Can a low-cost eye-tracker assess the impact of a valent stimulus? A study replicating the visual backward masking paradigm

Article

Accepted Version

Vlastos, D. D., Kyritsis, M. ORCID: https://orcid.org/0000-0002-7151-1698, Varela, V.-A., Gulliver, S. R. ORCID: https://orcid.org/0000-0002-4503-5448 and Papaioannou-Spiroulia, A. (2020) Can a low-cost eye-tracker assess the impact of a valent stimulus? A study replicating the visual backward masking paradigm. Interacting with Computers, 32 (2). pp. 132-141. ISSN 0953-5438 doi: https://doi.org/10.1093/iwc/iwaa010 Available at https://centaur.reading.ac.uk/90601/

Abstract

Capturing affective response to valent stimuli using eye-tracking is not only of interest to academic research, but also to commercial equipment developers (e.g., car dashboards). In order to investigate whether a low-cost eye-tracker can effectively detect participants' physiological response to negatively valent stimuli, forty-four participants aged 19-24 (mean = 24.7, SD = 5.8) were recruited to complete the visual backward masking paradigm in a repeated-measures experimental design. Saccadic duration and pupil sizes were recorded using a lower-end 60-hz tracker. Data was analysed using a mix of parametric and non-parametric tests. Our results suggest that valence in the form of fearful vs neutral faces has a significant main effect on both saccadic duration [V = 931, $p < 0.001$, d = 0.96] and pupil size [$t(43)$ = 29.81, $p < 0.001$, d = 3.91)]. Our findings were further supported by Bayes Factor analysis, which showed that saccadic duration data was 24 times more likely to occur, and pupil size measurement data was 89 times more likely, under the alternative hypothesis, showing that differences in valence had a main effect. The combined evidence produced by our Bayesian analysis, the large effect sizes of our frequentist analysis, and the significant effect on two separate measurements lead us to suggest that, under the right conditions, low-cost eye trackers can successfully detect changes in saccadic duration and pupil sizes as a result of physiological responses to threat-relevant visual stimuli.

Research Highlights

- Valent stimuli cause a measurable physiological response in the form of pupil dilation and saccadic delay, even in the absence of conscious processing.

- A low-cost eye-tracker was successfully used to detect affective response to valent stimuli by looking at relative differences in pupillometric and saccadic measurements.

Introduction

Eye-trackers, applying a range of technologies, have been used for over a century to record users' gaze and eye fixations. Eye- tracking is commonly used in user experience studies for A/B testing by collecting data related to users' fixations and further analysing the data using heatmaps (Djamasbi, 2014). Interestingly, eye-tracking is increasingly being used to detect cognitive and emotional response to valent stimuli, i.e., stimuli that evoke a positive or negative emotional response (Bailenson et al., 2008; Charoenpit & Ohkura, 2015; Guo, Ding, Liu, Liu, & Zhang, 2016; Guo, Ding, Wang, Liu, & Jin, 2016; Kalliatakis, Stergiou, & Vidakis, 2017). Since affective response to particularly negative valent stimuli can modulate cognition and impact both user experience and performance (Hudlicka, 2003), the capture and analysis of cognitive and emotional response, by using eye-tracking, is a useful tool within user experience studies. For example, the capture of response data is particularly important in areas such as transportation, where a fearful response by the driver could alert a vehicle system that something is wrong (Zhang, Ihme, & Drewitz, 2019).

Modern eye-tracking equipment varies significantly in functionality and cost. Most academic studies, especially in the field of interface evaluation, use mid-to-high-end eye-trackers, which cost tens of thousands of dollars, in order to measure gaze fixations, saccadic durations, and pupil dilations (Djamasbi, Siegel, Tullis, & Dai, 2010; Goldberg, Stimson, Lewenstein, Scott, & Wichansky, 2002; Kyritsis, Gulliver, & Feredoes, 2016; Russell, 2005). High-end equipment with sampling rates higher than 500Hz can support detailed measurement of emotional, cognitive, and

saccadic movement, yet the high cost makes high-end equipment practically unobtainable for most non-specialist researchers, e.g., undergraduate and masters level students. Moreover, the significant equipment and software cost involved in using high-end eye-tracking equipment practically limits incorporation within commercial interface/device designs (e.g., car dashboards, computer monitors, laptop interfaces, home appliances, etc.). Low-cost eye-trackers, with sampling rates of between 30 to 60Hz, cost a few hundred dollars and are primarily developed to capture participant gaze fixation data (Guasconi et al., 2017). Even though the accuracy of such devices is limited, the inexpensive nature of low-end eye-tracking devices, and the common availability of relevant Software Development Kits (SDKs), makes them attractive to cost aware users. But can low cost eye-tracker devices truly distinguish the presence of valent stimuli?

There is strong evidence that humans prioritise the processing of emotional information. For example, when asked to detect a specific visual stimulus (target) amongst distractors, stimuli with emotional valence (i.e., fearful faces, spiders, snakes, violent scenes, and others) cause a measurable physiological response (Lundqvist, Juth, & Öhman, 2014). This response results in an increase in saccadic duration, or even an increase in saccadic latency (the time it takes for onset of eye movement), which can be reliably detected by high-end eye-tracking equipment (McSorley & Van Reekum, 2013). The ability to quickly identify dangerous stimuli, i.e., stimuli provoking fear or anger, is a significant evolutionary advantage, since it supports individuals when approaching, or experiencing certain dangerous or life-threatening situations (Öhman, Flykt, & Esteves, 2001). On the other hand, this same process can also distract participants when undertaking a task. A physiological response to emotional valence has been measured even when stimulus is presented subliminally to participants (Del Zotto & Pegna, 2015). If emotional valence is present whilst undertaking a task, even if the stimulus is "subliminal" (i.e., processed unconsciously), then the

participant will be slower at performing the task. This is due to additional prioritised cognitive processing caused by an unconscious response, which has some interesting implications if a required response is time dependent.

Kahneman & Beatty (1966) demonstrated that pupil dilation closely follows an increased level of cognitive processing (100 ms - 200 ms), and that the level of pupillary dilation relates closely to 'mental effort', e.g., increases in use of short-term memory, language processing, reasoning, and perceptual discrimination (Beatty, 1982). May et al. (1990) showed that increased mental effort also impacts the extent of saccadic movements, with movements being restricted as the complexity of the task is increased. Although Dalmaijer (2014) demonstrated that low-end eye-trackers are poor at measuring saccades accurately (i.e. they have a low precision), it is contextually important to note that determining the relative presence (or absence) of valent stimuli does not require capture of precise saccadic duration. Instead the focus should be on the relative differences between saccadic durations.

Backward masking aims to facilitate the measurement of unconscious physiological responses, without the participant being consciously aware of the stimulus content. Visual backward masking involves the brief (i.e., 15-30 ms) presentation of a specific visual stimulus, which is subsequently masked from user gaze to avoid participants being consciously aware of the stimulus (Mogg & Bradley, 1999, 2002; Fox, 2002; Carlson & Reinke, 2008). Literature suggests that amygdala activation, from pre-attentive exposure to fearful stimuli, generates a measurable physiological response; i.e., i) pupil size measurements from masked fearful face trials will be significantly larger than pupil size measurements from masked neutral face trials, and ii) saccadic duration from masked fearful face trials will be significantly longer than saccadic duration from masked neutral face trials, as cognitive load increases due to the prioritised processing of the

fearful stimulus.

In this study, the authors aim to replicate the popular visual backward masking paradigm using fearful and neutral faces in order to determine whether physiological differences in pupil size and saccadic duration, between fearful and neutral faces, can be effectively measured and distinguished using a low-cost eye-tracker. Our hypothesis is that despite the inaccuracy of low-cost systems, the relative difference in physiological measurements, between fearful and neutral faces, will remain significant, allowing the presence (or absence) of valent stimuli to be identified.

Related Work

Measuring saccadic durations and pupil sizes using lower-end eye-tracking

Coyne & Sibley (2016) shared our concerns regarding the barriers formed by the high costs of eye-tracking, and subsequently used low-end specification eye-tracking devices to detect changes in pupil size in response to increased cognitive load. Coyne & Sibley (2016) recruited ten participants to engage with the digit span task (DST), which involves participants hearing a sequence of integers, and then repeating them back to the experimenter. The size of the sequence increases by two, starting with three numbers, and ending with nine numbers. With increases in sequence size, there is an expected increase in cognitive load, which has been shown to cause a measurable increase in pupil diameter (Klingner, Tversky, & Hanrahan, 2011). Despite using a low-end eye-tracking device (30hz to 60hz sampling rates), the results showed significant differences in pupil diameters between low levels of cognitive load (sequence of three numbers), and higher levels of cognitive load (sequences higher than three).

Bradley et al. (2008) invited eleven participants to view emotionally charged pictures (pleasant, neutral, and unpleasant) in a lab-controlled experiment. A low-end 60Hz eye-tracker

was used to record pupil size data and was correlated to skin conductance. The authors reported a significant main effect of valence on pupil diameter, heart rate, and skin conductance. These findings were also confirmed by Steenbergen, Band, & Hommel (2011), who asked eleven participants to look at emotionally charged pictures (neutral, positive, or negative), and then either look towards (prosaccade) or away (antisaccade) from a target. The authors reported a significant main effect of image content on pupil sizes, with negative pictures producing larger pupil diameter and saccadic latency compared to positive pictures.

Differences in pupil size measurements in high versus low working memory conditions were also studied by Tsukahara, Harrison, & Engle (2016). Using a low-end 60hz tracker, the team found that during several working memory tasks there was a significant increase in pupil size as memory load increased. This result was also affected by working memory capacity and fluid intelligence, with higher scoring participants more able to maintain their mental effort when placed under higher memory loads.

Despite some studies suggesting that low-cost eye-tracking devices can successfully detect pupil size differences in response to changes in mental effort, other studies are less optimistic. Sibley et al. (2018) recruited 90 Navy and Marine Corps officers to take part in an operation span task (OSPAN). The OSPAN is a working memory test that begins by asking participants to solve a simple mathematical problem. Participants are then asked to read a word and retain it in memory. The words and maths problems alternate and at certain intervals the participant is asked to recall the words that they were asked to retain. The authors tracked pupil sizes using a low-cost eye-tracker and found an unexpected significant negative correlation between pupil diameter and working memory load. The authors conclude that the result may have been the result of an error during the conversion of measurements from pixels to millimetres, but also raised caution against

using low-cost eye-tracking when capturing psychophysiological measurements.

For saccadic measurements, there is contradictory evidence on the validity of low-end eye-tracking equipment for such tasks. Leube, Rifai, & Wahl (2017) recruited eleven participants to perform a reading task and recorded saccadic durations using two eye-trackers (60Hz and 120Hz). It is important to note that saccadic durations during reading are much shorter than those found in fixation-to-target event-related experimental designs. The study reported a small but significant difference in the reliability of estimating saccadic duration between the 60Hz and 120Hz eye-trackers. The authors concluded that reliable detection of small saccadic durations, as occurs during a reading task, requires a sampling rate higher than those currently captured by low-cost eye-trackers.

On the other hand, Gibaldi, Vanegas, Bex, & Maiello (2017), validated the use of a lower-end Tobii EyeX eye-tracking device (set at 55Hz sampling rate) while developing a Matlab Toolkit that could interface with the device for use with behavioural experiments. The authors recruited fifteen participants to take part in a series of tasks that measured latency and duration of participant saccades in a repeated-measures experimental design. One of the tasks required participants to follow a target stimulus around the screen as closely as possible. The authors found that both saccadic latency, as well as saccadic durations during the task were consistent with literature on relevant tasks.

Measuring affective response to valent stimuli using medium-to-higher-end eye-trackers

Medium-to-higher-end eye-trackers are commonly used as tools for measuring physiological responses to threat-relevant stimuli. For example, Morris, McSorley, & Van Reekum (2018) recruited forty-two participants to engage with a 'follow the cross' task while their eye movements were recorded by a 250Hz eye-tracker. The task required participants to fixate on a cross that

changed position on the screen, while a picture of emotional facial expressions (neutral, happy, or angry face) appeared either on top or below the target. The authors reported a significant decrease in saccadic latency to targets when angry faces vs happy faces were used as distractors. Interestingly this effect was subject to intolerance of uncertainty test scores (a 27-item self-reporting measure to test whether the participant negatively reacts to emotional uncertainty), with high scoring participants showing significant effects towards valent stimuli, and low scoring participants not affected by the valent distractors.

Additional evidence supporting the use of eye-tracking for detecting affective responses comes from Hepsomali et al. (2017). The team recruited twenty-seven participants to undertake a task that involved fixating on an emotional facial expression (either angry, fearful, happy, or neutral) in the centre of the screen for 1000ms before the target changed position. The participants were then asked to either look towards (prosaccade) or away (antisaccade) from the target. Pupillometric and saccadic movements were recorded using a 1000Hz eye-tracker. The results of the study showed a significant effect of valence on pupil size, saccadic latency, and saccadic errors (i.e., the eye moving in the wrong direction). Differences in pupil sizes, in particular, showed a large effect, with angry and fearful (i.e., threat-relevant) stimuli causing the largest dilations in pupils when compared to happy faces.

Finally, Bodenschatz, Kersting, & Suslow (2019) used a 250Hz eye-tracker to investigate the effects of masked emotional facial expressions on gaze behaviour. Seventy-three participants were recruited to evaluate the valance of masked facial expressions (i.e., happy, neutral, angry, sad, and fearful). Participants were asked to look at a fixation point. After 1000ms the user was then presented with, for 50ms each: a forward mask containing just textual characters, the emotional facial image, and a backwards mask containing only textual characters. After the

backwards mask the user was presented with a neutral face for 1850ms, which was used to evaluate user fixation. The authors reported that when a fearful facial expression was used there was i) early saccadic movement towards regions of the neutral face that support diagnosis of fearful faces, and ii) more time spent fixating on fearful face regions. In both cases the effect sizes of the results were moderate. Awareness checks confirmed that participants were not consciously aware of the facial expressions of the masked stimuli, so were not consciously aware of the fearful face. Accordingly, the results of the study provide evidence that fearful faces will lead to an affective response, which in turn will lead to an increase in saccadic duration and latency.

In summary, medium-to-higher-end eye-trackers have been successfully used to measure physiological responses to valent (particularly threat-relevant) stimuli. There are some contradictions in literature, however, regarding the validity of low-cost eye-tracking as a suitable method of detecting physiological responses to lab-controlled working memory tasks. There is evidence to suggest medium-to-higher-end eye-trackers can be used reliably to detect physiological effects of valent stimuli, but as far as the authors are aware, this study is the first to examine whether low-cost eye-tracking can successfully detect pupillary and saccadic changes as a result of introducing lab-controlled masked threat-relevant stimuli to participants.

Method

Participants

Power analysis was used to determine a reasonable sample size for our experimental study. Our dependent variables (DV) for this study were pupil size and saccadic duration. Our independent variable (IV) was the stimulus valence (two independent groups: fearful to neutral and neutral to

neutral). Using a repeated measures MANOVA, with two measurements and two groups, a medium effect size (0.3), a statistical power of 0.8 (i.e., 80% chance of finding an effect when one exists), and a type I error of α = 0.05, analysis indicates that a sample size of approximately 45 participants was deemed suitable. 44 volunteers were recruited from computing and psychology BSc programmes (13 males and 31 females), aged 19–48 ($\mu = 24.7$, $\sigma = 5.8$), from CUC (Athens, Greece). Informed consent was obtained from all individual participants included in the study. Participants had normal or corrected-to-normal vision, were in good health, and none of the participants reported a history of psychiatric and neurological disorders. The experiment was approved by the Cardiff Metropolitan University Research and Ethics Committee and was conducted in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

Materials

The experiment was developed in Java and was set up to, within a Windows-based machine, support full compatibility with the eye-tracker's SDK. A low-cost USB bar eye-tracker was used to record both saccadic movements and pupil diameter (i.e., The EyeTribe, 2016). According to the manufacturer, the sampling rate of the eye-tracker is 60 samples every second, providing a latency of less than 20ms. The angular resolution, calculated using the root mean squares of successive samples, is 0.1º, and the mean accuracy is 0.5º (see Table 1 for a summary of technical specifications of the eye-tracking system).

(Table 1 about here, caption at the end of the manuscript)

Fearful and neutral face images were taken, with permission, from the Warsaw Set of Emotional Facial Expression Pictures (WSEFEP) library. WSEFEP images were used since the image set had previously been validated by Olszanowski et al. (2014). A total of 19 images were converted to grayscale to limit unconscious capture of attention due to colour effects, and then rescaled (490px * 640px) in order to fit the screen. Nine fearful faces were selected, which were previously rated with the highest "fearful" scores, and ten neutral faces, which were previously rated with the highest "neutral" scores (see Fig. 1 for example images). The number of fearful and neutral images was not arbitrary, rather the face validation in the WSEFEP image sets for these images decreased significantly after the ninth and tenth faces. Finally, the luminance of face stimuli was made consistent to decrease any effect that luminance might have on pupillary light response (Bradley et al., 2008).

(Fig. 1 about here, caption at the end of the manuscript)

The selected eye-tracker supported screen sizes of up to 24 inches. Accordingly, a 22" LCD monitor (HP, Compaq LA2205WG) was used, with an aspect ratio of 16:10, resolution of 1680px * 1050px, and a refresh rate of 60Hz. The task relevant stimuli, i.e., the set of images that participants were consciously attending during trials, were made up of eight 'cards' containing distinctly different shapes and colours (see Fig. 2). Stimuli differed in shape and colour to avoid bottom-up confounding effects due to feature similarities during the selective attention task. The stimuli were not converted to grayscale (unlike the distractors), in order to facilitate the attention task by providing additional features for target discrimination. Each card was 200px * 320px with a 24-bit colour depth.

(Fig. 2 about here, caption at the end of the manuscript)

Experimental Procedure

Participants were brought into the lab one at a time. All participants sat ~70 cm away from the screen, which was inside the operating range of the tracker (45 cm – 75 cm). The eye-tracker was placed under the monitor and was calibrated for each participant using a standard 5-point grid at the start of each experiment, which is an integrated function included with the eye-tracker's SDK. Calibration was repeated until optimal calibration was achieved, leading to ~0.5 degrees of error, which is approximately half a thumb's width at an arm's length. The laboratory was luminated by uniform neon lights to avoid optical artefacts. Sun blinds were also used at all times to minimise impact by external light sources.

In more detail, the experiment followed a repeated measures design with two conditions: i) fearful distractor to neutral mask (fearful) and ii) neutral distractor to neutral mask (neutral). Andersson, Nyström, & Holmqvist (2010) provide a formula (and a graph) for deciding the minimum number of trials for an alpha threshold of 0.05. For a 60Hz tracker it was calculated that approximately 30 trials were needed for each effect. Since only half the trials would contain a neutral-neutral mask, and the other half would contain a fearful-neutral mask, the total number of trials was set to 60. The choice of distractors and targets was completely randomised. In each trial, the participant started by looking at a cross at the centre of the screen, i.e., the fixation point. Participants were instructed to focus on the fixation point and the cards, but to ignore the distractor faces. After 900 ms, participants were shown a shape, presented on a target card, which was visible at the centre of the screen for 500 ms. The target card was replaced by another cross at the fixation

point. Participants were encouraged not to blink when the fixation cross appeared on the screen, and to react as soon as the cards appeared. After a further 900 ms the participant was presented with a selection of two cards, and a face distractor (see figure 3), with the face distractor presented randomly 5.5 degrees of visual angle above or below the fixation point. One card contained the target shape, and one card contained a non-target shape. Participants were asked to find and fixate on the card that contained the target shape. Participants were also asked to ignore the face distractors, i.e., the visual backward masked faces that appeared on the screen positioned centrally between the two cards (see figure 3). The final masking face distractor appeared neutral in extrafoveal vision, but half the trials contained a subliminal fearful face, which was 'concealed' after 17 ms by the neutral face (see Fig. 3).

(Fig. 3 about here, caption at the end of the manuscript)

The target and non-target shape cards were separated by approximately 14° of visual angle, which induced saccadic eye movements between the card fixation points, in order to process the shapes, i.e., to fixate on the correct card in order to move on to the next trial.

To assess whether fearful faces were consciously processed by participants, at the end of the experiment, all participants were asked if they noticed anything "strange" during the experiment. None of the participants reported seeing fearful faces, although they were aware that there were neutral faces visible at the end of each sample. Most participants reported some picture movement, but they attributed this to a "glitch" in the experimental graphics. Data associated with this study is available at https://github.com/dvlastos/eye-tracking-data.

Results

Two continuous dependent variables (DV) were measured in our experiment: average pupil size measurements (which was in arbitrary units) from the start of the target detection till fixation on peripheral target or non-target, and total saccadic duration in milliseconds. We approximated saccadic duration by calculating the total time from the end of the eye fixation on the fixation cross and the start of the eye fixation on the peripheral target or non-target. The (categorical) independent variables (IV) were fear/neutral using two experimental conditions: fearful distractor to neutral mask or neutral distractor to neutral mask.

Saccadic duration and pupil size were averaged from 60 trials for each of the 44 participants, resulting in 88 observations in total, i.e., 44 fearful and 44 neutral, with 30 trials each. It is important to note that by adopting a repeated measures design there was no within-group variation of pupil diameter, and, therefore, the baseline pupil size differences that naturally exist between participants were not accounted for. Instead, the authors measured relative changes in pupil size between our two experimental conditions from each participant. A standard deviation filter was applied on the raw data (i.e., before it was aggregated for each participant) to remove any values that exceeded three standard deviations and could be indicative of errors with the equipment. Furthermore, we removed trials with zero values, as those are indicative of blinks. We note, however, that the change in the aggregated statistics before and after applying the filters was negligible, possibly due to the large number of trials that were aggregated for each participant.

The data were tested for the usual parametric assumptions for MANOVA using the Doornik-Hansen test for multivariate normality found in the 'MVN' library by Korkmaz, Goksuluk & Zararsiz (2014), and Levene's test for homogeneity of variances. The multivariate normality assumption was violated ($E(4) = 129.95$, $p < 0.001$). Shapiro-Wilk indicated that i) the

saccadic durations data was not normally distributed (W = 0.87, p < 0.001), yet ii) the pupil size data was normally distributed. The saccadic durations data was transformed using monotonic functions and found that normality was achieved when using the logarithm. Although transformed saccadic data was still slightly skewed (W = 0.96, p < 0.01), the authors argue that data was within the 'acceptable' range for the robust MANOVA test (see plots in Fig. 4). As there were no other violations to the parametric assumptions, with the help of the 'MANOVA.RM' library developed by Friedrich, Konietschke, & Pauly (2018), a repeated measures MANOVA was used as an omnibus test. The result of the omnibus test was significant (Wald-Type Statistic: W(2) = 38.2, p < 0.0001).

(Fig. 4 about here, caption at the end of the manuscript)

For the post-hoc tests the paired sample Wilcoxon test was used as a non-parametric alternative for the saccadic durations data, and a paired samples t-test for the pupil size data. Since two 'difference of means' tests are required, thus increasing the familywise error rate, multiple comparisons were controlled for by adjusting the p-values using the Holm method. The paired samples t-test showed that there was a significant difference in the means of pupil size measurements between the fearful (mean = 22.7, SD = 5.26) and neutral groups (mean = 18.17, SD = 4.39) [t(43) = 25.91, p < 0.001]. Wilcoxon signed rank test showed that the difference in the means for saccadic duration was also significant between the fearful (mean = 396 ms, SD = 97.26 ms) and neutral groups (mean = 331.99 ms, SD = 103.65 ms) [V=931, p < 0.001] (see Fig. 5).

(Fig. 5 about here, caption at the end of the manuscript)

Cohen's d for paired samples was used as a measure of effect size for both tests. The difference between saccadic duration had a large effect size (d = 0.96), as did the difference in pupil sizes (d = 3.91), indicating that the differences between the means of the two groups were not only statistically significant, but also 'meaningfully different'. Bayesian analysis of t-tests, using the 'BayesFactor' library by Morey, Rouder, & Jamil (2014), was used to further reinforce our effect findings. The Bayes Factors were extracted to investigate how much more likely the alternative hypothesis (i.e., there is a difference between the means of the fearful and neutral groups) is than the null (i.e., no difference in the means). For the saccadic duration, our results indicated that the data was 24 times more likely to occur under the alternative hypothesis than the null. For the pupil size data, our results indicate that the data was approximately 89 times more likely to occur under the alternative hypothesis than the null. According to the criterion by both Jeffreys (1961) and Kass & Raftery (1995), this is strong evidence (for saccadic duration) and very strong evidence (for pupil size) that the alternative hypothesis is supported by the data. Literature suggested that amygdala activation from pre-attentive exposure to fearful stimuli would result in i) pupil size measurement, from masked fearful face trials, being significantly larger than pupil size measurement from masked neutral face trials, and ii) saccadic durations, from masked fearful face trials, due to additional cognitive load, being significantly longer than saccadic durations from masked neutral face trials. This paper presents strong evidence that a low-cost eye-tracking device can detect relative physiological differences between trials that present fearful masked distractors, and trials that present neutral masked distractors during a target detection task.

General Discussion

In our analysis a frequentist approach was initially used to determine whether there were significant differences in the mean pupil sizes and saccadic duration, i.e., between fearful and neutral face trials. Findings indicate that participants took longer to saccade to the target when masked fearful faces were presented (~17 ms) than when a masked neutral face was presented. Furthermore, there was an increase in pupil dilation during fearful trials when compared to neutral trials. Bayes factor analysis was used to further support our findings and provided strong evidence in favour of the alternative hypothesis (difference in the means) for both eye-related measurements.

These results, along with the results of the significance tests, suggest that a low cost eye-tracker, with specifications as described in table 1, is (if calibrated) able to detect relative psychological differences in the visual backward masking experiment. The results presented in this paper agree with those predicted from literature, i.e., that participants would have an increase in saccadic duration and pupil size when presented with fearful stimuli prior to target fixation. Ultimately, data collected did show relative physiological differences, despite the eye-tracker only having a sampling rate of 60Hz, a latency of < 20 ms, an angular resolution of 0.1°, and a mean accuracy of 0.5°.

Conclusion

The use of eye-tracking in user experience studies has been well documented, with most researchers focusing on gaze-based A/B testing and heatmap analysis. There is a growing body of literature, however, that argues in favour of using eye-tracking to capture user affect data (e.g., Hudlicka, 2003; Krämer, Kopp, Becker-Asano, & Sommer, 2013; Tuch et al., 2009; Zhai & Barreto, 2006; Zhang, Ihme, & Drewitz, 2019). Affective responses to valent stimuli modulate

cognition (see Dominguez-Borras & Vuilleumier, 2013; Pourtois, Schettino, & Vuilleumier, 2013; Todd, Cunningham, Anderson, & Thompson, 2012) due to threat-relevant information activating the amygdala in the brain. Activation of the amygdala causes widespread physiological responses, which can interfere with target identification and task processing (Straube, Mentzel, & Miltner, 2006). Traditional approaches to measuring the impact of valent stimuli, e.g. electrodermal response (EDR), have significant limitations due to the intrusive nature of equipment setup. Measurement of pupil dilation has been shown to covary with many traditional approaches (Bradley, Miccoli, Escrig, & Lang, 2008), and occurs even when a stimulus is subliminal and images are masked from conscious awareness (Öhman, Esteves, & Soares, 1995). Although eye-tracking has already been validated as a tool for detecting affective response to valent stimuli (Bradley, Miccoli, Escrig, & Lang, 2008; Lundqvist, Juth, & Öhman, 2014; McSorley & Van Reekum, 2013; Zhai & Barreto, 2006), most previous studies favour the use of expensive high-end (research-grade) eye-tracking, which makes it difficult for interested researcher to engage in this type of research unless they have access to expensive equipment. This study investigates whether a low-end eye-tracker could be used to detect relative physiological differences between trials that present fearful masked distractors, and trials that present neutral masked distractors during a target detection task in an experiment known as the visual backward masking paradigm.

Our findings successfully showed significant differences between pupil sizes and saccadic duration in a repeated measures design. Our results suggest that a low-end eye-tracker, if used properly, can detect relative physiological changes that exist between trials that present fearful masked distractors, and trials that present neutral masked distractors during a target detection task. The significant effect of threat-relevant stimuli to changes in saccadic duration and pupil size measurements are in agreement with similar experiments that utilised mid- and high-end eye-

tracking equipment. Our results showed large effect size measurements for both the saccadic duration (d = 0.96), and the pupil size (d=3.91), which is consistent with the study by Hepsomali et al. (2017), and somewhat higher than the effects found by Bodenschatz, Kersting, & Suslow (2019).

It is important to state that low-end eye-trackers are not developed or intended to support detection of valent information, and that high-end eye-tracking equipment should ideally be used if available. Nevertheless, for those undertaking commercial research and/or research at lower levels (e.g., undergraduate and postgraduate students) a low-cost eye-tracker could be considered under controlled conditions. No claims can be made, as a result of this study, about using low-cost equipment within usability experiments, yet this study shows that low-cost eye-tracking, under the right conditions, can be used to detect relative affective response to valent stimuli. Therefore, the authors conclude that low-cost eye-tracking could be considered for use in studies that aim to capture relative variance in user emotion during interaction.

Limitations

A chin rest was not used for our eye-tracking study. Although this reduces participant feeling of intrusion it risks introducing motion-related error in our measurement. To address this issue, the authors relied on multiple trials for each participant (60 each) and averaged their output. State of the art algorithms for measuring saccadic duration such as ID-T and IV-T were not used. Instead saccadic duration was approximated by taking the total time from the end of a fixation on the fixation cross till the beginning of a fixation on the target or non-target. This research did not compare the results of the low-cost eye-tracker against any other trackers, which means this paper is unable to make any claims about its comparative accuracy in the visual backward masking

experiment. Although the authors support consideration of use, additional research is required to validate use in a range of experimental contexts. Finally, despite making all efforts to control for the effects of luminosity, it is possible that there were uncontrolled luminosity effects produced by differences in the target stimuli. Our analysis does not show evidence to support this, but it is still a possibility that needs to be considered. "

Competing Interest Statement

The authors declare no conflict of interest.

Supplemental Online Material

Data associated with this study has been deposited at https://github.com/dvlastos/eye-tracking-data.

References

Andersson, R., Nyström, M., & Holmqvist, K. (2010). Sampling frequency and eye-tracking measures: how speed affects durations, latencies, and more.

Bailenson, J., Pontikakis, E., & Mauss, I. (2008). Real-time classification of evoked emotions using facial feature tracking and physiological responses. *International journal of human-computer studies, 66*(5), 303-317. http://doi.org/10.1016/j.ijhcs.2007.10.011

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*. http://doi.org/10.1037/0033-2909.91.2.276

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*. http://doi.org/10.1111/j.1469-8986.2008.00654.x

Bodenschatz, C. M., Kersting, A., & Suslow, T. (2019). Effects of briefly presented masked emotional facial expressions on gaze behavior: An Eye-Tracking study. Psychological reports, 122(4), 1432-1448.

Carlson, J. M., & Reinke, K. S. (2008). Masked fearful faces modulate the orienting of covert spatial attention. *Emotion, 8*(4), 522.

Charoenpit, S., & Ohkura, M. (2015). Exploring emotion in an e-learning system using eye-tracking. *International Journal of Affective Engineering, 14*(4), 309-316. https://doi.org/10.5057/ijae.IJAE-D-14-29

Coyne, J., & Sibley, C. (2016). Investigating the use of two low cost eye-tracking systems for detecting pupillary response to changes in mental workload. In Proceedings of the Human

Factors and Ergonomics Society Annual Meeting (Vol. 60, No. 1, pp. 37-41). Sage CA: Los Angeles, CA: SAGE Publications.

Dalmaijer, E. S. (2014). Is the low-cost EyeTribe eye-tracker any good for research? *PeerJ PrePrints*. http://doi.org/10.7287/peerj.preprints.141v2

Del Zotto, M., & Pegna, A. J. (2015). Processing of masked and unmasked emotional faces under different attentional conditions: An electrophysiological investigation. *Frontiers in Psychology*. http://doi.org/10.3389/fpsyg.2015.01691

Djamasbi, S. (2014). Eye-tracking and web experience. *AIS Transactions on Human-Computer Interaction, 6*(2), 37-54.

Djamasbi, S., Siegel, M., Tullis, T., & Dai, R. (2010). Efficiency, trust, and visual appeal: Usability testing through eye-tracking. In *Proceedings of the Annual Hawaii International Conference on System Sciences*. http://doi.org/10.1109/HICSS.2010.171

Dominguez-Borras, J., & Vuilleumier, P. (2013). Affective biases in attention and perception. In *The Cambridge Handbook of Affective Neuroscience* (pp. 331–356). http://doi.org/10.1017/CBO9780511843716.018

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

Friedrich, S., Konietschke, F., & Pauly, M. (2018). Analysis of Multivariate Data and Repeated Measures Designs with the R Package MANOVA. RM. arXiv preprint arXiv:1801.08002.

Goldberg, J. H., Stimson, M. J., Lewenstein, M., Scott, N., & Wichansky, A. M. (2002). Eye-tracking in web search tasks: design implications. *Proceedings of the 2002 Symposium on Eye-tracking Research & Applications*. http://doi.org/10.1145/507072.507082

Gibaldi, A., Vanegas, M., Bex, P. J., & Maiello, G. (2017). Evaluation of the Tobii EyeX Eye

tracking controller and Matlab toolkit for research. *Behavior research methods, 49*(3), 923-946.

Guasconi, S., Porta, M., Resta, C., & Rottenbacher, C. (2017). A low-cost implementation of an eye-tracking system for driver's gaze analysis. In 2017 10th International Conference on Human System Interactions (HSI) (pp. 264-269). IEEE.

Guo, F., Ding, Y., Wang, T., Liu, W., & Jin, H. (2016). Applying event related potentials to evaluate user preferences toward smartphone form design. *International Journal of Industrial Ergonomics*. http://doi.org/10.1016/j.ergon.2016.04.006

Guo, F., Ding, Y., Liu, W., Liu, C., & Zhang, X. (2016). Can eye-tracking data be measured to assess product design?: Visual attention mechanism should be considered. *International Journal of Industrial Ergonomics*. http://doi.org/10.1016/j.ergon.2015.12.001

Hepsomali, P., Hadwin, J. A., Liversedge, S. P., & Garner, M. (2017). Pupillometric and saccadic measures of affective and executive processing in anxiety. Biological psychology, 127, 173-179.

Hudlicka, E. (2003). To feel or not to feel: The role of affect in human-computer interaction. International Journal of Human Computer Studies. http://doi.org/10.1016/S1071-5819(03)00047-8

Jeffreys, H. (1961). *Theory of Probability*. Oxford University Press, London.

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*. http://doi.org/10.1126/science.154.3756.1583

Kalliatakis, G., Stergiou, A., & Vidakis, N. (2017). Conceiving Human Interaction by Visualizing Depth Data of Head Pose Changes and Emotion Recognition via Facial Expressions. *Computers*. http://doi.org/10.3390/computers6030025

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*. http://doi.org/10.1080/01621459.1995.10476572

Klingner, J., Tversky, B., & Hanrahan, P. (2011). Effects of visual and verbal presentation on cognitive load in vigilance, memory, and arithmetic tasks. Psychophysiology, 48(3), 323-332.

Korkmaz, S., Goksuluk, D., & Zararsiz, G. (2014). MVN: An R package for assessing multivariate normality. *The R Journal*. http://doi.org/10.1.1.661.7273

KräMer, N., Kopp, S., Becker-Asano, C., & Sommer, N. (2013). Smile and the world will smile with you—The effects of a virtual agent's smile on users' evaluation and behavior. International Journal of Human-Computer Studies, 71(3), 335-349.

Kyritsis, M., Gulliver, S. R., & Feredoes, E. (2016). Environmental factors and features that influence visual search in a 3D WIMP interface. *International Journal of Human Computer Studies*. http://doi.org/10.1016/j.ijhcs.2016.04.009

Leube, A., Rifai, K., & Wahl, S. (2017). Sampling rate influences saccade detection in mobile eye-tracking of a reading task. J. Eye Mov. Res., 10, 3.

Lundqvist, D., Juth, P., & Öhman, A. (2014). Using facial emotional stimuli in visual search experiments: The arousal factor explains contradictory results. *Cognition and Emotion*. http://doi.org/10.1080/02699931.2013.867479

May, J. G., Kennedy, R. S., Williams, M. C., Dunlap, W. P., & Brannan, J. R. (1990). Eye movement indices of mental workload. *Acta Psychologica*. http://doi.org/10.1016/0001-6918(90)90067-P

McSorley, E., & Van Reekum, C. M. (2013). The time course of implicit affective picture processing: an eye movement study. *Emotion (Washington, D.C.)*. http://doi.org/10.1037/a0032185

Mogg, K., & Bradley, B. P. (1999). Orienting of attention to threatening facial expressions presented under conditions of restricted awareness. *Cognition & Emotion, 13*(6), 713-740. https://doi.org/10.1080/026999399379050

Mogg, K., & Bradley, B. P. (2002). Selective orienting of attention to masked threat faces in social anxiety. *Behaviour research and therapy, 40*(12), 1403-1414.

Morey, R. D., Rouder, J. N., & Jamil, T. (2014). BayesFactor: Computation of Bayes factors for common designs. *R Package Version 0.9.*

Morriss, J., McSorley, E., & Van Reekum, C. M. (2018). I don't know where to look: the impact of intolerance of uncertainty on saccades towards non-predictive emotional face distractors. *Cognition and emotion, 32*(5), 953-962.

Öhman, A., Esteves, F., & Soares, J. J. F. (1995). Preparedness and preattentive associative learning: Electrodermal conditioning to masked stimuli. In *Journal of Psychophysiology*.

Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. *Journal of Experimental Psychology: General*. http://doi.org/10.1037/0096-3445.130.3.466

Olszanowski, M., Pochwatko, G., Kuklinski, K., Scibor-Rylski, M., Lewinski, P., & Ohme, R. K. (2014). Warsaw set of emotional facial expression pictures: A validation study of facial display photographs. *Frontiers in Psychology*. http://doi.org/10.3389/fpsyg.2014.01516

Pourtois, G., Schettino, A., & Vuilleumier, P. (2013). Brain mechanisms for emotional influences on perception and attention: What is magic and what is not. *Biological Psychology*. http://doi.org/10.1016/j.biopsycho.2012.02.007

Russell, M. C. (2005). Hotspots and Hyperlinks: Using Eye-tracking to Supplement Usability Testing. *Usability News*.

Sibley, C., Foroughi, C., Brown, N., & Coyne, J. T. (2018, September). Low Cost Eye-tracking: Ready for Individual Differences Research?. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 62, No. 1, pp. 741-745). Sage CA: Los Angeles, CA: SAGE Publications.

Straube, T., Mentzel, H. J., & Miltner, W. H. R. (2006). Neural mechanisms of automatic and direct processing of phobogenic stimuli in specific phobia. *Biological Psychiatry*. http://doi.org/10.1016/j.biopsych.2005.06.013

Todd, R. M., Cunningham, W. A., Anderson, A. K., & Thompson, E. (2012). Affect-biased attention as emotion regulation. *Trends in Cognitive Sciences*. http://doi.org/10.1016/j.tics.2012.06.003

Tsukahara, J. S., Harrison, T. L., & Engle, R. W. (2016). The relationship between baseline pupil size and intelligence. Cognitive psychology, 91, 109-123.

Tuch, A. N., Bargas-Avila, J. A., Opwis, K., & Wilhelm, F. H. (2009). Visual complexity of websites: Effects on users' experience, physiology, performance, and memory. International journal of human-computer studies, 67(9), 703-715.

Van Steenbergen, H., Band, G., Hommel, B. (2011). Threat but not arousal narrows attention: evidence from pupil dilation and saccade control. Frontiers in psychology, 2, 281.

Zhai, J., & Barreto, A. (2006, August). Stress detection in computer users based on digital signal processing of noninvasive physiological variables. In 2006 international conference of the IEEE engineering in medicine and biology society (pp. 1355-1358). IEEE.

Zhang, M., Ihme, K., & Drewitz, U. (2019). Discriminating drivers' emotions through the dimension of power: Evidence from facial infrared thermography and peripheral physiological measurements. Transportation research part F: traffic psychology and behaviour, 63, 135-143.

Table Captions

Table 1. Technical Specifications of the Eye-tracking System.

Figure Captions

Figure 1. Examples of Neutral and Fearful Stimuli (Distractors), taken from the Warsaw Set of Emotional Facial Expression Pictures (WSEFEP).

Figure 2. List of task-relevant (target) stimuli.

Figure 3. Task schematic illustrating each trial. Note, faces randomly appeared above or below the central fixation point.

Figure 4. Q-Q plot diagram (left) and density plot (right) of saccadic measurements after applying a logarithmic transformation.

Figure 5. Differences between pupil sizes (top) and saccadic times in ms (bottom) when comparing fearful-to-neutral and neutral-to-neutral groups. Note that pupil size measurements are in arbitrary units. Error bars are 95% CI.