the *spatial* distribution of fixations. Smith and Henderson (2011) replicated Klein and MacInnes's Waldo study (1999) confirming elevated reflexive saccade latencies to probes at previously visited locations. However, they concluded that IOR was not responsible for the foraging facilitation. They instead referred to their findings of higher proportion of forward proceeding saccades, as a saccadic momentum (i.e., repetition of the saccadic program, rather than a consequence of IOR (Smith & Henderson, 2009). Regardless of the differences in the theoretical explanations of this phenomenon, the existence of a spatial bias in the selection of the foraging saccade remains uncontended. However, no study has directly compared the *spatial* distribution of saccade selection between WC and EA observers yet. Thus, it remains to be determined whether fine-grained measures of the *spatial* dynamics of eye movements during visual search are modulated by culture.

To this aim, we recorded the eye movements of WC and EA observers while they were solving one of the most famous visual search problems: *Where's Waldo*. *Where's Waldo* provides a unique active visual search task with parametrical difficulty and complexity. The vignettes contain rich high-spatial frequency information (edges and details), with an overall uniform background (relatively unbiased spatial layout with a unique clutter scene), representing an excellent proxy to probe the existence of perceptual cultural differences between the two cultures beyond simple geometric stimuli. All the observers shared a comparable level of visual experience with the famous book and no particular expertise in solving this visual problem. In addition, importantly, having previous experience with Waldo books hardly presents an advantage in the localization of Waldo (Olson, 2015). We selected 30 vignettes with a parametric level of difficulty (see Figure 1 for an example of the stimuli). Note that although the paradigm and analyses we used were heavily inspired by the IOR literature, we did not include an additional cuing task as in Smith and Henderson (2009). Instead, we mainly focus on the distribution mapping of *n*-back fixations, with *spatial* IOR as one of the potential factors modulated by culture, while also performing a temporal analysis on the fixation duration prior to return saccades compared to other saccades types (i.e., foregoing, over- and undershoot).

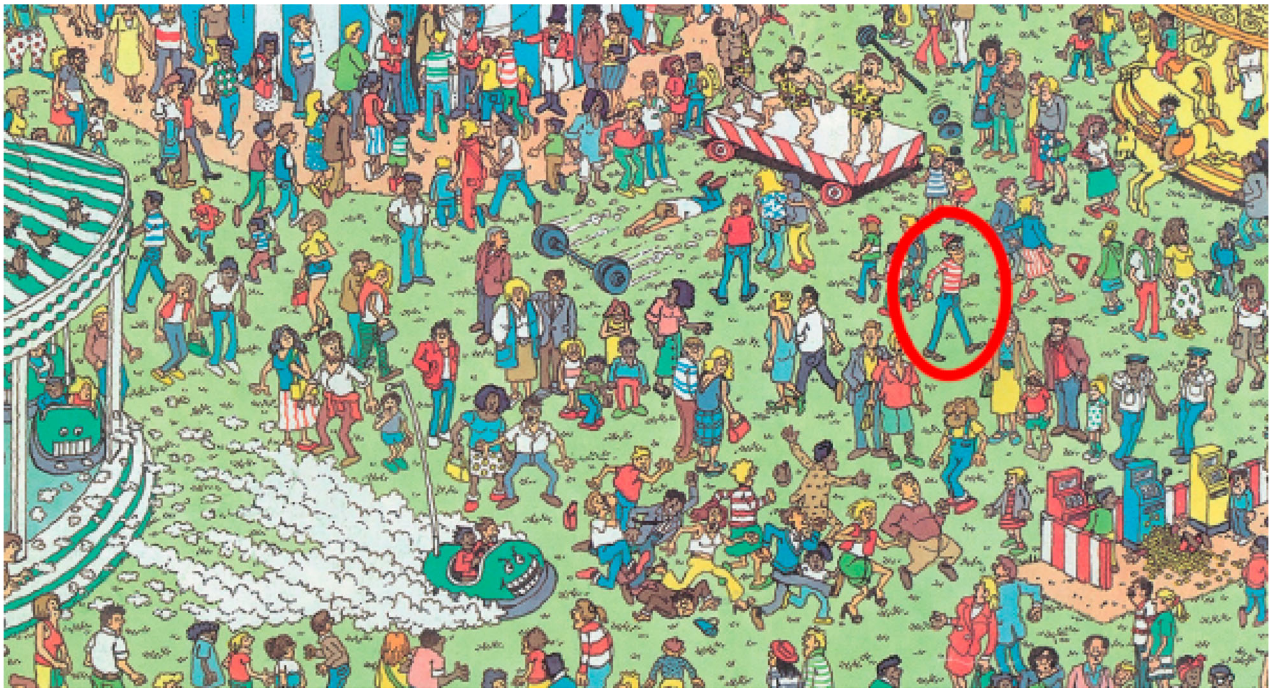

Figure 1. One of the *Where's Waldo* illustrations presented to the participants. Stimuli were scanned from the Solid Gold Collection (Handford, 2008). The red circle, which was not visible during the experiment, indicates a Waldo target.

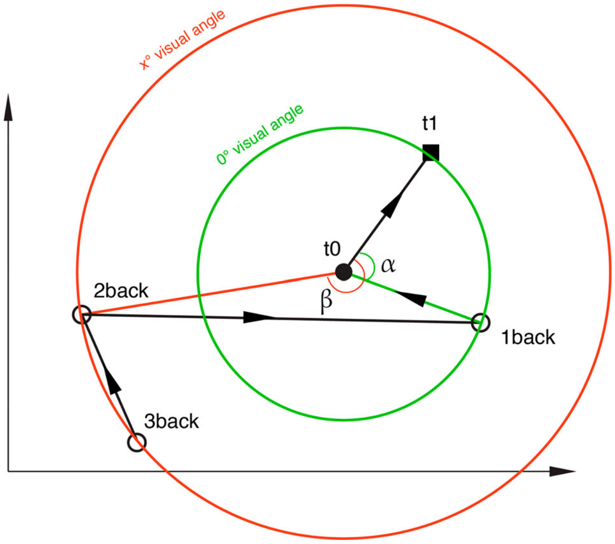

Figure 2. Illustration of the calculation of relative saccade orientation and amplitude (in reference to the 1-back or 2-back saccade). The black solid dot marks the penultimate fixation t_0 (i.e., the current fixation location, second last in the sequence); black open dots mark the previously visited locations in sequence depicted by the black arrows; the black solid square marks a potential future fixation location at t_1. The green circle marks a distance of 0° of visual angle; the exterior red circle marks a larger distance at x° of visual angle. To calculate the angle between the current fixation (t_1) and any previous fixation (1-back and 2-back), the penultimate fixation location (t_0) serves as the pivot. Angle α is calculated at t_0 between the current fixation (t_1) and the 1-back fixation. Angle β is calculated between the current fixation (t_1) and the 2-back fixation.

We analysed the characteristics of the scan paths by projecting the spatial fixation distributions in reference to the 1-back and 2-back fixations, by using the relative saccade orientation and amplitude, similar to Smith and Henderson (2011; see also Figure 2). Importantly, to investigate the fine-grained differences between WC and EA observers, we applied Kernel density estimations on the empirical joint likelihood of the relative saccade orientation and amplitude. We then used a robust data-driven approach to statistically isolate the differences between WC and EA observers (Lao, Miellet, Pernet, Sokhn, & Caldara, 2017). Our results show that EA observers are more likely to orientate their gaze to a previously fixated location compared to WC observers, while the temporal aspects of inhibition of return were not modulated by culture. This fixation pattern might relate to a reduced sensitivity towards foveal information in the EA observers. Such an eye movement sampling strategy might also impair their performance in this visual search task, as finding *Waldo* involves the processing of high-spatial filtering information.

Methods

The experiment material, raw data, and analysis scripts are openly available at https://osf.io/2q8jw/.

Participants

Twenty WCs (11 females, mean age = 24.70 (SD = 2.78)) and 20 EA (11 females, mean age = 22.25 (SD = 1.77)) from the University of Fribourg participated in the current study. All participants were right-handed with normal or corrected-to-normal vision. Participants were interviewed before and after the experiment to ensure they were familiar with the Waldo task and had no memory of the exact images presented during the experiment, or had a particular expertise with this task. All the EA participants were from Mainland China, had never been to a Western country before, and had not spent more than three months in Switzerland. The experiment was approved by the local ethics committee and participants provided written informed consent upon arrival to the experiment.

Eye tracking

Eye movements were recorded by means of a Desktop-Mount EyeLink 2K eye tracker (SR Research Ltd., Mississauga, Ontario, Canada) with a temporal resolution of 1000 Hz and a spatial resolution of 0.01° of visual angle. The average gaze position error was about 0.25°. We recorded observers' eye movements from the dominant eye monocularly via Matlab (R2006a), using the Psychophysics (PTB-3) and EyeLink Toolbox extensions (Brainard, 1997; Cornelissen, Peters, & Palmer, 2002; Pelli, 1997). Standard calibration and validation of the EyeLink were conducted before the experiment and repeated whenever necessary during the procedure.

Stimuli and procedure

Stimuli were 30 unique full-colour Where's Waldo[1] illustrations scanned from the Solid Gold Collection (Handford, 2008). Images were typical Waldo illustrations, which contained an overly crowded scene with very rich background/foreground elements (see Figure 1). In each of the images, there is one distinctively dressed Waldo character (red circle, Figure 1). The exact gesture and size of the target (Waldo) varied from scene to scene, yet never exceeded 1° of visual angle. The images were presented on a Dell P1130 19″ CRT monitor with a 1280 × 960 pixel white background at a refresh rate of 170 Hz. The size of each image was 1280 × 795 pixels, subtending 31.13° × 19.63° of visual angle at a viewing distance of 70 cm.

Participants were instructed to search for Waldo in each of the images. At the beginning of each trial, we used a central fixation cross as an automatic drift correction. If the participant's eye gaze was more than 1° of visual angle away from the fixation cross, an additional calibration followed by validation was conducted. After the fixation cross, a Waldo illustration was presented at random until response. The participant responded by pressing the space bar upon the detection of Waldo. The same image would appear again after a blank screen (100 ms), and the participants were instructed to indicate the location of the target using their gaze. The next trial began after the experimenter confirmed the target detection. In case of unsuccessful detection of the target at hand within a maximum duration of 2 min, participants could either skip the current stimuli completely, or elect to resume searching at a later stage. They could try up to three times on the vignettes they were struggling with. However, only the trials with successive identification on the first presentation were included in the eye movement analysis. The experimenter was in the same room with the subject throughout the experiment to closely monitor the eye drift and the accuracy of the subjects' detection. After the experiment, the participants verbally confirmed that they had never seen these particular Waldo images before.

Eye movement preprocessing

Eye movements were preprocessed using a custom MATLAB script applying the same default threshold parameters as in the EyeLink software (Miellet et al., 2010). Gaze velocity above 30°/s were determined as saccades. Fixations falling within a spatial range of <0.3° visual angle and a temporal range of <20 ms were merged. We extracted saccade information (onset, offset, amplitude and orientation) for each participant. We analysed the following oculomotor characteristic variables typically reported in eye movement literatures: number of fixations per second, average single fixation duration, and average saccade length.

One-back and 2-back fixation distribution

During the search for Waldo, all fixations were described by the means of their variation from the previous fixation. In other words, for each saccade we computed the angle and amplitude difference in relations to the 1-back and 2-back position of the eye (Smith & Henderson, 2011). As shown in Figure 2, saccade vector from t_0 to t_1 is compared with the vector of t_0 to 1-back (α) or 2-back (β) fixation. For example, if the eye fixation at t_1 perfectly overlaps with the 1-back fixation, the angle and the amplitude between the two vectors would be [0°, 0°]. For n fixations in a given trial from one observer, we computed n-1 saccades, which return in n-2 pairs of 1-back saccadic information (relative orientation and amplitude) and n-3 pairs of 2-back saccadic information.

Temporal inhibition of return

Temporal inhibition of return was investigated by analysing the fixation duration prior to return saccades compared to other saccades. We fitted a linear mixed effect model for fixation durations to test whether fixations prior to returns were longer than foregoing, over- or undershooting fixation from the *n*-back location. Thus, for each participant and each trial, we computed the average fixation duration for the preceding: return saccades, foregoing saccades, and saccades that over- or undershoot return locations. Return saccades were defined as those with a saccade vector formed by an angle within 5° with the *n*-back saccade and a distance within 1° of visual angle; over- or undershoot saccades are those with a distance larger than 1°. Similarly, foregoing saccades are defined as those with saccade vector formed an angle between [175° 185°] with the *n*-back saccade. The definition of return saccades used here is *ad-hoc*, but we tested a range of other values (0.5°–2°) and obtained similar results.

Statistics on 1-back and 2-back fixation distribution

The statistical analysis on the relative fixation distribution was performed on the joint empirical likelihood function of the *n*-back saccade orientation and amplitude. We first projected the orientation and amplitude information into a two-dimensional space independently for each participant at a single-trial level. We applied kernel density estimation for the joint likelihood of relative fixation orientation and amplitude. The bandwidth of the kernel is determined automatically using Scott's Rule (Scott, 1992). We then performed statistical mapping on the Gaussian smoothed 2D histogram. Moreover, in order to avoid bias in the spatial mapping and smoothing, the angular–amplitude representation was modeled on a warp two-dimensional space as shown in Figure 3(a). Intuitively, this is equivalent of applying a convolution on the surface of an open cylinder, with the smoothing kernel being a wrapped normal distribution on the azimuthal angle and a normal distribution on the polar angle. Individual angular–amplitude map was calculated by taking the 5% trimmean across trials. To assess the statistical differences between WC and EA, we applied a data-driven approach based on the Linear Mixed Model and bootstrap clustering implemented in *i*Map4 (Lao et al., 2017). This is equivalent to a 2D kernel regression. We applied a random intercept model with the subject as random effect and the culture as fixed effect. The result is then presented in a radial histogram similar to Figure 3(b). The colours in Figure 3 depict the frequency of relative saccade angle and amplitude. The colour blue represents low frequencies; the colour yellow represents high frequencies. Return to a previously fixated location has a value of [0°, 0°] (heat maps are inspired by the visualization used by Hooge, Over, van Wezel, & Frens, 2005; Motter & Belky, 1998).

Hypothesis testing (i.e., the linear contrast of the coefficient maps between WC and EA observers) is performed using a bootstrap spatial clustering method in *i*Map4 (Lao et al., 2017). The original statistical values (F-values) were thresholded at $p < .05$. *i*Map4 computes the size of the clusters and later compares them with a bootstrap distribution obtained under the null hypothesis (H_0). To construct the bootstrap null distribution, we removed the mean from each group. This procedure ensures the null hypothesis, in which no difference between WCs and EAs, is true. *i*Map4 then bootstraps the subject with replacement within each group to create 1000 null response matrices, and performs the same linear mixed model and contrasts (WCs – EAs). For each bootstrap, *i*Map4 computes the maximum cluster size of the *F*-value map at $p < .05$ and saves it in a vector. The bootstrapped cluster distribution under H_0 is the sorted vector resulting from each hypothesis testing. Significant clusters are defined as the one from the original contrast with size larger than the 95th bootstrapped

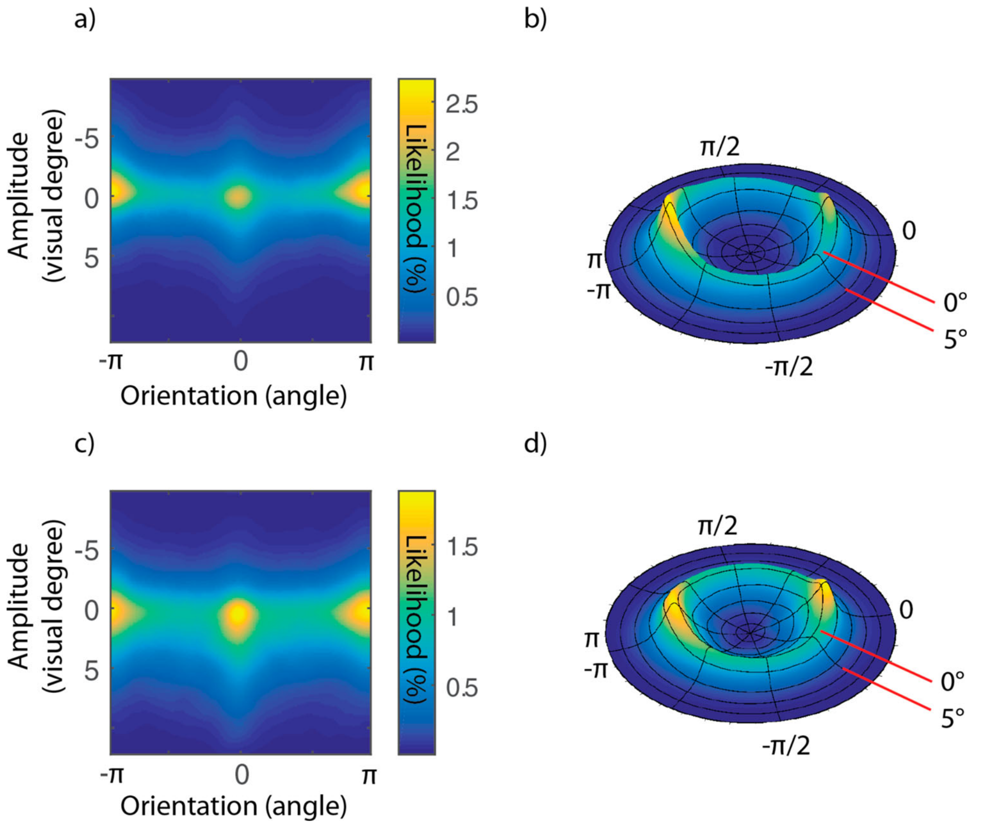

Figure 3. 2D representation of the joint likelihood function of the relative saccade angular and amplitude for 1-back (a, b) and 2-back (c, d) saccades across both groups of observers. In (a) and (c), saccade information is presented in Cartesian coordinate system with the relative orientation on the *x*-axis and amplitude in degree of visual angle on the *y*-axis. Importantly, the yellow area in the middle [0°, 0°] shows *n*-back saccades that deviated at an angle of 0° and an amplitude difference of 0° (a return to the *n*-back fixated location). (b) and (d) are the polar coordinate representation of (a) and (c). Colour map indicates joint likelihood of orientation and amplitude in percentage.

cluster size in the null distribution (for the details of the algorithm, see Lao et al., 2017).

Results

Search performance

Before investigating the presence and impact of oculomotor IOR on eye movements, we assessed the participant's performance in the *Where's Waldo* search task. Statistical analysis was performed using the generalized linear mixed model in MATLAB with subject and stimulus as random intercepts to account for repeated measures (Baayen, Davidson, & Bates, 2008). As expected, participants had difficulties in finding the target, but showed a good performance on average: success rate of 93.67%. In most cases, Waldo was located at the first attempt (86.92%). There was no significant difference on the success rate of finding Waldo at the first attempt between WC (87.93% [84.07, 90.96], square bracket shows 95% confidence intervals) and EA (86.52% [82.43, 89.77], $F(1,38) = 0.326$, $p = .571$). The average search duration for locating Waldo at the first attempt was 42.53s. The WC observers ($M = 37.40$s [21.72, 53.08]) were significantly faster to locate the target than EAs ($M = 59.81$s [44.10, 75.52], $F(1,1041) = 12.94$, $p = .00034$).

Eye movement results

Only the trials in which participant gave correct answers at the first attempt were analysed for eye movements. Overall, both groups of observers fully explored the whole image and there is no apparent bias in the area searched (supplementary figure 1). We did not find any significant difference between the two groups of observers for eye movement descriptive indices (see supplementary figure 2). Both groups showed a comparable number of fixations per second ($M_{WC} = 3.57$ [3.37, 3.77]; $M_{EA} = 3.46$ [3.26, 3.66];

$F(1,1041) = 0.55$, $p = .458$), similar mean fixation duration ($M_{WC} = 244.92$ ms [233.58, 256.26]; $M_{EA} = 245.20$ ms [233.85, 256.55]; $F(1,1041) = 0.001$, $p = .970$), and comparable mean saccade amplitudes ($M_{WC} = 3.36°$ of visual angle [3.06, 3.66]; $M_{EA} = 3.05°$ [2.75, 3.35]; $F(1,1041) = 2.28$, $p = .131$).

To explore the effect of image difficulty on oculomotor characteristics, we separated the Waldo stimuli into two levels of difficulty (*easy* or *hard*) independently for each participant. We first sorted the search durations of all the stimuli within one single observer, and then labeled the first 15 stimuli as *easy* and the rest as *hard*. Eye movement characteristics were then modeled as functions of the image difficulty (*easy* or *hard*), culture of the observer (WC or EA), and their interaction. The number of fixations per second and the mean fixation duration were not significantly modulated by the image difficulty, as neither the main effect nor its interaction with the group predictor was significant (F's < 0.647, $p > .05$). However, both the main effect of the task difficulty ($F(1, 1039) = 49.63$, $p = 3.3699\text{e}{-12}$) and the image difficulty x group interaction ($F(1, 1039) = 12.09$, $p = 5.2858\text{e}{-4}$) were significant for mean saccade amplitude. *Post hoc* comparisons showed that the significant effects were driven by the EA observers, who made significantly smaller saccades during the viewing of hard ($M = 2.76°$ [2.456, 3.072]) than easy ($M = 3.27°$ [2.971, 3.578]) stimuli, whereas the WCs displayed similar mean saccade amplitudes during the viewing of both easy ($M = 3.45°$ [3.149, 3.756]) and hard ($M = 3.26°$ [2.948, 3.562]) stimuli (see supplementary figure 3). It is worth noting that our result does not change even if we treat the image difficulty as a continuous predictor: using the z-scored search duration for each image as predictor we found the same significance only in the mean saccade amplitude (significant main effect of search duration/image difficulty: $F(1, 1039) = 79.702$, $p = 1.934\text{e}{-18}$, and significant interaction: $F(1, 1039) = 7.00$, $p = .00827$).

Analysis on the 1-back and 2-back fixation distribution

The *n*-back fixation distribution is represented as a smoothed 2D map of the relative saccade orientation and amplitude for both groups of observers (Figure 3). As shown in Figure 3, the forgoing saccades were not distributed uniformly. Two local maxima are presented in the angular-amplitude map (Figure 3(b) and (d)). This result indicates that the subsequent saccade performed by human observers was most likely to be either exactly the same as the previous saccade, or directed towards the N-back fixation location. More importantly, in general, the local maximum located around [0°, 0°] shows that there was no strong evidence of spatial inhibition of return, as overall all the observers showed a high likelihood to return to the previous fixated location.

We then performed statistical mapping using *i*Map4 to compare the cultural difference of the 2D angular-amplitude map between WC and EA (Lao et al., 2017). The result is shown in Figure 4. For both 1-back and 2-back saccades, significant differences are revealed after non-parametric bootstrap clustering test: EA showed higher saccade likelihood on the [0°, 0°] compare to WC observers, indicating a directional effect towards previously fixated locations for EA observers (1-back: maximum F value: $F(1, 38) = 30.23$, $p = .00003$; 2-back: maximum F value: $F(1, 38) = 29.35$, $p = .000004$).

To further quantify the cultural fixation return bias and explore its relationship with individual visual search performance, we conducted a *post hoc* analysis within the significant cluster. Using the sum likelihood value within the significant region (i.e., around [0°, 0°]), we estimated that the EA observers were about 4% more likely to saccade to a previously visited fixation location than WC observers (1-back: EA – 19.27% [18.23, 20.31], WC – 15.53% [14.49, 16.56]; 2-back: EA – 19.56% [18.43, 30.68], WC – 15.02% [13.90, 16.14]). Importantly, this returning strategy was positively related to the search performance (i.e., how long it took for each participant to find Waldo on average, see Figure 5). The longer it took for an individual to find Waldo, the more likely for the observer to perform a return saccade (1-back: regression coefficient *beta* = 0.0915 [0.0570, 0.1259], $t(38) = 5.38$, $p = 4.079\text{e}{-6}$; 2-back: *beta* = 0.1071 [0.0722, 0.1419], $t(38) = 6.22$, $p = 2.849\text{e}{-7}$).

An additional *post-hoc* analysis was performed to assess whether the search strategy of EA observers relates to suboptimal foveal sampling. For both groups, we compared the number of the proportion of trials that were terminated after the first fixation to the target. The fixation to the target is defined as within the 1° visual angle radius, but the resulting pattern is the same within 0.5° to 2° of visual angle. We then fitted a mixed effect model with subject

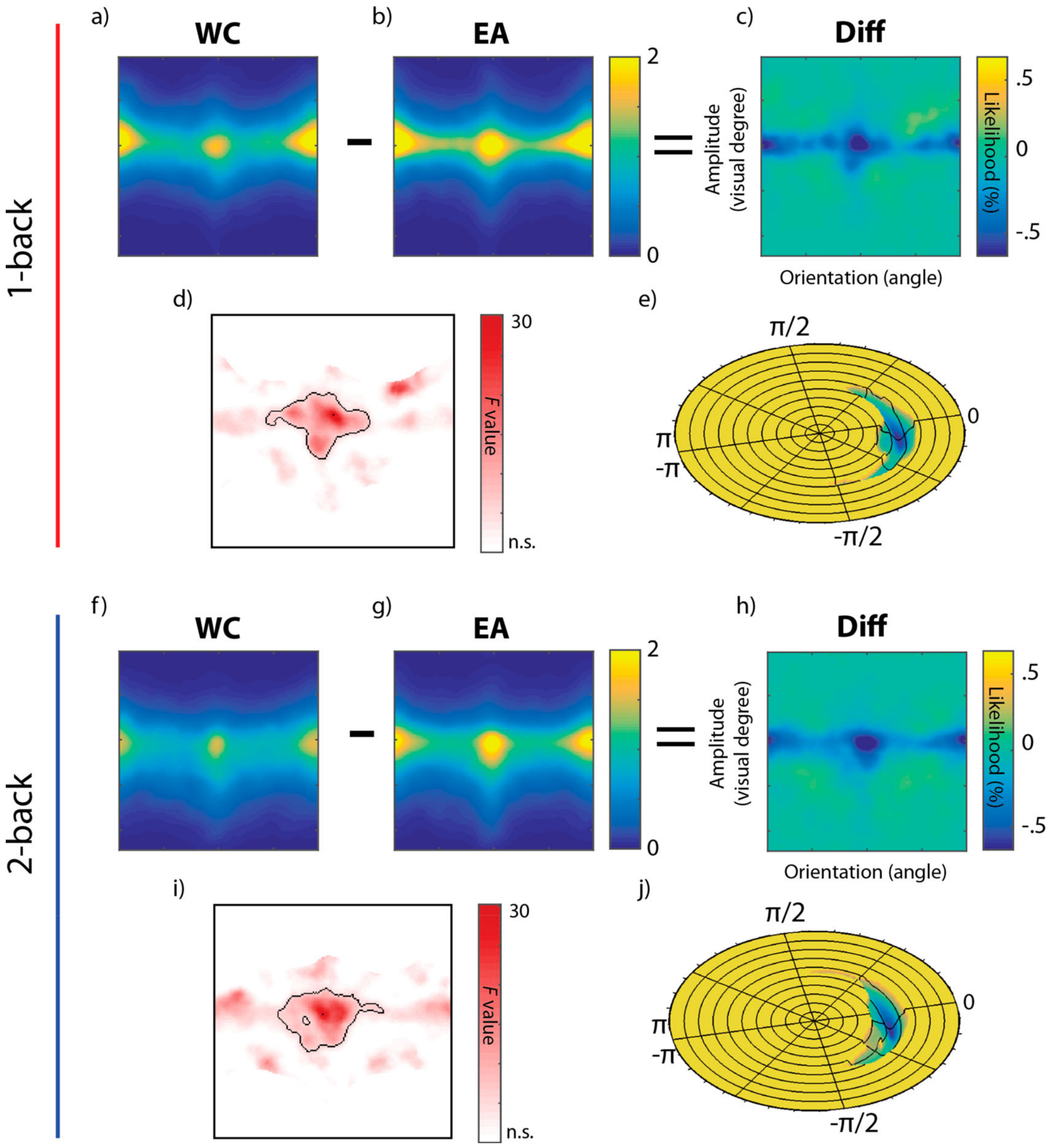

Figure 4. Statistical analyses on the differences between WC and EA on the angular-amplitude joint likelihood map. (a) Angular-amplitude map for WC in 1-back condition. (b) Angular-amplitude map for EA in 1-back condition. (c) Angular-amplitude differences between WC and EA in 1-back condition (WC minus EA). (d) Statistical value map (*F*-value) for the difference showed in (c), black line indicates the significant area. (e) The polar representation of the significant different area. (f–j) are the result for the 2-back condition.

and stimuli as random group effect. This analysis revealed a significant difference between WC and EA observers: the proportion of trials terminated after the first fixation to the target is lower in EA than WC observers (EA – 47.36 [42.59, 52.18], WC – 66.05% [53.10, 76.98], $t(1039) = 4.35$, $p = 1.5208\text{e-}05$).

Temporal inhibition of return

While our spatial analysis showed that both groups of observers performed return saccades with a high likelihood, a temporal effect on IOR could still be present if fixation durations prior to a return saccade are longer than saccades performed for foregoing, over- or undershoot saccades from prior fixation locations compared to saccades to the return locations. We thus performed an additional temporal analysis on fixation duration on the saccades performed from the *n*-back location in both groups. The linear mixed effect model showed a significant main effect of Saccade type on fixation duration ($F(2, 2170) = 6.922$, $p = .0010$), whereas neither the main

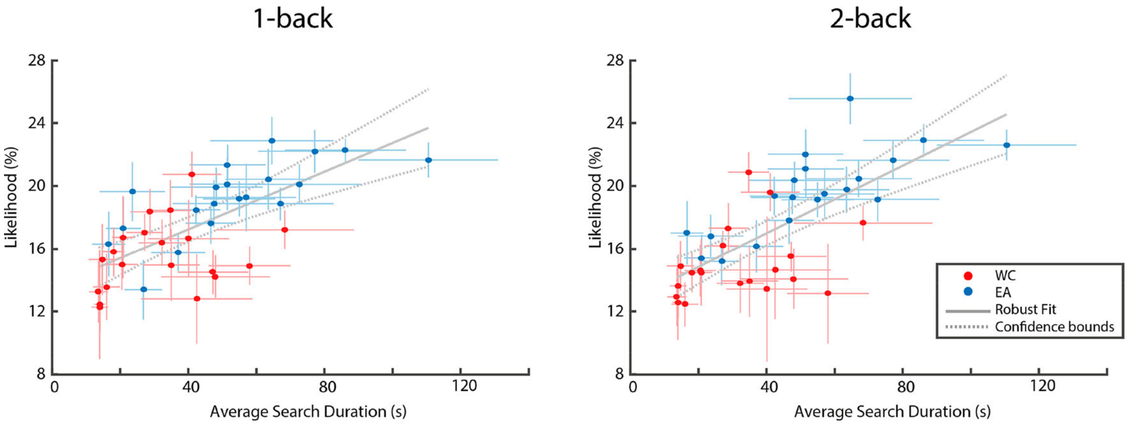

Figure 5. *Post-hoc* analysis within the significant cluster. The scatter plots show the likelihood to orientate the gaze towards previously fixated locations (sum likelihood within the significant cluster, *y*-axis) as a function of individual search performance (average search duration, *x*-axis). A robust linear regression was performed in Matlab using *fitlm* with a Cauchy weight function. Red dot: WC observers; Blue dot: EA observers. Error bar shows 95% confidence interval of the mean.

effect of Culture or the Culture X Saccade type interaction was significant. The temporal IOR as the fixation duration for return saccades (241.83 ms [233.50, 249.21]) was significantly larger than foregoing, over- or undershoot saccades (227.83 ms [223.07, 232.49]), but this effect was not modulated by the culture of the observer.

Discussion

The present study investigated whether culture modulates the oculomotor strategies during the visual search of *Waldo*. We modeled the fixation distribution by quantifying the stochastic differential of the gaze temporal dynamics (1-back and 2-back fixation distributions). We had a particular interest in mapping potential cultural differences in the *spatial* distribution of saccade selection, by considering *spatial* inhibition of return (IOR – an inhibitory process that facilitates eye gazes to novel locations in the visual field) as one of the potential markers for such effect. WC and EA observers found *Waldo* with similar accuracy and displayed similar eye movement properties (i.e., fixation duration, saccade length) and distributions. We found support of *temporal* IOR for both groups of observers. Regardless of the culture of the observer, fixation durations were longer when preceding return saccades compared to foregoing, over- and undershoot saccades from *n*-back locations. However, neither groups of observers showed *spatial* IOR, as revelead by a data-driven non-parametric analysis: return fixations occurred very frequently during the search of *Waldo* equally for both groups. However, the novelty of the current observation is that EA observers were even more likely to gaze towards a previously fixated location compared to WC observers. This “double-checking” visual search strategy also correlated with a decrease in performance, as the longer an observer spent on finding *Waldo*, the more return fixations they performed. Overall, our results highlight this very distinct and time-consuming visual search strategy in EA observers, a strategy that might be related to a decreased sensitivity in processing foveal fine-grained information. Waldo images represent a crowded scene where the dissociation between foreground and background is parametrically hampered. The task of finding Waldo is then straightforward and neutral, unlike judging the appeal of a picture or reporting the content of an image, which could easily be affected by subjective affective decisions like in previous studies (Chua et al., 2005; Goh et al., 2009; Miellet et al., 2010). Importantly, both cultural groups had comparable familiarity with Waldo’s vignettes and the paradigm used here.

In line with Smith and Henderson (2011), our results did not show strong evidence for *spatial* IOR in either group of observers during the visual search of *Waldo*. Smith and Henderson (2011) argued that such complex visual stimuli necessitate observers to perform return saccades for a second inspection. In the literature of reading , the increase in reflexive saccades with increasing text difficulty is well documented (Rayner & Pollatsek, 1989; Vitu & Underwood, 2005). Interestingly, we found that EA observers returned more to previously foveated locations than for WC observers for both 1-back and 2-back fixations. Since EA observers have a more *global* distribution of attention relative to WC observers, they might be less efficient in the processing of foveal information (Miellet et al., 2013). This cultural difference in the scan path strategy might thus relate to the encoding of the high-spatial frequency details of the Waldo image and the location of the target. EA observers revisited the previous fixation location more often than WC observers, thus extending the search for Waldo. On the contrary, WC observers were better in encoding *local* information and in asserting the absence of the target in the fixated areas and propagate their eye gaze forward by then locating Waldo faster than EA observers. Such a lower rate of return fixations observed in Westerns engenders a better monitoring of the eye movement scan paths. As detailed before, evidence for a larger perceptual window in EA observers comes from previous studies that used different tasks and visual problems (Boduroglu et al., 2009; Miellet et al., 2013; Petrova et al., 2013). Thus, ultimately, a future study using a gaze-contingent design that parametrically manipulates the quantity of foveal information in this task (see, Miellet et al., 2013) is necessary to provide direct evidence for a larger attentional window in the EA observers.

Potential differences in spatial attention between WCs and EAs observers should have been probed by using a more conventional cueing task, as widely used in the IOR literature (Posner & Cohen, 1984). While the primary focus of the current study was to statistically map the spatial distribution of saccade selection between WC and EA observers, from our results we could reasonably predict a cultural difference in responding to cued *versus* uncued locations during spatial attention. Surprisingly, to the best of our knowledge, no study has directly compared participants from both cultures with cueing paradigms yet. Future studies are necessary to investigate this important issue.

It is worth noting that to quantify the *spatial* inhibition or facilitation of return, a baseline frequency must be properly estimated (Bays & Husain, 2012). In the current study, we focused on the cultural difference between WC and EA observers on saccade facilitation. The lack of spatial inhibition of return is indicated by the local maximum around the point [0°, 0°] in the smoothed spatial distribution of saccade. However, to better estimated the bias in saccade facilitation and its potential cultural effect, a baseline frequency should be introduced by either presenting the same participant with naturalistic stimulus in a free-viewing setting, or estimated by shuffling saccade within the same trials (Bays & Husain, 2012; Wilming, Harst, Schmidt, & König, 2013). Future studies are necessary to clarify this issue.

Inhibition of return is one of the mechanisms facilitating visual search in humans. It can be measured *spatially*, as the probability of the next search location conditioned by the previously fixated area; and/or *temporally*, as the delay response to targets presented at previously sampled locations (Smith & Henderson, 2009, 2011). Here, we tested the effect of temporal IOR using oculomotor kinematics, by measuring changes in fixation duration conditioned by subsequent saccades. Our results provided evidence for temporal IOR, with longer fixation durations for return saccade compared to other saccades (i.e., foregoing, over- and undershoot). However, ideally, an additional behavioral measure should have been used to quantify the efficiency of processing information in the return location (for instance, see Smith & Henderson, 2009, 2011). Future studies comprising an appropriate design including additional task demands would be necessary to further clarify the cultural effect of the temporal IOR.

Finally, the larger return fixation rate observed in the EA observers increased with image difficulty. EA performed shorter saccades when the individual specific image difficulty increased, which could be used as an indicator of the area inspected during a single fixation. However, it is worth noting that in line with previous studies (Evans et al., 2009; Rayner et al., 2007, 2009), we did not observe a significant cultural difference on the global eye-tracking measures. The mean fixation duration and the fixations per second in both groups are comparable and similar even when the difficulty level of the vignettes increased.

Conclusions and future work

During the last two decades, many cultural differences in diverse perceptual (Nisbett & Masuda, 2003; Nisbett & Miyamoto, 2005) and face processing tasks (e.g., Blais et al., 2008; for a review see Caldara, 2017) have been reported in the literature. EA observers sample and process visual information more with a *global* strategy, compared to a *local* processing style typical of WC observers. However, the results reported so far for tasks involving the processing of natural visual scenes have provided mixed evidence of an cultural effect (Chua et al., 2005; Goh et al., 2009; Masuda et al., 2008; Miellet et al, 2010). Here, we investigated whether finding *Waldo* with parametrical scene complexity would elicit culturally specific eye movement patterns. Our robust data-driven analyses have clearly shown that visual search strategies are not falling into *universal* rules, with EA observers being slower and having more return fixations compared to WC observers. Future studies are necessary to clarify whether and how the *global/local* cultural perceptual bias observed for face processing contributes to this effect. Ideally, the individual variability across trials should also be properly considered by using a hierarchical model directly modeling on the *spatial–temporal* saccade data (e.g., a Poisson point process model).

One prediction from the current result is that *local* eye movement strategy typical of Western observers could lead to a more efficient information encoding of the inspected area thus preventing a return fixation. Whereas, on the contrary, the *global* strategy, typical of Eastern observers compensates for a suboptimal foveal sampling, by reprocessing previously visited locations (Findlay & Brown, 2006). In the future, the use of a design with a gaze-contingent technique is necessary to verify this hypothesis, as visual information intake can be controlled parametrically with such experimental techniques. Regardless of these potential explanations, the present data feed the literature by describing culturally dependent cognitive and visual strategies, which emphasize the need to report the observers' cultural background as a crucial variable for the understanding of visual cognition.

Note

1. We followed the convention of previous visual search literature and refer to the task as "Where's Waldo?". Note the original book from which the stimuli are taken from is the British version "Where's Wally?".

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Schweizerischer Nationalfonds zur Förderung awarded to R. C. [Grant Number IZLJZ1_171065].

ORCID

Roberto Caldara http://orcid.org/0000-0002-8260-9949

References

Andrews, T. J., & Coppola, D. M. (1999). Idiosyncratic characteristics of saccadic eye movements when viewing different visual environments. *Vision Research*, *39*(17), 2947–2953.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

Bays, P. M., & Husain, M. (2012). Active inhibition and memory promote exploration and search of natural scenes. *Journal of Vision*, *12*(8), 8.

Blais, C., Jack, R. E., Scheepers, C., Fiset, D., & Caldara, R. (2008). Culture shapes how we look at faces. *Plos One*, *3*(8), e3022–e3029. doi:10.1371/journal.pone.0003022

Boduroglu, A., Shah, P., & Nisbett, R. E. (2009). Cultural differences in allocation of attention in visual information processing. *Journal of Cross-Cultural Psychology*, *40*(3), 349–360. doi:10.1177/0022022108331005

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436.

Buswell, G. T. (1935). How people look at pictures: a study of the psychology and perception in art.

Bylinskii, Z., DeGennaro, E. M., Rajalingham, R., Ruda, H., Zhang, J., & Tsotsos, J. K. (2015). Towards the quantitative evaluation of visual attention models. *Vision Research*, *116*(Part B), 258–268. doi:10.1016/j.visres.2015.04.007

Caldara, R. (2015). *Mapping the impact of culture and race in visual processing (pp. XVIII–XVIII)*. Presented at the 2015 7th International Conference on Knowledge and Smart Technology (KST), IEEE, Chonburi, Thailand. http://doi.org/10.1109/KST.2015.7051447

Caldara, R. (2017). Culture reveals a flexible system for face processing. *Current Directions in Psychological Science*, *26(3)*, 249–255.

Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience*, *6*(8), 641–651. doi:10.1038/nrn1724

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 of the United States of America*, *102*(35), 12629–12633. doi:10.1073/pnas.0506162102

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. doi:10.1167/14.11.8

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.

Cramer, E. S., Dusko, M. J., & Rensink, R. A. (2016). Group-level differences in visual search asymmetry. *Attention, Perception, & Psychophysics*, 1–18.

Evans, K., 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. doi:10.1080/17470210802373720

Findlay, J. M., & Brown, V. (2006). Eye scanning of multi-element displays: I. Scanpath planning. *Vision Research*, *46*(1–2), 179–195. doi:10.1016/j.visres.2005.06.010

Friston, K., Adams, R., Perrinet, L., & Breakspear, M. (2012). Perceptions as hypotheses: Saccades as experiments. *Frontiers in Psychology*, *3*(151), 1–20. doi:10.3389/fpsyg.2012.00151

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*, 663–664.

Goh, J. O., Tan, J. C., & Park, D. C. (2009). Culture modulates Eye-movements to visual novelty. *Plos One*, *4*(12), e8238–e8239. doi:10.1371/journal.pone.0008238

Greene, M. R., Liu, T., & Wolfe, J. M. (2012). Reconsidering Yarbus: A failure to predict observers' task from eye movement patterns. *Vision Research*, *62*(C), 1–8. doi:10.1016/j.visres.2012.03.019

Haji-Abolhassani, A., & Clark, J. J. (2014). An inverse Yarbus process: Predicting observers' task from eye movement patterns. *Vision Research*, *103*(C), 127–142. doi:10.1016/j.visres.2014.08.014

Handford, M. (1987). *Where's Waldo?* Boston: Little, Brown.

Handford, M. (2008). *Where's Wally? The solid gold collection*. London: Walker Books.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, *7*(11), 498–504. doi:10.1016/j.tics.2003.09.006

Hooge, I. T. C., Over, E. A. B., van Wezel, R. J. A., & Frens, M. A. (2005). Inhibition of return is not a foraging facilitator in saccadic search and free viewing. *Vision Research*, *45*(14), 1901–1908. doi:10.1016/j.visres.2005.01.030

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2*(3), 194–203. doi: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. doi:10.1109/34.730558

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*(18), 1543–1548. doi:10.1016/j.cub.2009.07.051

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*(C), 67–76. doi: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. doi:10.1111/j.1467-7687.2011.01067.x

Kelly, D. J., Miellet, S., & Caldara, R. (2010). Culture shapes eye movements for visually homogeneous objects. *Frontiers in Psychology*, *1*(6), 1–7. doi:10.3389/fpsyg.2010.00006

Klein, C. (2001). Developmental functions for saccadic eye movement parameters derived from pro-and antisaccade tasks. *Experimental Brain Research*, *139*(1), 1–17.

Klein, R. M., & MacInnes, W. J. (1999). Inhibition of return is a foraging facilitator in visual search. *Psychological Science*, *10*(4), 346–352.

Kümmerer, M., Wallis, T. S., & Bethge, M. (2016). DeepGaze II: Reading fixations from deep features trained on object recognition. *arXiv preprint arXiv:1610.01563*.

Kümmerer, M., Wallis, T. S. A., & Bethge, M. (2015). Information-theoretic model comparison unifies saliency metrics. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(52), 16054–16059. doi:10.1073/pnas.1510393112

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*, 559–575. doi:10.3758/s13428-016-0737-x

Lao, J., Vizioli, L., & Caldara, R. (2013). Culture modulates the temporal dynamics of global/local processing. *Culture and Brain*, *1*(2–4), 158–174. doi:10.1007/s40167-013-0012-2

Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences*, *4*(1), 6–14. doi:10.1016/S1364-6613(99)01418-7

Masuda, T., Akase, M., Radford, M. H. B., & Wang, H. (2008). Cross-cultural research on the pattern of eye-movement: Comparing the level of concentration between Japanese and Western participants. *Japanese Psychological Research*, *79*, 35–43.

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), 6. doi:10.1167/14.7.6

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), 1–12. doi:10.1691

Miellet, S., Vizioli, L., He, L., Zhou, X., & Caldara, R. (2013). Mapping face recognition information use across cultures. *Frontiers in Psychology*, *4*(34),

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. doi:10.1167/10.6.21

Motter, B. C., & Belky, E. J. (1998). The guidance of eye movements during active visual search. *Vision Research*, *38*(12), 1805–1815.

Nisbett, R. E., & Masuda, T. (2003). Culture and point of view. *Proceedings of the National Academy of Sciences of the*

United States of America, *100*(19), 11163–11170. doi:10.1073/pnas.1934527100

Nisbett, R. E., & Miyamoto, Y. (2005). The influence of culture: Holistic versus analytic perception. *Trends in Cognitive Sciences*, *9*(10), 467–473. doi:10.1016/j.tics.2005.08.004

Olson, R. (2015, February 3). *Here's Waldo: Computing the optimal search strategy for finding Waldo [Web log post]*. Retrieved from http://www.randalolson.com

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

Petrova, K., Wentura, D., & Fu, X. (2013). Cultural influences on oculomotor inhibition of remote distractors: Evidence from saccade trajectories. *Vision Research*, *84*, 43–49. doi:10.1016/j.visres.2013.03.008

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

Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. *Attention and Performance X: Control of Language Processes*, *32*, 531–556.

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), 254–259. doi:10.1037/a0013508

Rayner, K., Li, X., Williams, C. C., Cave, K. R., & Well, A. D. (2007). Eye movements during information processing tasks: Individual differences and cultural effects. *Vision Research*, *47*(21), 2714–2726. doi:10.1016/j.visres.2007.05.007

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice Hall.

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. doi:10.1068/p6750

Scott, D. W. (1992). *Multivariate density estimation: Theory, practice, and visualization*. New York, Chichester: John Wiley & Sons.

Smith, T. J., & Henderson, J. M. (2009). Facilitation of return during scene viewing. *Visual Cognition*, *17*(6-7), 1083–1108. doi:10.1080/13506280802678557

Smith, T. J., & Henderson, J. M. (2011). Looking back at Waldo: Oculomotor inhibition of return does not prevent return fixations. *Journal of Vision*, *11*(1), 3. doi:10.1167/11.1.3

Tardif, J., Fiset, D., Zhang, Y., Estéphan, A., Cai, Q., Luo, C. … Blais, C. (2016). Culture shapes spatial frequency tuning for face identification. *Journal of Experimental Psychology. Human Perception and Performance*, *43*(2), 294–306. doi:10.1037/xhp0000288

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2003). Top-down control of visual attention in object detection. *International Conference on Image Processing*, *1*, I–253–6 vol.1.

Vitu, F., & Underwood, G. (2005). Visual extraction processes and regressive saccades in reading. In G. Underwood (Ed.), *Cognitive Processes in Eye Guidance* (pp. 1–32). Oxford: Oxford University Press.

Wilming, N., Harst, S., Schmidt, N., & König, P. (2013). Saccadic momentum and facilitation of return saccades contribute to an optimal foraging strategy. *PLoS Computational Biology*, *9*(1), e1002871.

Yarbus, A. L. (1967). *Eye movements and vision* (Vol. 160). New York: Plenum Press.