# Guiding Visual Attention on 2D Screens: Effects of Gaze Cues from Avatars and Humans

Julius Albiz
jalbiz@kth.se
KTH Royal Institute of Technology
Stockholm, Sweden

Olga Viberg
oviberg@kth.se
KTH Royal Institute of Technology
Stockholm, Sweden

Andrii Matviienko
andriim@kth.se
KTH Royal Institute of Technology
Stockholm, Sweden

## ABSTRACT

Guiding visual attention to specific parts of an interface is essential. One powerful tool for guiding attention is gaze cues, that direct visual attention in the same direction as a presented gaze. In this paper, we explored how to direct users' visual attention on 2D screens using gaze cues from avatars and humans. For this, we conducted a lab experiment (N = 30) based on three independent variables: (1) *stimulus* shown either as avatars or human faces, (2) *target direction* with a target appearing left or right from a stimulus, and (3) *gaze validity* indicating whether a stimulus' gaze was directed towards a target (valid gaze) or not (invalid gaze). Our results show that participants' total and average fixation on a target lasted longer in the presence of the human image than the avatar stimulus when a target appeared on the right side and when a stimulus' gaze was towards the target. Moreover, participants' average fixation was longer on the human than avatar stimulus gazing in the opposite direction from a target than towards it.

## CCS CONCEPTS

• **Human-centered computing → User studies**; **Empirical studies in HCI**.

## KEYWORDS

visual attention, gaze cues, attention guidance, virtual avatars

**ACM Reference Format:**
Julius Albiz, Olga Viberg, and Andrii Matviienko. 2023. Guiding Visual Attention on 2D Screens: Effects of Gaze Cues from Avatars and Humans. In *The 2023 ACM Symposium on Spatial User Interaction (SUI '23), October 13–15, 2023, Sydney, NSW, Australia.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3607822.3614529

## 1 INTRODUCTION

Within User Experience Design, effectively guiding users' attention in an interface is essential [9, 22, 37]. Visual Hierarchy is one of the established methods defined as the order in which the human eye is guided to consume each design element of an interface in the intended way [9]. To facilitate a specific order of visual guidance, researchers have introduced different visual cues such as arrows, words, and gaze to encourage users to interact more with an interface to complete tasks [31]. Following another human gaze is a clear example of Cialdini's Social Proof [8] saying that if an object has caught the attention of others, it will probably be of interest to us as well. However, with constantly increasing attention-hungry user interfaces on 2D screens, e.g., websites and smartphone applications, guiding visual attention via *gaze cues* becomes even more complex and requires a better understanding of efficient strategies.

Gaze cues are social cues used to direct visual attention based on the direction of other people's gaze. It is often used in websites to allocate users' visual attention to parts of the website that are important to interact with [25]. Websites typically use gaze cues by displaying images of real-life people with a certain gaze direction to guide visual attention to call-to-action objects such as buttons. With the emergence and fast expansion of technology, more and more people interact daily with computers and virtual environments that include virtual characters, i.e., "avatars". Therefore, previous research focused on making these avatars more human-like by adding gaze cues [3, 18]. When researching gaze cues and how they affect users, eye tracking is often used as an evaluation method, as it provides designers a more psychological approach to usability testing [35]. Expanding the usage of gaze cues to avatars can be beneficial for settings where avatars need to communicate, e.g., video games [16] or immersive learning settings [20]. As the goal of avatars in these settings is to make the interactions as natural as possible, being able to use human-like social patterns such as gaze cues will help fulfill this goal. Therefore, this study evaluates the effects of gaze cues, i.e., eye movements indicating a certain direction, provided by avatars compared to humans.

In this paper, we explored how to direct users' visual attention on 2D screens using gaze cues from avatars and humans. For this, we conducted a controlled lab experiment (N = 30) based on three independent variables: (1) *stimulus* shown either as avatars or human faces, (2) *target direction* with a target appearing either on the left or right side from a stimulus, and (3) *gaze validity* indicating whether a stimulus' gaze was directed towards a target (valid gaze) or in the opposite direction from it (invalid gaze). Our results show that participants' total and average fixation on a target lasted longer in the presence of the human than the avatar stimulus when a target appeared on the right side and when a stimulus' gaze was looking in the target's direction. Moreover, participants' average fixation was longer on a human than on the avatar stimulus with an eye gaze in the opposite direction from a target than in the target's direction. Our research contribution includes an empirical evaluation and design guidelines for directing visual attention on 2D screens using gaze cues from avatars and humans.

---

*SUI '23, October 13–15, 2023, Sydney, NSW, Australia*

ACM ISBN 979-8-4007-0281-5/23/10…$15.00
https://doi.org/10.1145/3607822.3614529

## 2 RELATED WORK

This section provides an overview of three main pillars of related work: (1) visual hierarchy in web design, (2) eye tracking and visual attention, and (3) gaze cues on avatars and humans.

### 2.1 Visual hierarchy in web design

Since the birth of the internet, websites have been a medium for communicating information and have found various areas of use [22]. Due to websites being often task-oriented, how a website can communicate information to users in a way beneficial for the task at hand is essential for web designers [22]. Within a website, designers present the visual elements to the users, with the users mentally assembling the elements to uncover the meaning behind them. Good web design particularly focuses on how efficiently it guides users' visual attention from one element to another and does it in the correct order [22]. This is also known as Visual Hierarchy and is an essential concept in the advertisement field [9]. The main goal of visual advertisement design is (1) visual communication and (2) visual attention allocation. Several layout patterns take advantage of people's scanning and reading patterns on a visual interface [9]. The three main are the Guttenberg diagram, the z-pattern layout, and the f-pattern layout. The Z-pattern and F-pattern layouts facilitate interaction with elements in a pattern that follows the letters Z or F. On the contrary, the Guttenberg Diagram is a pattern that follows an even distribution of information and suggests that the user's attention sweeps across the interface in a series of horizontal movements called axes of orientation. Each of these axes starts increasingly from the left edge and moves towards the right edge. This pattern, therefore, suggests that users pay the least attention to the bottom-left part of the interface [9].

### 2.2 Eye tracking and visual attention

Eye tracking is typically employed to assess users' cognitive processes and distribution of their visual attention [24, 35]. Eye tracking is an experimental method that records users' eye movements and gaze locations across time and task [6], such as fixations and saccades. For example, Eraslan et al. [10] used eye tracking to investigate if participants with ASD had different strategies for processing information on websites compared to participants with no neurological disabilities. Results showed that the participants with ASD tended to look more at irrelevant visual elements, had shorter fixation durations, and had longer scan paths. Eye tracking also allows designers to assess users' visual attention to understand why certain elements are not interacted with. For example, Boardman and Mccormick [5] used eye tracking to understand consumer viewing patterns on shopping websites. The results showed that users' attention was directed in a different pattern than in product listing or information, supported by another study about viewing strategies on a Facebook website [29]. It indicates that users' viewing strategy depends on motivation, unlike previous research that claims that people use specific patterns to scan textual websites [9, 23]. Another example includes the assessment of visual attention to the website advertisements [21, 36]. One way of advertising focuses on *banner-ads* – a rectangular display embedded into a website that redirects users to the sponsors' website if clicked. Previous research has focused on the design of banner ads to avoid banner blindness, i.e., when users ignore banner ads consciously or unconsciously and increase the effectiveness of advertisements [21, 36]. Their results indicate that ads should have visual elements that stand out to grab visual attention [36]. Therefore, in this work, we employed eye tracking to better understand the allocation of users' visual attention in the presence of virtual and human-like avatars and directional cues, which we outline in the following subsection.

### 2.3 Gaze cues on avatars and humans

A gaze cue is a visual cue that can be a visual display of a human or a virtual avatar [25] to provide social information and influence human behavior [8]. As discussed in the previous subsection, eye tracking helps in understanding the effectiveness of advertisements, and one way of guiding users' attention to ads is by using gaze cues [12, 32]. For instance, Sajjacholapunt and Ball [28] measured the effectiveness of banner ads in three conditions: no face, mutual (the gaze direction is towards the users), and averted (a gaze cue is applied) gaze. Their results showed that dwell time on vertical banner ads was higher with averted gaze, and both averted and mutual gaze led to higher dwell times compared to the no face, with the averted gaze accumulating the highest dwell time. Another study explored the influence of gaze direction on food preferences [17]. Participants had to write down their willingness to pay, taste, and health preferences before the test, in which they looked at images of people with food and varying gaze directions. The results showed participants' willingness to pay more for the images that utilized gaze cues in the direction of the food and lower preference for the food on the images with gaze cues in the opposite direction from the food. Arrow cues are another visual cue to allocate visual attention on websites [7, 13, 19]. For example, Joseph et al. [13] used an fMRI scan to measure brain activity when presented with gaze and arrow cues and found that humans direct visual attention more unconsciously when presented with a gaze cue than with an arrow cue. Although humans might direct attention unconsciously regarding gaze cues, the overall cueing effect is similar to when presented with arrow cues [7, 19]. Moreover, arrow cues efficiently direct attention to a group of objects, whereas gaze cues allocate attention to the specific object [7]. These findings show no significant differences in the cueing effect, but gaze cues are more efficient in providing social aspects compared to arrows [17]. Thus, this work investigates gaze cues on human-like and virtual avatars to understand better how they affect visual attention allocation.

Social signals, e.g., eye gaze, play an important role in human-to-human [3, 15, 18] communication, which led to the implementation of social signals on virtual avatars [3]. In the presence of social cues on virtual avatars, users showed faster task-completion times [2, 14, 18], higher preferences for virtual avatars [3], and lower error rates [2]. Realistic and engaging avatars increase immersiveness, learning among users, and overall enjoyment [27]. For instance, Khoramashi et al. [14] employed gaze behaviors to make avatars realistic and engaging. In an experiment where participants completed a simple task of mirroring the avatar, two conditions were set, one in which the avatar used gaze cues and one in which it did not. Results showed that gaze cues significantly improved participants' reaction time to the avatar's movements, made the task feel less difficult, and showed that the avatar's gaze movements

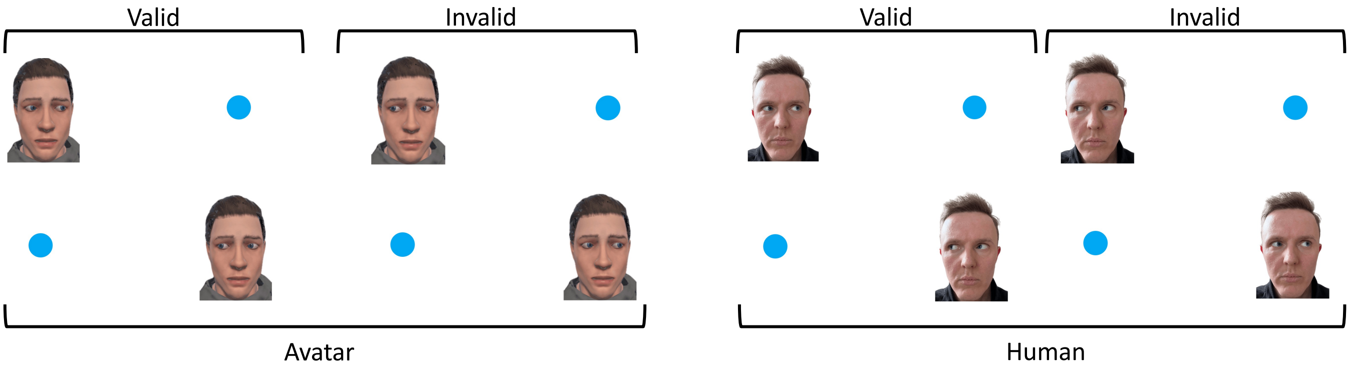

**Figure 1: Overview of eight experimental conditions: four conditions (left) included a virtual avatar and four conditions (right) a human-like avatar. They were further split into target direction, i.e., left or right from the avatar, and validity, i.e., the eyes looking in the same direction where a target appears or the opposite.**

were cooperative, human-like, and realistic. In summary, previous research has shown that gaze cues are important in guiding users' visual attention. However, there is a limited understanding regarding their effectiveness in directing visual attention using avatars and humans as stimuli since previous research shows that gaze cues are effective with humans and avatars. In this paper, we compare these types of stimuli, particularly in the presence of valid and invalid gaze guidance, which we describe in detail in the evaluation section in the following.

## 3 EVALUATION

We conducted a controlled lab experiment to assess the influence of virtual and human-like avatars on users' visual attention guidance. The research question for this experiment is: *How can we direct users' visual attention using virtual and human-like stimuli for 2D interaction?*

### 3.1 Participants

We recruited 30 participants (15 male, 15 female) aged between 21 and 50 ($M = 34.1, SD = 7.9$). Participants had previous experience with virtual avatars from series and movies (N = 13), social media (N = 12), video games (N = 10), commercials (N = 3), internet browsing (N = 3), and work (N = 3). We recruited the participants through the advertising channels of our institution. We excluded participants who had eye surgery, wore glasses with more than one power, or had any eye movement or alignment abnormalities, such as lazy eyes, Strabismus, or Nystagmus. Participants did receive any compensation for their participation.

### 3.2 Study design

The study was designed to be within-subject with three independent variables: *stimulus*, *target direction*, and *gaze validity*. The stimulus consisted of two levels which included (1) a human and (2) a virtual avatar face. Since daily interaction with computers shifts towards virtual environments that include virtual characters [2, 14, 15, 18, 26, 30], i.e., "avatars", we explore human and virtual faces to understand better their influence on guiding visual attention. The target direction consisted of two levels: (1) left – with a target appearing on the left, and (2) right – with a target on the right. The selection of target appearance on the left and right sides was based on the people's scanning and reading patterns on visual interfaces, which typically follow side-wise movements rather than up and down or in-between directions [9]. Lastly, the gaze validity also had two levels: (1) valid – with the eyes of the stimuli moving in the same direction where a target appeared and (2) invalid – with the eyes of the stimuli moving in the opposite direction where a target appeared. We explore gaze validity to investigate distraction and focus introduced by gaze cues often employed to direct visual attention based on the direction of people's gaze, e.g., to allocate users' visual attention to content important to interact with [25]. To explore all levels of independent variables, we created eight experimental conditions (2 stimuli x 2 target directions x 2 gaze validities) (Figure 1). The sequence of eight conditions was counterbalanced using a Balanced Latin square.

At the beginning of each condition, a fixation cross appeared in the middle of the screen, which was replaced by an avatar (300x455 pixels) after 670 ms. The starting avatar's gaze was directed toward the participant. Afterward, the stimulus's gaze changed after 900 ms to either left or right. After 300 ms, a target shaped as a circle (610 pix) appeared at a distance 100 pixels away from the edge of the screen, either to the left or right, depending on the experimental condition. Each trial could therefore be a valid gaze cue, in which the object appears in the gaze direction of the stimulus, or an invalid, in which the object appears in the opposite gaze direction. The consequent condition started in 2000 ms with a fixation cross to regain participants' gaze to the middle of the screen. Participants sat in a chair 60-65 cm away from the screen.

### 3.3 Apparatus

We employed screen-based eye tracker Tobii Pro Nano [1] to assess participants' attention. We chose a screen-based eye tracker to create natural interaction for the participants and remove possible distractions caused by wearable eye-tracking glasses. The study

---

[1] https://www.tobii.com/products/eye-trackers/screen-based/tobii-pro-nano

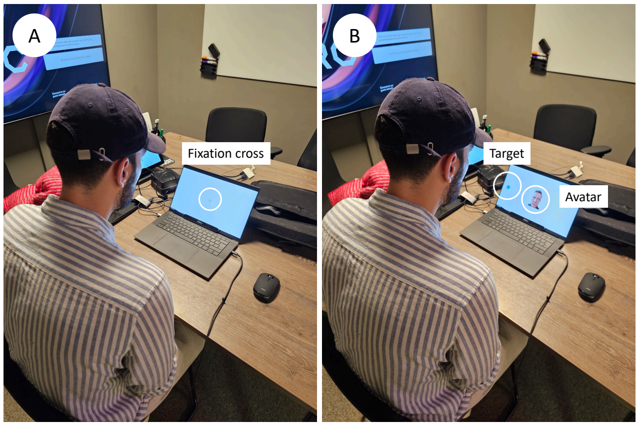

**Figure 2: Study setup: (a) a participant is looking at the fixation cross in the middle of the screen before the target appears, and (b) a participant is looking at the target circle on the left side of the screen with a human avatar in the middle of the screen.**

was conducted using a Dell XPS 15 laptop with a screen size of 15.6 inches and the screen-based eye tracker placed at the bottom of the screen (Figure 2). The study design followed three earlier developed approaches [1, 4, 25], which have been used to measure the gaze cueing effect of participants. The virtual avatar was created using the Unity Asset UMA 2 Multipurpose Avatar[2], which was customized to look similar to the human one. Before the experiment, we defined Areas of Interest (AOI) to create areas for collecting data on participants' gaze. We analyzed the data gathered from the eye tracking session using the Analyze tab of Tobii Pro Lab.

## 3.4 Measurements

To compare participants' attention based on different stimuli, target direction, and gaze validity, we measured the following dependent variables using the AOI tool in Tobii Pro Lab. By creating an AOI, the users ensure all gaze data within the AOI is registered and available for analysis. Metrics such as fixations, saccades, and visits are registered within the AOI. In this study, fixations are the main metrics, and to measure these, Tobii Pro Lab has metrics such as total duration, average duration, time to first fixation, and frequency of fixations. The following measures were gathered in this study:

- **Duration of fixations on a target** (in ms): we measured how long participants fixated on a target in total and on average.
- **Duration of fixations on a stimulus** (in ms): we measured how long participants fixated on a stimulus in total and on average.
- **Frequency of fixations**: we measured how often participants fixated on a target and a stimulus on average.
- **Time to the first fixation on a target** (in ms): we measured the time between the target's appearance and participants' gaze landing on it.

## 3.5 Procedure

After obtaining informed consent, we explained the experiment's goal and allowed the participant to test the setup for familiarization purposes and calibrate the eye tracker. Their task was to look at the middle of the screen, the beginning, and a circular target upon its appearance. At the end of the study, the participants reflected on their experience of the stimuli and the gaze cues. The entire study lasted approximately 15 minutes per participant.

# 4 RESULTS

We discovered that participants' total and average fixation on a target lasted longer with a human image than an avatar stimulus when a target appeared on the right side and when a stimulus' gaze was looking toward the target. Moreover, participants' average fixation was longer on a human than an avatar stimulus with an eye gaze in the opposite direction from a target than toward it. Lastly, participants glanced more often at a stimulus when a stimulus' was looking toward the target, and it took longer for participants to glance at a target for the first time if a stimulus' gaze was not looking toward the target.

## 4.1 Duration of fixations on a target

*4.1.1 Total duration.* We discovered that participants' *total* fixation on a target lasted longer in the presence of a human stimulus ($Md = 1630\ msec, IQR = 305$) than a virtual avatar ($Md = 1529\ msec, IQR = 992$). As for the direction, participants' *total* fixation on a target lasted longer when it appeared on the right ($Md = 1647\ msec, IQR = 262$) than on the left ($Md = 1475\ msec, IQR = 999$). Lastly, participants' *total* fixation on a target lasted longer when the stimulus' eye gaze was looking in a target's direction ($Md = 1658\ msec, IQR = 308$) than in the opposite ($Md = 1496\ msec, IQR = 983$). These findings were supported by the statistically significant main effects for the stimulus type ($F(1, 29) = 165, p < 0.001, \eta^2 = 0.85$), target direction ($F(1, 29) = 99, p < 0.001, \eta^2 = 0.77$), and gaze validity ($F(1, 29) = 96, p < 0.001, \eta^2 = 0.77$). The post-hoc analysis has shown statistically significant differences between all pairs ($p < 0.001$) for all independent variables.

Our statistical analysis revealed three statistically significant interaction effects for stimulus*target direction ($F(1, 29) = 82.7, p < 0.001, \eta^2 = 0.74$), stimulus*gaze validity ($F(1, 29) = 46.7, p < 0.001, \eta^2 = 0.61$), and target direction*gaze validity ($F(1, 29) = 88.7, p < 0.001, \eta^2 = 0.75$). As for the first interaction effect, the post-hoc analysis showed that participants' total fixation on a target lasted longer when it appeared on the right than on the left with a virtual avatar ($p < 0.001$) and a human stimulus ($p < 0.001$). The total fixation on a target was also longer in the presence of a human stimulus than a virtual avatar when a target appeared on the left ($p < 0.001$) and on the right ($p < 0.001$). As for the second interaction effect, participants' total fixation on a target lasted longer with a valid than invalid gaze for a virtual avatar ($p < 0.001$) and a human stimulus ($p < 0.001$). The total fixation on a target was also longer with a human stimulus than in a virtual avatar when a gaze was valid ($p < 0.001$) and invalid ($p < 0.001$). As for the third interaction effect, participants' total fixation on a target lasted longer with a valid gaze than with invalid ($p < 0.001$) when it appeared on the left, but when it appeared on the right, it

---

[2] https://assetstore.unity.com/packages/3d/characters/uma-2-unity-multipurpose-avatar-35611

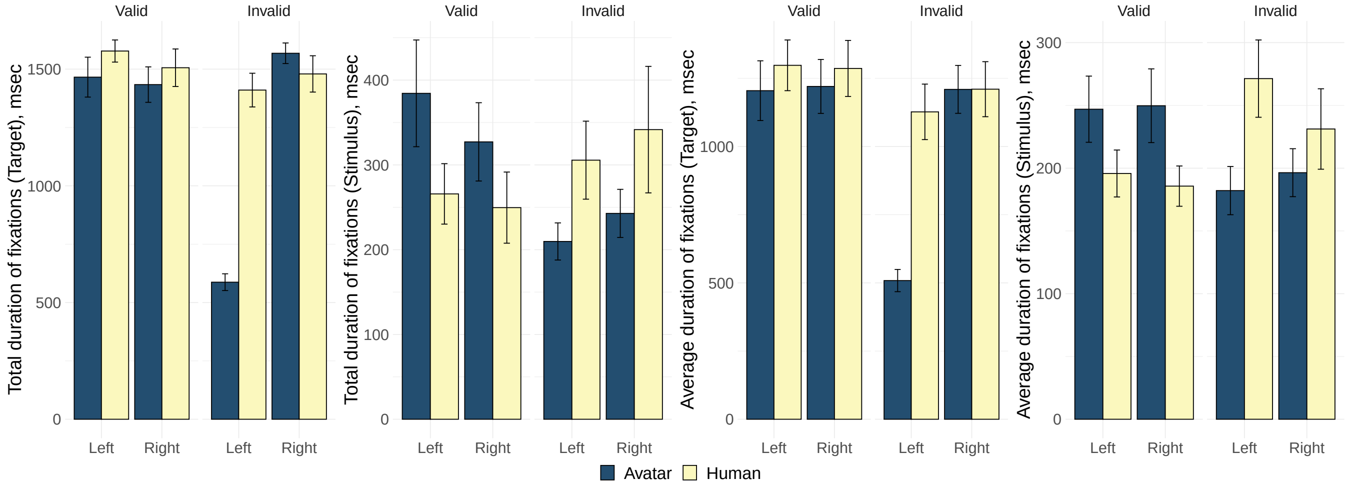

**Figure 3: Overview of the results: total and average fixations on a target and stimulus for the combination of independent variables: stimulus (avatar/human), target direction (left/right), and gaze validity (valid/invalid).**

lasted longer for the invalid than valid gaze ($p < 0.001$). The total fixation on a target was also longer with invalid gaze when a target appeared on the right than left ($p < 0.001$) and with valid gaze when it appeared on the left than right ($p < 0.001$). The remaining pairwise comparisons were not statistically significant ($p > 0.05$).

*4.1.2 Average duration.* Participants' *average* fixation on a target lasted longer in the presence of a human stimulus ($Md = 1486\ msec, IQR = 893$) than a virtual avatar ($Md = 893\ msec, IQR = 1059$). As for the direction, participants' *average* fixation on a target lasted longer when it appeared on the right ($Md = 1516\ msec, IQR = 904$) than on the left ($Md = 858\ msec, IQR = 1077$). Lastly, participants' *average* fixation on a target lasted longer when the stimulus' eye gaze was looking in a target's direction ($Md = 1516\ msec, IQR = 917$) than in the opposite ($Md = 841\ msec, IQR = 1015$). These findings were supported by the statistically significant main effects for the stimulus type ($F(1, 29) = 28, p < 0.001, \eta^2 = 0.49$), target direction ($F(1, 29) = 20, p < 0.001, \eta^2 = 0.41$), and gaze validity ($F(1, 29) = 36, p < 0.001, \eta^2 = 0.56$). The post-hoc analysis has shown statistically significant differences between all pairs ($p < 0.001$) for all independent variables.

Our statistical analysis revealed three statistically significant interaction effects for stimulus*target direction ($F(1, 29) = 16.9, p < 0.001, \eta^2 = 0.37$), stimulus*gaze validity ($F(1, 29) = 7.9, p < 0.001, \eta^2 = 0.21$), target direction*gaze validity ($F(1, 29) = 19, p < 0.001, \eta^2 = 0.4$). As for the first interaction effect, the post-hoc analysis showed that participants' average fixation on a target lasted longer when it appeared on the right than on the left in the presence of a virtual avatar ($p < 0.05$) and a human stimulus ($p < 0.05$). The average fixation on a target was also longer in the presence of a human stimulus than a virtual avatar when a target appeared on the left ($p < 0.05$) and on the right ($p < 0.05$). As for the second interaction effect, participants' average fixation on a target was comparable for all pairs ($p > 0.05$). As for the third interaction effect, participants' average fixation on a target lasted longer with a valid gaze than with invalid ($p < 0.05$) when it appeared on the left, but when it appeared on the right it lasted longer for the invalid than valid gaze ($p < 0.05$). The average fixation on a target was also longer with invalid gaze when a target appeared on the right than left ($p < 0.05$) and with valid gaze when it appeared on the left than right ($p < 0.05$). The remaining pairwise comparisons were not statistically significant ($p > 0.05$). Figure 3 provides a detailed overview of the results.

## 4.2 Duration of fixations on a stimulus

*4.2.1 Total duration.* We discovered that participants' *total* fixation on a stimulus was comparable for a human stimulus ($Md = 221\ msec, IQR = 172$) and a virtual avatar ($Md = 148\ msec, IQR = 236$). As for the direction, participants' *total* fixation on a stimulus was comparable when a target appeared on the right ($Md = 231\ msec, IQR = 168$) and on the left ($Md = 232\ msec, IQR = 159$). Lastly, participants' *total* fixation on a stimulus was comparable when the stimulus' eye gaze was looking in a target's direction ($Md = 226\ msec, IQR = 176$) and in the opposite direction ($Md = 236\ msec, IQR = 149$). These findings were supported by the statistically non-significant main effects for the stimulus type ($F(1, 29) = 1.6, p > 0.05, \eta^2 = 0.05$), target direction ($F(1, 29) = 0.27, p > 0.05, \eta^2 = 0.001$), and gaze validity ($F(1, 29) = 0.11, p > 0.05, \eta^2 = 0.003$). None of the interaction effects were statistically significant ($p > 0.05$).

*4.2.2 Average duration.* We discovered that participants' *average* fixation on a stimulus was comparable to a human stimulus ($Md = 198\ msec, IQR = 129$) and a virtual avatar ($Md = 218\ msec, IQR = 125$). As for the direction, participants' *average* fixation on a stimulus was comparable when a target appeared on the right ($Md = 209\ msec, IQR = 127$) and the left ($Md = 217\ msec, IQR = 130$). Lastly, participants' *average* fixation on a stimulus was comparable when the stimulus' eye gaze was looking in a target's direction ($Md = 197\ msec, IQR = 130$) and in the opposite direction ($Md = 228\ msec, IQR = 119$). These findings were supported by the statistically non-significant main effects for the stimulus

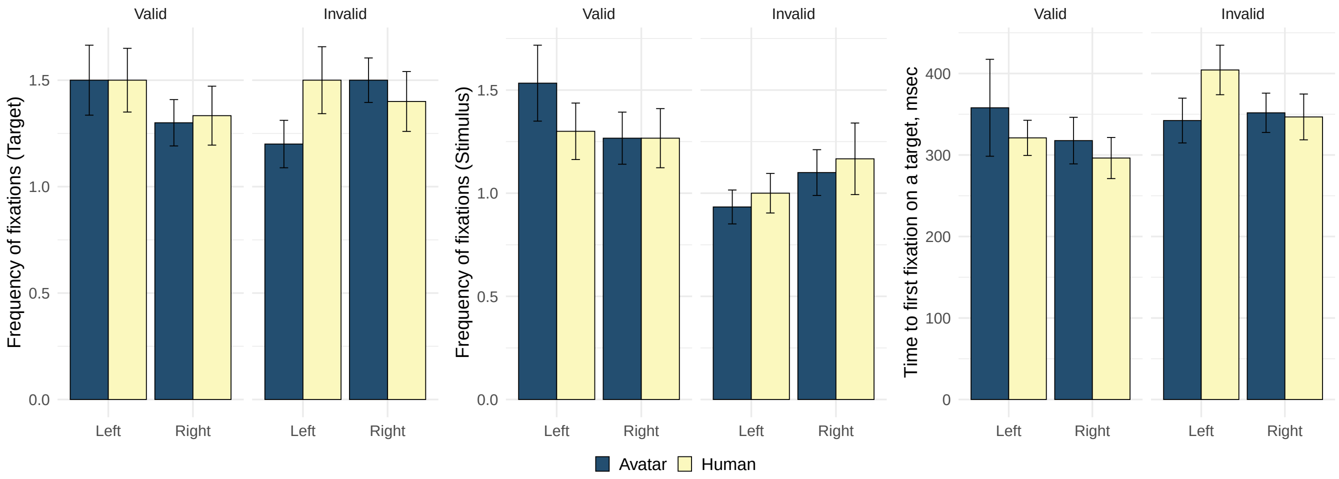

**Figure 4: Overview of the results: frequency of fixations on a target and stimulus, and time to the first fixation on a target for the combination of independent variables: stimulus (avatar/human), target direction (left/right), and gaze validity (valid/invalid).**

type ($F(1, 29) = 0.01, p > 0.05, \eta^2 = 0.0003$), target direction ($F(1, 29) = 0.7, p > 0.05, \eta^2 = 0.024$), and gaze validity ($F(1, 29) = 0.7, p > 0.05, \eta^2 = 0.024$). Additionally, there was one statistically significant interaction effect for stimulus*gaze validity ($F(1, 29) = 8.6, p < 0.001, \eta^2 = 0.23$). The post-hoc analysis has shown that participants' *average* fixation was longer on a human stimulus with an eye gaze in the opposite direction from a target than to the right direction ($p < 0.05$). The remaining interaction effects were not statistically significant ($p > 0.05$). Figure 3 provides a detailed overview of the results.

## 4.3 Frequency of fixations

*4.3.1 On a target.* We discovered that participants glanced at a target a comparable number of times in the presence of a human stimulus ($Md = 1, IQR = 1$) and a virtual avatar ($Md = 1, IQR = 1$). The same applies to the direction left ($Md = 1, IQR = 1$) and right ($Md = 1, IQR = 1$) of the target's appearance, and the correct ($Md = 1, IQR = 1$) and incorrect gaze direction ($Md = 1, IQR = 1$). These findings were supported by the statistically non-significant main effects for the stimulus type ($F(1, 29) = 1.1, p > 0.05, \eta^2 = 0.03$), target direction ($F(1, 29) = 1.2, p > 0.05, \eta^2 = 0.04$), and gaze validity ($F(1, 29) = 1.26, p > 0.05, \eta^2 = 0.04$). None of the interaction effects were statistically significant ($p > 0.05$).

*4.3.2 On a stimulus.* We discovered that participants glanced at a stimulus a comparable number of times in the presence of a human stimulus ($Md = 1, IQR = 1$) and a virtual avatar ($Md = 1, IQR = 1$). The same applies to the direction left ($Md = 1, IQR = 1$) and right ($Md = 1, IQR = 1$) of the target's appearance, and the valid ($Md = 1, IQR = 1$) and invalid gaze direction ($Md = 1, IQR = 1$). However, the main effect for the gaze validity was statistically significant ($F(1, 29) = 0.04, p < 0.05, \eta^2 = 0.14$), indicating that participants glanced more often at a stimulus looking in the direction of target's appearance ($p < 0.05$). The remaining main and interaction effects were not statistically significant ($p > 0.05$). Figure 4 provides a detailed overview of the results.

## 4.4 Time to the first fixation on a target

We discovered that it takes a comparable time to glance at a target after its appearance in the presence of a human stimulus ($Md = 315$ $msec, IQR = 172$) and a virtual avatar ($Md = 300$ $msec, IQR = 157$). The same applies to the direction left ($Md = 317$ $msec, IQR = 174$) and right ($Md = 301$ $msec, IQR = 145$) of the target's appearance, and the valid ($Md = 273$ $msec, IQR = 159$) and invalid gaze direction ($Md = 324$ $msec, IQR = 143$). However, the main effect for the gaze validity was statistically significant ($F(1, 29) = 16.7, p < 0.001, \eta^2 = 0.37$), indicating that it takes longer for participants to glance at a target for the first time if a stimulus' gaze is not looking toward target's appearance ($p < 0.05$). The remaining main and interaction effects were not statistically significant ($p > 0.05$). Figure 4 provides a detailed overview of the results.

## 4.5 Qualitative feedback

When it came to the gaze cues provided by the avatar, eleven (out of 30) participants mentioned that the gaze cue was similar to the human gaze cue, as it was a similar experience following the gaze direction. As P4 noted: *"The animated figure seemed more tired, but it was equally simple to follow the gaze direction of them both."*. Additionally, participants commented that even though the avatar seemed a little more tired, following the gaze direction of both stimuli was as simple. For example, P28 mentioned that *"it felt natural when the avatar glanced at the object and therefore felt natural to follow the gaze direction."* However, other participants commented that *"I thought they were somewhat neutral in relative to one another. They did not stand out against each other. Same kind of facial expressions."* [P7] and *"I thought the avatar had a very unclear facial expression"* [P3]. Two participants were curious about why they were tricked when the gaze was in the wrong direction. For example, some participants noted *"The avatar looked diagonally to the side, which made me want to look there instead. I was more interested in seeing what it was looking at"* [P8] and *"It was harder to* *be "prepared," you always wanted to glance in the direction the person was looking in"* [P20].

Three participants (P2, P7 & P17) underlined that after some time, they stopped focusing as much on their eyes and used their peripheral vision to identify where the object would appear quickly. P7 further explained that at first, they trusted the human stimulus more and were more fooled by it, but after a while, even the trust aspect disappeared, and they learned not to always focus on the eyes. Trust was a factor that many participants stressed in different ways. For example, P19 and P29 thought both stimuli aimed to trick them *"It felt like they were trying to trick me, I was tricked the first time, but after that, I did not look where they looked on purpose. I did not trust them."* [P29] and P23 chose to differentiate the stimuli *"I trusted the human face more when the animated face came I first thought it would look in the right direction. I had more trust in the human face and was tricked more by it."*. This opinion of being tricked by the human stimulus was also shared by P14, who explained it in the following way: *"You look at the eyes more on the human as you recognize it to be more human eyes. With the avatar, it was easier to look away and at the object."*.

## 5 DISCUSSION AND FUTURE WORK

By exploring visual attention guidance using humans and avatars under valid and invalid gaze directions on 2D screens, we have shown conceptual differences and derived a set of design guidelines and practical implications discussed in the following.

### 5.1 Human to focus and avatar to distract

With this experiment, we have shown fundamental differences in guiding visual attention using avatars and humans. We discovered that human gaze cues facilitate focus on a target, while avatars are better at distracting users' attention. This outcome is based on the result that participants' total and average fixation on a target lasted longer with human images than with an avatar stimulus. This applies to the situations when a target appears on the right side relative to the avatar and when a stimulus' gaze is looking in the target's direction, which is in line with previous work on gender differences in gaze cueing [1, 4]. These implications can go beyond the interaction on laptops but also on smartphones. For example, during the interview, participants mentioned Snapchat as a social media platform where they interact with avatars since every person has a small avatar representing themselves. When interacting with other people on Snapchat or browsing the map, one can see other people's avatars. Therefore, participants may deem this an interaction with avatars. Still, it might be a completely different type of interaction compared to watching avatars in animated series/movies or video games. It can provide more or less experience of interacting with avatars.

As for the fixations on a stimulus, our findings indicate that participants fixated on them for a comparable amount of time. This highlights that the focus on a target was stronger and marginally affected by the appearance of stimuli, e.g., participants tended to focus better on human images than virtual avatars. Similarly, fixations on a target were comparable between human images and virtual avatars, indicating one glance on average. This implies that participants' gaze did not tend to jump between a stimulus and a target. However, we observed that participants glanced more often at a stimulus when their gaze was toward a target. Most likely, after quickly and successfully spotting a target, participants were looking for a confirmation of their selection by looking at a stimulus again. Lastly, participants did not feel that the gaze cues provided by the stimuli differed greatly, which is in line with previously mentioned studies [2, 14, 18] on decreased task completion time with avatar gaze cues. Our evaluation aimed to add to the research by taking a well-known gaze cue method and exploring its effect on humans and avatars. Results thus show differences and avatars' gaze cues can allocate visual attention differently to human gaze cues, depending on the situation. Even though previous research, such as [2, 14, 18, 26, 30], have proven that gaze cues from avatars work, this study compares the gaze cues from the two stimuli. It shows that there is no statistically significant difference in task completion. This can lead to designers having avatars as a choice of providing gaze cues, given that it is viable in the specific setting. Therefore, we derived a set of design guidelines, which we list in the following subsection.

As for the target direction, our results indicated that participants fixated longer on targets when they appeared on the right side regardless of stimulus type. This could be explained by the cultural background of our sample, which fully consisted of participants who read and write from left to right. Thus, moving their eyes from left to right could have been a natural and habitual movement, which introduced a directional bias in their spatial cognition [11]. This finding and explanation also align with other results demonstrating that participants fixate more on a target on the left in the presence of a valid gaze, i.e., a stimulus looking toward a target. Still, it has the opposite effect when a target appears on the right. This implies that participants' reading and writing direction possibly overwrites the gaze cues presented by avatars, also known as script directionality effects that arise from left to right reading and writing habits [33].

### 5.2 Design guidelines

Based on the results of our study and the discussion above, we derived the following design guidelines (GL) guiding visual attention on 2D screens using gaze cues with avatars and humans:

- **GL1**: Human images are better suited to facilitate focus on a target and avatars to guide focus away.
- **GL2**: Gaze cues of a stimulus looking toward a target lead to better focus on it.
- **GL3**: Using human images increases the speed of the first fixation.
- **GL4**: Frequency of glances at a target is not affected by the avatar and human images and direction of a target's appearance.
- **GL5**: Target appearance on the right side leads to longer total and average duration of fixations on it independently of stimulus type.
- **GL6**: Target appearance on the *left* combined with stimulus' eyes looking in the target's direction leads to *longer* total and average duration of fixations on it.
- **GL7**: Target appearance on the *right* combined with stimulus' eyes looking in the target's direction leads to *shorter* total and average duration of fixations on it.

## 5.3 Practical implications

Gaze cues with avatars can find use in settings where avatars are used to communicate, such as video games or immersive learning, and not only in 2D interfaces, such as websites and smartphone apps. Being able to use gaze cues in these settings can help in making interactions feel more human-like, but at the same time decrease task-completion times [2, 14, 15, 18] and error rates [2]. With participants sharing that following the gaze direction of the avatar felt natural, this is a good sign that avatars can appear more realistic and engaging with gaze cues. Therefore, this can lead to increased immersiveness when presented in a 3D space, e.g., in virtual reality, increased user learning, and overall enjoyment [27].

Gaze cues are often used in advertisement settings, where the goal is to allocate visual attention to the desired parts of the advertisements, which often is the product or brand [28, 32]. Based on our findings, game designers could, for example, use their avatars in their advertisements and provide gaze cues, increasing exposure to their brand and product. At the same time, using their already branded avatars instead of unknown humans also may increase their branding, as the avatars may pique the interest of the people looking at the advertisements. Similarly, animated movies or series may also use their already-created avatars instead of taking images of unknown humans for marketing purposes. It is also important to distinguish the settings where gaze cues with avatars might produce different results for humans. Suppose avatars are used in settings where they feel misplaced. In that case, they may allocate attention to the avatar instead of in the gaze direction and, therefore, not produce the intended results, which they might do if the setting is relevant. In the after-study interviews, some participants mentioned that they stopped looking at the stimulus's eyes after several trials. Doing the same test twice but with different time pauses would affect the viability of the gaze cues and lead to more participants strictly using their peripheral vision to locate the object instead of using the gaze direction.

Previous work [1, 4] also included a test task meant to serve as a test run and, combined with the instructions, prepare the participants for the test. Due to this test only running eight trials compared to the other studies that had significantly more, in this study, we did not choose to have test tasks to minimize the learning effect on the participants. Instead, the instructions were tested on the pilot study and refined to explain better the test and how the trials would work. As for the first trial of every participant, there was no clear difference in the time to first fixation, and the trials that were randomly ordered as the first trial were evenly distributed across the different eight trials, meaning that roughly the same amount of invalid trials and valid trials were the first trial the participants saw. For future studies, a test task might still be viable to prepare the participants better to avoid large differences in time to the first fixation.

## 5.4 Future work

Further research on whether participants' previous experience with avatars may affect their interaction with them would be an interesting continuation of this line of research. In this study, there was no formal questionnaire or form to measure their previous experience and no way of determining how each media influences the experience gained. Finding a way of ranking experiences and having a tested method to measure experiences would be a valuable contribution. In this study, we used only a male avatar and male human image to provide the gaze cues, leaving the gender aspect out of the scope. However, there might be a correlation between a participant's gender and gaze behaviors [4]. Future work could expand the gender aspect to focus on the stimuli that present the gaze cues to see if there is a difference in how people allocate visual attention based on the gender of the stimuli and if the stimulus type can affect this difference. Moreover, this evaluation focused on the interaction on a single screen and focus on a target without an implicit interaction. Future work can further explore setups with multiple screens and use participants' gazes as input for entertainment, visual attention guidance, or eye-tracking calibration methods [34].

## 6 LIMITATIONS

The experiment presented in this work focused on systematically evaluating three aspects of visual attention guidance on a laptop screen due to the availability of powerful mobile eye-tracking systems. However, we should have investigated other types of devices with 2D user interfaces, e.g., smartphones and tablets, that should be considered in future work. Another limitation of this work is that avatars and targets were presented to participants in isolation from a context, e.g., a website. The presence of other user interface elements would compete for users' attention even more and lead to lower duration and frequency of fixations on a target. However, we aimed to provide an initial empirical evaluation of the effects of avatars and humans on visual attention guidance to create a baseline, and future work should consider more complicated and realistic scenarios for attention guidance. Within the scope of this study, we investigated one representation of an avatar and one representation of a human. Other representations might lead to different results and should be more systematically explored in future work, especially since some avatars might create better emotional connections with users than others, e.g., based on familiarity or cultural background. Moreover, only eye gaze played a role of guidance in this experiment, and adding mimics, e.g., lip or eyebrow movements, is a better way to guide users' attention due to a higher level of avatars' expressiveness.

## 7 CONCLUSION

In this work, we investigated how to direct users' visual attention on 2D screens using gaze stimuli from avatars and humans. We found that participants' overall and average fixations on a target lasted longer when a human than an avatar stimulus was present, the target appeared on the right side, and the stimulus' gaze was directed toward the target. In addition, participants' average fixation on a human than an avatar stimulus was longer when the gaze was directed in the opposite direction of the target than in the direction of the target. Finally, participants glanced more frequently at a stimulus when the stimulus's gaze was directed in the direction of the target's appearance, and it took longer for participants to first glance at a target when the stimulus's gaze was not directed in the direction of the target's appearance.

# ACKNOWLEDGMENTS

We would like to thank all participants who took part in our experiment.

# REFERENCES

[1] N. Alwall, D. Johansson, and S. Hansen. 2010. The gender difference in gaze-cueing: Associations with empathizing and systemizing. *Personality and Individual Differences* 49, 7 (Nov. 2010), 729–732. https://doi.org/10.1016/j.paid.2010.06.016

[2] Sean Andrist, Michael Gleicher, and Bilge Mutlu. 2017. Looking Coordinated: Bidirectional Gaze Mechanisms for Collaborative Interaction with Virtual Characters. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 2571–2582. https://doi.org/10.1145/3025453.3026033

[3] Sean Andrist, Tomislav Pejsa, Bilge Mutlu, and Michael Gleicher. 2012. Designing effective gaze mechanisms for virtual agents. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. Association for Computing Machinery, New York, NY, USA, 705–714. https://doi.org/10.1145/2207676.2207777

[4] Andrew P. Bayliss, Giuseppe di Pellegrino, and Steven P. Tipper. 2005. Sex differences in eye gaze and symbolic cueing of attention. *The Quarterly Journal of Experimental Psychology Section A* 58, 4 (May 2005), 631–650. https://doi.org/10.1080/02724980443000124

[5] Rosy Boardman and Helen Mccormick. 2021. Attention and behaviour on fashion retail websites: an eye-tracking study. *Information Technology & People* 35, 7 (Jan. 2021), 2219–2240. https://doi.org/10.1108/ITP-08-2020-0580

[6] Benjamin T. Carter and Steven G. Luke. 2020. Best practices in eye tracking research. *International Journal of Psychophysiology* 155 (Sept. 2020), 49–62. https://doi.org/10.1016/j.ijpsycho.2020.05.010

[7] Jeanette A. Chacón-Candia, Juan Lupiáñez, Maria Casagrande, and Andrea Marotta. 2023. Eye-Gaze direction triggers a more specific attentional orienting compared to arrows. *PLOS ONE* 18, 1 (Jan. 2023), e0280955. https://doi.org/10.1371/journal.pone.0280955

[8] Robert B. Cialdini and Robert B. Cialdini PH.D, PhD. 1993. *Influence (rev): The Psychology of Persuasion.* HarperCollins. Google-Books-ID: mTYj9XUlYvMC.

[9] Doaa Farouk Badawy Eldesouky. 2013. Visual Hierarchy and Mind Motion in Advertising Design. *Journal of Arts and Humanities* 2, 2 (2013), 148–162. https://doi.org/10.18533/journal.v2i2.78

[10] Sukru Eraslan, Victoria Yaneva, Yeliz Yesilada, and Simon Harper. 2019. Web users with autism: eye tracking evidence for differences. *Behaviour & Information Technology* 38, 7 (July 2019), 678–700. https://doi.org/10.1080/0144929X.2018.1551933

[11] Nafiseh Faghihi and Jyotsna Vaid. 2023. Reading/writing direction as a source of directional bias in spatial cognition: Possible mechanisms and scope. *Psychonomic Bulletin & Review* 30, 3 (2023), 843–862. https://doi.org/10.3758/s13423-022-02239-1

[12] Kassandra Friebe, Sabína Samporová, Kristína Malinovská, and Matej Hoffmann. 2022. Gaze Cueing and the Role of Presence in Human-Robot Interaction. In *Social Robotics (Lecture Notes in Computer Science)*, Filippo Cavallo, John-John Cabibihan, Laura Fiorini, Alessandra Sorrentino, Hongsheng He, Xiaorui Liu, Yoshio Matsumoto, and Shuzhi Sam Ge (Eds.). Springer Nature Switzerland, Cham, 402–414. https://doi.org/10.1007/978-3-031-24667-8_36

[13] Robert M. Joseph, Zachary Fricker, and Brandon Keehn. 2015. Activation of frontoparietal attention networks by non-predictive gaze and arrow cues. *Social Cognitive and Affective Neuroscience* 10, 2 (Feb. 2015), 294–301. https://doi.org/10.1093/scan/nsu054

[14] Mahdi Khoramshahi, Ashwini Shukla, Stéphane Raffard, Benoît G. Bardy, and Aude Billard. 2016. Role of Gaze Cues in Interpersonal Motor Coordination: Towards Higher Affiliation in Human-Robot Interaction. *PLOS ONE* 11, 6 (June 2016), e0156874. https://doi.org/10.1371/journal.pone.0156874

[15] Simon Kimmel, Frederike Jung, Andrii Matviienko, Wilko Heuten, and Susanne Boll. 2023. Let's Face It: Influence of Facial Expressions on Social Presence in Collaborative Virtual Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 429, 16 pages. https://doi.org/10.1145/3544548.3580707

[16] Michael Lankes and Argenis Gomez. 2022. GazeCues: Exploring the Effects of Gaze-based Visual Cues in Virtual Reality Exploration Games. *Proceedings of the ACM on Human-Computer Interaction* 6 (Oct. 2022), 1–25. https://doi.org/10.1145/3549500

[17] Apoorva Rajiv Madipakkam, Gabriele Bellucci, Marcus Rothkirch, and Soyoung Q. Park. 2019. The influence of gaze direction on food preferences. *Scientific Reports* 9, 1 (April 2019), 5604. https://doi.org/10.1038/s41598-019-41815-9

[18] Santiago Martinez, Robin Sloan, Andrea Szymkowiak, and Kenneth Scott-Brown. 2011. Animated Virtual Agents to Cue User Attention. *International Journal On Advances in Intelligent Systems* 4 (Jan. 2011), 299–308.

[19] Takashi Mitsuda, Mio Otani, and Sayana Sugimoto. 2019. Gender and individual differences in cueing effects: Visuospatial attention and object likability. *Attention, Perception, & Psychophysics* 81, 6 (Aug. 2019), 1890–1900. https://doi.org/10.3758/s13414-019-01743-2

[20] Jewoong Moon and Jeeheon Ryu. 2021. The effects of social and cognitive cues on learning comprehension, eye-gaze pattern, and cognitive load in video instruction. *Journal of Computing in Higher Education* 33, 1 (April 2021), 39–63. https://doi.org/10.1007/s12528-020-09255-x

[21] Francisco Muñoz-Leiva, Janet Hernández-Méndez, and Diego Gómez-Carmona. 2019. Measuring advertising effectiveness in Travel 2.0 websites through eye-tracking technology. *Physiology & Behavior* 200 (March 2019), 83–95. https://doi.org/10.1016/j.physbeh.2018.03.002

[22] Xufang Pang, Ying Cao, Rynson W. H. Lau, and Antoni B. Chan. 2016. Directing user attention via visual flow on web designs. *ACM Transactions on Graphics* 35, 6 (Dec. 2016), 240:1–240:11. https://doi.org/10.1145/2980179.2982422

[23] Kara Pernice. 2019. Text Scanning Patterns: Eyetracking Evidence. https://www.nngroup.com/articles/text-scanning-patterns-eyetracking/

[24] Kara Pernice and Jakob Nielsen. 2009. How to Conduct Eyetracking Studies | Nielsen Norman Group Report. https://www.nngroup.com/reports/how-to-conduct-eyetracking-studies/

[25] Qian Qian, Miao Song, and Keizo Shinomori. 2013. Gaze cueing as a function of perceived gaze direction. *Japanese Psychological Research* 55, 3 (2013), 264–272. https://doi.org/10.1111/jpr.12001

[26] Radiah Rivu, Ken Pfeuffer, Philipp Müller, Yomna Abdelrahman, Andreas Bulling, and Florian Alt. 2021. Altering Non-Verbal Cues to Implicitly Direct Attention in Social VR. In *Proceedings of the 2021 ACM Symposium on Spatial User Interaction* (Virtual Event, USA) *(SUI '21)*. Association for Computing Machinery, New York, NY, USA, Article 18, 2 pages. https://doi.org/10.1145/3485279.3485309

[27] K. Ruhland, C. E. Peters, S. Andrist, J. B. Badler, N. I. Badler, M. Gleicher, B. Mutlu, and R. McDonnell. 2015. A Review of Eye Gaze in Virtual Agents, Social Robotics and HCI: Behaviour Generation, User Interaction and Perception. *Computer Graphics Forum* 34, 6 (2015), 299–326. https://doi.org/10.1111/cgf.12603

[28] Pitch Sajjacholapunt and Linden Ball. 2014. The influence of banner advertisements on attention and memory: human faces with averted gaze can enhance advertising effectiveness. *Frontiers in Psychology* 5 (2014). https://www.frontiersin.org/articles/10.3389/fpsyg.2014.00166

[29] Graham G. Scott and Christopher J. Hand. 2016. Motivation determines Facebook viewing strategy: An eye movement analysis. *Computers in Human Behavior* 56 (March 2016), 267–280. https://doi.org/10.1016/j.chb.2015.11.029

[30] William Steptoe, Robin Wolff, Alessio Murgia, Estefania Guimaraes, John Rae, Paul Sharkey, David Roberts, and Anthony Steed. 2008. Eye-Tracking for Avatar Eye-Gaze and Interactional Analysis in Immersive Collaborative Virtual Environments. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work* (San Diego, CA, USA) *(CSCW '08)*. Association for Computing Machinery, New York, NY, USA, 197–200. https://doi.org/10.1145/1460563.1460593

[31] David Stevens. 2022. Optimizing Visual Cues in Educational Software. In *Augmented Cognition: 16th International Conference, AC 2022, Held as Part of the 24th HCI International Conference, HCII 2022, Virtual Event, June 26 – July 1, 2022, Proceedings*. Springer-Verlag, Berlin, Heidelberg, 287–303. https://doi.org/10.1007/978-3-031-05457-0_23

[32] Rita Ngoc To and Vanessa M Patrick. 2021. How the Eyes Connect to the Heart: The Influence of Eye Gaze Direction on Advertising Effectiveness. *Journal of Consumer Research* 48, 1 (June 2021), 123–146. https://doi.org/10.1093/jcr/ucaa063

[33] Jyotsna Vaid. 2011. Asymmetries in representational drawing: Alternatives to a laterality account. *Spatial dimensions of social thought* 18 (2011), 231–255. https://doi.org/10.1515/9783110254310.231

[34] Simon Voelker, Andrii Matviienko, Johannes Schöning, and Jan Borchers. 2015. Combining Direct and Indirect Touch Input for Interactive Workspaces Using Gaze Input. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction* (Los Angeles, California, USA) *(SUI '15)*. Association for Computing Machinery, New York, NY, USA, 79–88. https://doi.org/10.1145/2788940.2788949

[35] Jiahui Wang, Pavlo Antonenko, Mehmet Celepkolu, Yerika Jimenez, Ethan Fieldman, and Ashley Fieldman. 2019. Exploring Relationships Between Eye Tracking and Traditional Usability Testing Data. *International Journal of Human–Computer Interaction* 35, 6 (April 2019), 483–494. https://doi.org/10.1080/10447318.2018.1464776

[36] Qiang Yang, Yuanjian Zhou, Yushi Jiang, and Jiale Huo. 2021. How to overcome online banner blindness? A study on the effects of creativity. *Journal of Research in Interactive Marketing* 15, 2 (Jan. 2021), 223–242. https://doi.org/10.1108/JRIM-12-2019-0212

[37] Yanxia Zhang, Ken Pfeuffer, Ming Ki Chong, Jason Alexander, Andreas Bulling, and Hans Gellersen. 2017. Look together: using gaze for assisting co-located collaborative search. *Personal and Ubiquitous Computing* 21 (2017), 173–186. https://doi.org/10.1007/s00779-016-0969-x