The Stare-in-the-Crowd Effect in Virtual Reality

Pierre Raimbaud, Alberto Jovane, Katja Zibrek, Claudio Pacchierotti, Marc Christie, Ludovic Hoyet, Julien Pettre, Anne-Hélène Olivier

▶ **To cite this version:**

Pierre Raimbaud, Alberto Jovane, Katja Zibrek, Claudio Pacchierotti, Marc Christie, et al.. The Stare-in-the-Crowd Effect in Virtual Reality. VR 2022 - IEEE Conference on Virtual Reality and 3D User Interfaces, Mar 2022, Christchurch, New Zealand. pp.281-290, 10.1109/VR51125.2022.00047 . hal-03581916v1

HAL Id: hal-03581916

https://hal.science/hal-03581916v1

Submitted on 20 Feb 2022 (v1), last revised 23 Mar 2023 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The Stare-in-the-Crowd Effect in Virtual Reality

Pierre Raimbaud*
Inria, Univ Rennes, CNRS, IRISA - France

Alberto Jovane
Inria, Univ Rennes, CNRS, IRISA - France

Katja Zibrek
Inria, Univ Rennes, CNRS, IRISA - France

Claudio Pacchierotti
CNRS, Univ Rennes, Inria, IRISA - France

Marc Christie
Univ Rennes, CNRS, IRISA, Inria - France

Ludovic Hoyet
Inria, Univ Rennes, CNRS, IRISA - France

Julien Pettré
Inria, Univ Rennes, CNRS, IRISA - France

Anne-Hélène Olivier
Univ Rennes, Inria, CNRS, IRISA, M2S - France

Figure 1: The stare-in-the-crowd effect describes the tendency of humans in noticing and observing, more frequently and for longer time, gazes oriented toward them (directed gaze) than gazes directed elsewhere (averted gaze). This work analyses the presence of such an effect in VR and its relationship with social anxiety levels. The figure above shows an example of the user's view during our experiment. All agents, except the woman in the front row wearing a black jacket, have their gaze averted.

ABSTRACT

Nonverbal cues are paramount in real-world interactions. Among these cues, gaze has received much attention in the literature. In particular, previous work has shown a search asymmetry between directed and averted gaze towards the observer using photographic stimuli, with faster detection and longer fixation towards directed gaze by the observer. This is known as the stare-in-the-crowd effect. In this study, we investigate whether stare-in-the crowd effect is preserved in Virtual Reality (VR). To this end, we designed a within-subject experiment where 30 human users were immersed in a virtual environment in front of an audience of 11 virtual agents following 4 different gaze behaviours. We analysed the user's gaze behaviour when observing the audience, computing fixations and dwell time. We also collected the users' social anxiety score using a post-experiment questionnaire to control for some potential influencing factors. Results show that the stare-in-the-crowd effect is preserved in VR, as demonstrated by the significant differences between gaze behaviours, similarly to what was found in previous studies using photographic stimuli. Additionally, we found a negative correlation between dwell time towards directed gazes and users' social anxiety scores. Such results are encouraging for the development of expressive and reactive virtual humans, which can be animated to express natural interactive behaviour.

Index Terms: Human-centered computing—Visualization—Visualization design and evaluation methods; Human-centered computing—User studies; Applied computing—Law, social and behavioral sciences;

1 INTRODUCTION

Immersion in virtual worlds populated by autonomous humans necessarily raises the question of their interactions with the user. It is now widely demonstrated that the realism of the immersive world and the level of presence felt by the user is, at least partially, correlated to the interaction capabilities of the virtual humans [7, 45, 50].

The field covered by these interaction capacities and by their different modalities is very broad, therefore, we focus here on the non-verbal interaction capacities of virtual humans, to communicate a message by their attitudes, their gestures, their positions in relation to a user. More specifically, we are interested in the gaze behaviour of virtual humans and their perception by the users (see for example the overview of eye-gaze behaviour for virtual agents in [44]).

In the current research, we are interested in the initiation of an interaction between virtual humans and a user, and we ask whether the virtual humans' gaze behaviour can be useful in initiating it. Can the gaze trigger a mutual gaze between the user and the virtual human? Can it focus the user's attention on it? Can this constitute the starting of an interaction through nonverbal communication?

These issues have been addressed outside the field of Virtual Reality (VR) through a protocol called the "stare-in-the-crowd effect" [51], which demonstrated that when multiple faces are exposed to a subject during a visual search task, the detection of faces whose gaze is directed towards the subject is faster (vs. averted ones). It has also been shown that in free visual tasks, visual attention is affected by the presence of directed gaze among averted ones [15].

While these observations were made primarily on the basis of photographic stimuli, in our work we question whether such effects are maintained when a user is immersed in VR. To answer that, we replicate the experiment of Crehan et al. [15], adapting it to VR. We aim at analysing whether the presence of the stare-in-the-crowd effect is retained when observing a virtual crowd, as well as how it is affected by the self-assessed social anxiety levels of the users.

In Section 2, we give an overview of the related literature, and situate our work accordingly. In Section 3, we introduce the objective and hypotheses about our VR replication of the stare-in-the-crowd effect. Section 4 details the experiment, and Section 5 our results. We finally discuss them in Section 6 before concluding in Section 7.

2 RELATED WORK

2.1 Nonverbal communication behaviours

By definition, nonverbal communication refers to any interpersonal exchange that does not use spoken language. It relies on different

*Corresponding author - email: pierre.raimbaud@inria.fr

interaction aspects, notably described in the studies of proxemics and kinesics, where posture and motion of the head, body, and limbs, and gaze behaviours are relevant cues of investigation [23]. Since nonverbal cues are paramount for humans in their daily social interactions [10], previous studies have investigated (i) how to reproduce these cues in virtual environments, to make users interact with virtual agents [5,40,48], and (ii) to what extent effects induced by nonverbal communication cues could be reproduced in VR [2,4,9,34].

When assessing VR users' social behaviours, it has been shown that the design of nonverbal behavioural cues of virtual characters can influence users' feelings in such populated environments. To measure these responses, estimations of presence and engagement have been used as an evaluation of the naturalness of users' interactions [1,24]. In social VR, the design of such cues also contributes to this naturalness. In this respect, Li et al. [29] evaluated the presence, the interaction quality and its social meaning as felt by users, in a photo sharing task. They showed that VR users exhibited social interactions close to real-life behaviour, and that such behaviours were improved compared to only video-based exchange environments. Similarly, Roth et al. [43] showed in a conversational task situation the importance of designing realistic gaze cues to improve presence.

Moreover, authors have shown the ecological validity of VR to reproduce real-life nonverbal communication human behaviours [39]. Iachini et al. [26] showed that proximity behaviour to virtual agents in VR resembles the proximity people exhibit in the real-world. For gaze behaviours in dyadic interactions, Bailenson et al. [3] showed the preservation of the equilibrium between mutual gaze and personal space distance in VR. Additionally, Garau et al. [20] showed the effect of an inferred-gaze model on perceived quality of communication in VR, compared to a random-gaze model. In line with this, Nummenmaa et al. [35] showed the importance of VR users' interpretation of virtual agents' gaze cues in order to avoid collisions when navigating towards another them. For user-virtual agent interactions in the context of a crowd, Narang et al. [34] also confirmed the importance of the modelling of gaze interactions, reflected by an increase of the believability of the interaction when comparing it with and without using gaze models in the virtual environment.

To conclude, previous studies have shown the importance of nonverbal communication during interactions between a user and virtual agents, where VR is able to preserve real world behaviour, as assessed by the social behaviours of VR users among a virtual audience [26,35] and by users self-assessments such as presence and engagement [11,21]. In this paper, we are notably interested in gaze behaviour in the frame of its ability to initiate an interaction [33], which has received little attention in the VR literature. We will describe in the next section the so-called stare-in-the-crowd effect.

2.2 The stare-in-the-crowd effect

The stare-in-the-crowd effect is a gaze behaviour effect that reflects the existence of a search asymmetry between directed and averted gazes when users face a crowd: directed gazes are detected faster than averted ones and cause more frequent and longer fixations [51]. Won Grunau and Anston [51] showed this effect through the evaluation of a visual search task, i.e., detection of an averted gaze onto directed ones and vice versa, using photographic stimuli. Other previous work showed it with similar experimental designs [17,42]. Additionally, the extendability of this effect has been evaluated using other stimuli. In this regard, Colombatto et al. [13] showed that 3D geometric shapes that appear to be facing users (cones) activate a stare-in-the-crowd effect, and that this was caused by their orientation and not by asymmetry nor contrast differences. They concluded that this effect is not necessarily due to a particular saliency of the eyes but rather due to the intentionality conveyed by the stimulation.

It should also be mentioned that some studies have mitigated the existence of the stare-in-the-crowd effect, notably by refuting the fact that this effect occurs in any configuration [14,37,38]. Palanica et al. [38] found that this effect was strongly modulated by gazes positions in the crowd and by interactions of target position and gaze direction. It was also demonstrated that the nature of the task is relevant [37], i.e., between a visual search task where it is asked to detect the specific gaze and report its position in the crowd and one with detection only. For both tasks, a stare-in-the-crowd effect was found but not on the same aspects (accuracy vs. rapidity), while the effect of gaze positions in the crowd on the results was still present.

In this regard, to our knowledge, Cooper et al. [14] found the most contrasting results in the literature regarding the stare-in-the-crowd effect. They focused on the original experimental design of Grunau and Anston [51], addressing the study of the expected asymmetry about users' reaction time in favour of direct gazes. First, they succeeded in replicating Won Grunau and Anston's results [51], using photographs– here with heads only. Then, their other experiments considered other gaze condition comparisons, notably with several kinds of averted gazes: these supported the absence of the stare-in-the-crowd effect as defined previously [51]. As a result, this study can be considered for future studies as a caveat in favour of (i) the control on averted gaze stimuli, and against (ii) the use of the visual search paradigm to study the stare-in-the-crowd effect and (iii) the evaluation of the effect only over its rapidity asymmetry aspect.

Following Cooper et al. recommendations, Crehan et al.'s study is therefore of the utmost importance, since it proposes a new evaluation paradigm [15]: the visual search task is replaced by an observation task, and the effect is measured through eye-tracking. With this paradigm and its associated metrics such as total gazing time – dwell time, and first fixation time and duration, they also observed the stare-in-the-crowd effect, still using photographs – here with complete bodies. Moreover, they studied dynamic conditions, where gazes changed from averted to direct ones and vice-versa, replicating some natural eye-gaze interactions: *being caught staring at someone* and *catching someone else staring*. They found that these dynamic conditions affect user gaze behaviour similarly to directed gazes.

2.3 Social anxiety and gaze behaviour

Social anxiety is related to discomfort and avoidance of social situations due to one's fear of negative evaluation from the side of others [12]. Such anxiety can be reflected by body cues such as heart rate increase, both in real-life situations [41] and virtual ones [27]. Another common trait of socially anxious people is a high sensibility to eye-contact, particularly elevated fear of direct gaze signal and avoidance of the eye-region of others [6]. This trait is pronounced for faces with an angry expression, as this signifies negative evaluation [25]. For neutral expressions, the fear can also be elevated when the gazes are averted, since this could potentially signify that a person is disinterested in them [54]. In summary, the tendency of socially anxious people to avoid eye-contact could be indicated either by directed or averted gaze of the other, depending on whether it is signalling a negative response or not. Such behaviours have been also shown in virtual reality with groups of virtual agents [52]. In this regard, Lange and Pauli [28] found that for neutral conditions, socially anxious individuals gaze less at the head of the agent, confirming the expected avoidance of gaze contact for virtual humans.

3 Objective and hypotheses

In the present study, we aim at investigating the perception of nonverbal cues when a user is immersed in a virtual environment populated with virtual agents. Our main objective is to study the reaction of users, through their gaze behaviour, when facing a virtual crowd where agents can either look at them or look away.

Previous studies using eye-tracking investigated user's gaze when observing photographs depicting a seated audience. They showed users' preference for gazing at individual subjects in these photographs, whose gaze was directed towards them rather than averted from them, also called the stare-in-the-crowd effect [51].

According to the literature, VR can be used to depict social interactions with user's behaviours that are close to real-life ones. We are thus interested in the presence of this effect in VR – an environment more adapted to natural human interactions than photographs.

Towards this objective, we propose two hypotheses, H1 and H2.

First, we expect that we will observe the same effect as reported in Crehan et al. [15] using a series of photographs, but in VR.

- **H1**: The stare-in-the-crowd effect is preserved with virtual agents in VR.

This means that eye-tracking data will show more saliency characteristics (number of fixations, gaze duration) towards the agent who is directing its gaze towards the user as opposed to when the agent is not looking at the user. Moreover, we also expect the same effect comparing the static averted condition to each dynamic one, i.e., during the phenomena *being caught staring* and *catching someone else staring*. However, for these gazing conditions we expect a lower magnitude of effect than for the static directed gaze one, since the time when the agent is looking at the user is shorter. Finally, we are also interested in the comparison between the behaviour of the user in the dynamic conditions as opposed to static directed one.

Moreover, it has been shown previously that social anxiety influences VR users' gaze behaviours towards a virtual crowd, in a similar way to when interacting with humans in physical reality [28, 55]. Indeed, a higher social anxiety is typically correlated with a lower rate of mutual eye contact towards directed gazes than in the case of socially non-anxious individuals [6, 46]. Therefore, we expect that:

- **H2**: There will be a negative correlation between the time spent gazing towards the agents who are staring at the user and the user's level of social anxiety.

This suggests a possibility that the stare-in-the-crowd effect will depend on the amount of socially anxious individuals in our test sample. With many users scoring high on social anxiety this effect could disappear completely, thus, it is relevant to explore this relationship. It is also important to note that in some cases lack of gaze towards a socially anxious individual can be more frightening, as it can signal disinterest. However, we created the experimental conditions where the context of the averted gaze would not be interpreted like this.

4 Experiment

4.1 Overview

To study the stare-in-the crowd effect in VR, we designed an experiment inspired by Crehan et al. [15], which demonstrated the presence of this effect using photographs. In our experiment, the user is asked to observe a virtual crowd where the gazes of the virtual agents are manipulated according to a series of target conditions/behaviours, similarly to Crehan et al. [15]. These crowd gaze conditions are:

- Averted - **A**: no virtual agent looks towards the human user during the observation task (see Fig. 2.1);
- Directed - **D**: one virtual agent, referred to as the "active agent", stares at the user at the beginning of the observation task and will keep staring at him or her until the end of the task, while no other virtual agent stares at the user (see Fig. 2.2);
- Averted-then-Directed - **AD**: no virtual agent looks towards the user at the beginning of the observation task, but the active agent will start staring at the user once looked at and will continue to stare until the end of the task (see Fig. 2.3);
- Directed-then-Averted - **DA**: the active agent stares at the user at the beginning of the observation task, but will stop once looked at, while no other virtual agent staring (see Fig. 2.4).

Examples of such gaze behaviours in our VR implementation can be seen in the supplementary video.

We asked users to observe the virtual crowd, without telling them to actively searching for directed or averted gazes. Such indications

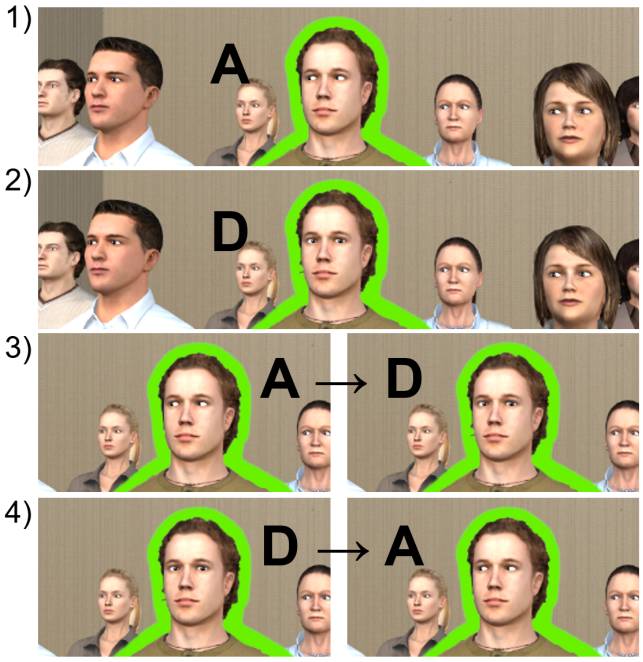

Figure 2: Our four crowd gaze conditions (active agent in green): 1) averted gaze - **A**, 2) directed gaze - **D**, 3) averted-then-directed gaze - **AD**, and 4) directed-then-averted gaze - **DA**. See details in Sec. 4.1.

are different with respect to some previous studies [13, 17, 42], but consistent with Crehan et al. [15, 16]. In line with Crehan et al. [15], we also propose to use an eye-tracking system to evaluate the users' gaze behaviours instead of using a search task, which would be less natural. However, opposite to previous studies [13–17, 19, 42], we use a crowd of virtual agents in VR as visual stimuli (see Fig. 3).

4.2 Virtual environment and stimuli creation

The virtual environment we used here, shown in Fig. 3, was created using Unity 2021.2.0b9. It is composed of a room, resembling a classroom or a conference room, equipped with standard pieces of furniture as well as individual chairs placed on a wooden stage. Virtual agents (our virtual crowd) are seated on these chairs, like an audience, 1m away from the user at the minimum. All virtual agents are clearly visible to the user, without any occlusion between their heads. The wooden stage hides part of the virtual agents' bodies, so as to make the user focus on their faces. Similarly to the photographic stimuli used in Crehan et al. [15], the virtual audience was slightly (10°) oriented to the right, as well as the user (20°). Moreover, the user was placed slightly on the right of the virtual crowd. Such position/orientation choice was chosen for two main reasons: (i) to have all the virtual characters in the user's initial field of view, since they appear at real scale (1:1)); and (ii) to allow virtual agents to look towards the user's position without needing to rotate their head, but only their eyes, while maintaining a natural gaze behaviour (e.g., horizontally rotating the eyes a maximum of 30° with respect to the head). These two aspects ensured that all virtual agents could be easily viewed, and that eyes orientation would be the main difference between them, with different gaze behaviours but similar head orientation, thus avoiding such kind of bias [32].

We considered eleven virtual agents' models from the Microsoft RocketBox adult avatars collection [22], including six females and five males. Fig. 3 shows this virtual audience from top and from the user's point of view. Additionally, we placed another male model in front of the crowd, as if he was giving a presentation to them. However, no speech could be heard by the user, it was only to provide a social setting, and to justify why the crowd was looking towards a common point away from the user. To increase the naturalness of

Figure 3: Virtual scene; the user faces eleven agents listening to a speaker standing behind the user. The inset shows the user's view point during the observation task. Active agents are noted by red dots

agents' behaviours, we applied simple blinking animation on their eyes. Then, a specific gaze behaviour was chosen according to the condition at hand, A, D, AD, or DA, as described in Sec. 4.1.

The virtual agent staring at the users, referred to as the "active agent", is chosen randomly among nine of the eleven agents of the crowd. These nine agents are highlighted with red dots in Fig. 3. This choice was driven by the need to have a balanced distribution of active gazing agents across the user's field of view, as suggested in [17, 37, 38] to test any potential position effects on the results.

It should be noted that for coherence with the other conditions and to enable a consistent comparison of our metrics (see Sec. 4.6), an active agent is also chosen in condition A (no agent looks at the user), although it does not behave differently to the rest of the crowd.

For agents' gaze behaviours we built a gaze mechanism, favouring eye rotations over head and torso rotations, while providing realistic results - e.g., the maximum angle of eye rotation was 30°. This way we could create realistic eye gaze for all positions and conditions.

Finally, in conditions AD and DA, where the agent's gaze dynamic behaviour (from averted to directed or vice-versa) is triggered by the user, we introduced a time limit as suggested by Crehan et al. [15]. If the user has not looked at the target agent within half of the total trial time, the agent's gaze changes anyway, without waiting for the user's gaze. Following Crehan et al. [15], each trial repetition (i.e., the user looking at the crowd) lasted 16 seconds. After this time, the environment fades out and fade in again to the same scene but featuring a new gazing behaviour and active agent (see Sec. 4.5).

4.3 Participants and Apparatus

30 participants (8 females, 22 males; age: aver. 30, SD: 9.5; VR experience from 1 to 5: aver. 3.4, SD: 1.4; computer games experience from 1 to 5: aver. 3.5, SD: 1.5) took part in our experiment, all with normal or corrected-to-normal vision. They voluntarily participate in the experiment and received no compensation for it. The study complied with the Declaration of Helsinki and was approved by local ethical committee (COERLE). Participants were asked to seat on a standard chair throughout the whole experiment, and to wear the VR head-mounted display FOVE, which has an embedded eye-tracking system. Its field of view for a user is 100°, as well as the one of eye-tracking. Its advertised spatial tracking accuracy is less than 1°, and its maximum eye-tracking sampling rate is 120 Hz.

4.4 Data collection

We collected two types of data: (i) continuous user's gaze behaviour during the VR experience, and (ii) social personality data after it.

For (i), gaze behaviour was collected using the embedded eye-tracking system of the VR headset. At each frame, the user's gaze information was logged along with the timestamp and the current gaze condition of the virtual crowd (A, D, AD, or DA). This gaze information was indicating the presence or the absence of a hit on the head of the "active agent", computed using the 2D screen position of the VR user's gaze and the current 2D scene viewed by the user.

For (ii), information about users' social anxiety was collected after the experiment through a questionnaire. We used the standardised questionnaire based on the Liebowitz Social Anxiety Scale [30]. This one allows for the evaluation of social anxiety through self-estimation of the levels of fear and avoidance of a person in determined social situations. A score can be computed from the answers, ranging from 0 (not socially anxious) to 144 (very socially anxious).

4.5 Experimental procedure

First, an informative document about the study was given to the users, along with the informed consent form and oral explanations to answer any questions. Once ready, users were seated on a chair and equipped with the FOVE headset. A calibration of the eye-tracking system was performed to ensure the quality of gaze data collection.

Then, the users were immersed in our virtual environment for a brief training phase, where they had time to familiarise with the environment and setup. During this phase, all agents of the virtual crowd were looking at the virtual speaker, were not changing their gazing behaviour over time, and random agents would be blinking in the crowd. Users were free to look both at the crowd and behind them to see the virtual speaker – which was not talking, to understand the context of the scene. It was explained to them that their task would be to face and observe the virtual audience, and to not look at the virtual speaker after the training phase. No information about gazing behaviours or any other specific tasks to complete were provided.

After this training phase, users were asked to perform 72 trials of this observation task, each lasting 16 seconds. All users were exposed to the same trials i.e., all the tested conditions described in Sec. 4.1. Each combination of "gaze condition/behaviour" per "active virtual agent positioning" was shown twice to each user, leading to: 4 gaze behaviours $\times$ 9 possible active agents $\times$ 2 repetitions = 72 trials in total. In order to make it possible for the user to rest during the experiment, the trials were ordered in 3 blocks, with equal number of gaze conditions presented in each block of 24 trials, as well as the distribution of the active virtual agent. Order of active agents was randomised inside each block. In averted conditions, an agent was chosen randomly and the position of these agents was balanced with the agents in the other conditions, which all include a directed gaze. Additionally, virtual agents' models were randomly switched between all eleven positions, so that the appearance of the models would not influence the results. A 3-seconds black screen was displayed to the users between each trial. During this pause, users were asked to re-position their head and gaze orientation towards the top-centre of the screen, by looking at a small geometric shape. This was done to ensure the same initial point for the user's gaze at each trial. Users were notified that the trials would be divided into three blocks of 24, so as to allow them to rest and remove the headset between each block to minimise fatigue. In addition, such breaks were also used to re-calibrate the eye-tracking system to ensure data quality. If needed, users could also stop within a block.

Finally, users were asked to fill a post-experiment questionnaire with the social anxiety questions, along with demographic ones (age, gender, experience with VR and games) and a free comment section.

4.6 Metrics

From the eye-tracking collected data, we computed different metrics related to the users' gaze towards the active agent of the crowd. Gaze activity was split between saccades when such activity was shorter than 150 ms, and fixations when it was longer [31,53]. For each trial, we considered the following metrics in line with Crehan et al. [15]

- Dwell time: the total time spent looking at the active virtual agent;
- Fixations count: the total number of fixations on the active virtual agent;
- First fixation time: the time of the first fixation on the active virtual agent, counted from the beginning of the trial;
- First fixation duration: the length of the first fixation;
- Second fixation time: the time of the second fixation on the active virtual agent, counted from the beginning of the trial;
- Second fixation duration: the length of the second fixation.

All the above metrics are used to identify the stare-in-the-crowd effect, particularly the dwell time and fixation count metrics that are computed even in absence of multiple fixations on the active agent. The analysis of first and second fixations are also important to better understand user's gaze behaviours, even though they are not always present in stare-in-the-crowd related studies. But they are particularly relevant for the dynamic conditions that we included here, where the user's first fixation on the active agent triggers the change in its gaze behaviour (from averted to directed or vice-versa).

5 Results and Discussion

5.1 Gaze behaviours

According to our objectives and hypotheses, we focused on five comparisons, related to three cases: (1) the stare-in-the-crowd effect in static conditions, (2) *catching someone else staring* and (3) *being caught staring* phenomena, in line with Crehan et al. [15]. For (1), we compared the averted to the directed gaze conditions – A vs. D. Then, we compared each static condition with each dynamic. For (2), averted versus averted-then-directed – A vs. AD, and directed versus averted-then-directed – D vs. AD. For (3), the averted versus directed-then-averted – A vs DA, and directed versus directed-then-averted – D vs. DA. For pairwise comparisons, we ran dependent paired samples t-tests on the six metrics we described in Sec. 4.6 as continuous variables. Such tests guarantee conservative results in the comparison between different gaze conditions. The normal distribution assumption was verified for 25 of our 30 dependent paired samples when running a Shapiro-Wilk test: we ran Student's t-tests for these samples, and Wilcoxon signed rank tests for the remaining ones. Due to our multiple comparison design, we conducted a Bonferroni correction which changed our target significance level from α=0.05 to α=0.00166.

Results are shown in Tables 1 to 5. For each metric, they contain the means and standard deviations, along with significance level, plus statistics and effect size (both when doing Student's t-test). They are shown by comparison of pairs, in Table 1 for conditions A vs. D, Table 4 for A vs. AD, Table 5 for D vs. AD, Table 2 for A vs. DA, and Table 3 for D vs. DA. These results are based on the averages obtained by each user *across all trials that share the same gazing conditions regardless of position*, i.e., 18 in total for each condition. In these tables, a symbol * indicates a p-value $<$0.00166, ** a p-value $<$0.00033, and *** a p-value $<$0.00003.

Comparison A vs D interpretation.

As shown in Table 1, p-values from the metrics *dwell time, fixation count, first and second fixation durations* were all significant, with higher values on the directed condition, which are all **indicators of the presence of a stare-in-the-crowd effect**. We also expected users to spot the active agent in the directed gaze condition sooner, which should be reflected through significantly earlier first fixation time. Such results have been reported and used to confirm the presence of a stare-in-the-crowd effect in previous studies with drawing or photographic stimuli [42, 51]. However, in our experiment, *first fixation time* results do not reveal such a significant difference. For this metric, we discuss our results later in this section (**see Active agent's position effect and Table. 7 with its analysis**). Based on the expectations of the stare-in-the-crowd effect, our results nonetheless show a significantly earlier *second fixation time* on the directed condition compared to the averted one, following the trend expected for the first fixation time. Fig. 4 .1) summarises the comparison between the results on averted and directed conditions and its interpretation for the stare-in-the-crowd effect.

Table 1: Gaze metrics results - comparison of A vs. D conditions

	Averted	Directed			
Metric	Mean (SD)	Mean (SD)	p-value	t	η_p^2
Dwell time	504 (175)	1570 (864)	$<$0.00001 ***	-6.75	0.61
Fixation count	1.15 (0.29)	2.35 (1.03)	$<$0.00001 ***	-6.45	0.59
1st fix. duration	332 (77)	552 (185)	$<$0.00001 ***	-5.61	0.52
1st fix. time	5173 (1213)	4969 (1402)	0.53119	0.63	0.014
2nd fix. duration	407 (282)	554 (214)	0.00158 *	wilc.	wilc.
2nd fix. time	8602 (1395)	6861 (1785)	0.00031 **	4.09	0.37

Time and duration in ms.

Table 2: Gaze metrics results - comparison of A vs. DA conditions

	A	DA			
Metric	Mean (SD)	Mean (SD)	p-value	t	η_p^2
Dwell time	504 (175)	808 (363)	0.00005 **	-4.78	0.44
Fixation count	1.15 (0.29)	1.56 (0.56)	0.00015 **	-4.37	0.40
1st fix. duration	332 (77)	483 (165)	0.00003 ***	-4.92	0.45
1st fix. time	5173 (1213)	4847 (1307)	0.32902	0.99	0.03
2nd fix. duration	407 (282)	374 (95)	0.34921	wilc.	wilc.
2nd fix. time	8602 (1395)	7773 (1724)	0.05175	2.03	0.12

Time and duration in ms.

When comparing with Crehan et al. [15], we found the same results on all our metrics, except for the significantly longer duration for the first fixation in the directed condition in our experiment. Nonetheless, this result is in line with other previous studies [42, 51] and the stare-in-the-crowd effect by definition. In addition, it could be explained by a stronger effect of VR to capture attention with directed gazes, as suggested by our larger effect size results for the other metrics, compared to Crehan et al.'s ones [15].

Comparison A vs DA interpretation. As shown in Table 2, *first fixation duration, dwell time and fixation count* were significantly different between averted and directed-then-averted conditions, with higher values in the latter. In contrast, *second fixation duration* and *second fixation time* were not significantly different between these conditions. *First fixation time* metric did not show significant differences either; for this result, see the discussion point **Active agent's position effect** later in this section. The results for all the other five metrics might be understood and explained according to the procedure of the directed-then-averted gaze trial. Indeed, in this condition, once the first fixation had started on the active agent, users could observe a dynamic gaze change. This might have captured their attention and could explain the fact that they stared significantly longer towards the active agent during the first fixation. After, the active agent entered the averted gaze condition: this could explain why the directed-then-averted condition results of second fixation duration and second fixation time were not significantly different compared to the averted condition ones. Finally, dwell time and fixation count were nevertheless significantly higher in the dynamic condition, which could be explained by the multiple rechecks by users towards the active agent during the remaining time of a trial, to see if the agent would look at them again. Fig. 4 .2) summarises the comparison between the results on the averted and directed-then-averted conditions and its interpretation in relation with the

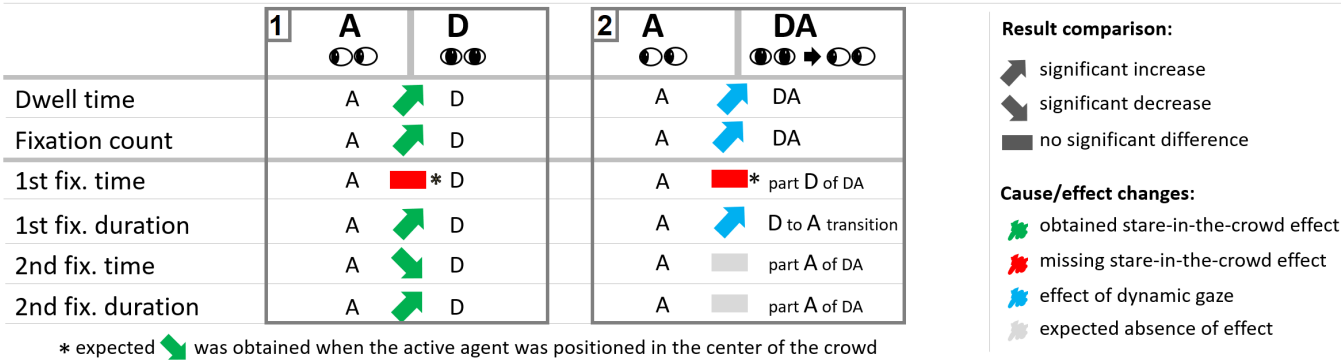

Figure 4: Summary of results for two representative cases of comparison: 1) A vs D reveals the presence of the stare-in-the-crowd effect in VR, and 2) A vs DA reveals effects of dynamic gazes

Table 3: Gaze metrics results - comparison of D vs. DA conditions

	Directed	Directed-then-Averted			
Metric	Mean (SD)	Mean (SD)	p-value	t	η_p^2
Dwell time	1570 (864)	808 (363)	<0.00001 ***	6.83	0.62
Fixation count	2.35 (1.03)	1.56 (0.56)	<0.00001 ***	6.17	0.57
1st fix. duration	552 (185)	483 (165)	0.07638	1.84	0.10
1st fix. time	4969 (1402)	4847 (1307)	0.69058	0.40	0.01
2nd fix. duration	554 (214)	374 (95)	0.00016 **	4.32	0.39
2nd fix. time	6861 (1785)	7773 (1724)	0.05246	-2.03	0.12

Time and duration in ms.

Table 4: Gaze metrics results - comparison of A vs. AD conditions

	Averted	Averted-then-Directed			
Metric	Mean (SD)	Mean (SD)	p-value	t	η_p^2
Dwell time	504 (175)	1503 (789)	<0.00001 ***	wilc.	wilc.
Fixation count	1.15 (0.29)	2.18 (0.79)	<0.00001 ***	-7.24	0.64
1st fix. duration	332 (77)	544 (176)	<0.00001 ***	-6.22	0.57
1st fix. time	5173 (1213)	5371 (1229)	0.87121	wilc.	wilc.
2nd fix. duration	407 (282)	644 (226)	0.00011 **	wilc.	wilc.
2nd fix. time	8602 (1395)	6978 (1544)	0.00032 **	4.08	0.36

Time and duration in ms.

stare-in-the-crowd effect and the effect of dynamic gaze changes.

In addition, when comparing our results to the ones of Crehan et al. [15], both studies found similar effects, except that in their case instead of finding a significant difference for the first fixation duration, they found it for the second fixation one.

Comparison D vs DA interpretation. As shown in Table 3, *dwell time, fixation count and second fixation duration* were significantly different between directed and directed-then-averted conditions, with lower values in the latter. **These results confirm the stare-in-the-crowd effect**: indeed, in a directed-then-averted condition, once the first fixation on the active agent had started, its gaze remained averted, thus significant differences are consistent with the ones observed for these metrics on the averted vs. directed comparison. In a similar way, in both conditions, the agent's gaze was directed before the first fixation started, which can explain the absence of the significant difference for the *first fixation time*. After that, for the *first fixation duration* and the *second fixation time*, the absence of significant difference between these two conditions is consistent with the interpretation given for averted vs. directed-then-averted conditions and could thus be explained the following: the gaze change of the active agent that occurred at the beginning of the first fixation could have captured the VR users' attention at a level not significantly different to the one caused by a directed gaze for the first fixation in terms of duration, and could have nonetheless made them check back towards this agent as soon as in the directed gaze condition, therefore through an early second fixation on it.

Table 5: Gaze metrics results - comparison of D vs. AD conditions

	Directed	Averted-then-Directed			
Metric	Mean (SD)	Mean (SD)	p-value	t	η_p^2
Dwell time	1570 (864)	1503 (789)	0.45729	0.75	0.02
Fixation count	2.35 (1.03)	2.18 (0.79)	0.15961	1.44	0.07
1st fix. duration	552 (185)	544 (176)	0.85695	0.18	0.00
1st fix. time	4969 (1402)	5371 (1229)	0.16703	-1.41	0.06
2nd fix. duration	554 (214)	644 (226)	0.13021	-1.56	0.08
2nd fix. time	6861 (1785)	6978 (1544)	0.75169	0.32	0.00

Time and duration in ms.

In addition, compared to our results, Crehan et al. [15] did not observe the stare-in-the-crowd effect in all the metrics, since they found no effect of dwell time or fixation count. However, as we did, they found a significant difference for the second fixation duration. Our differences may come from the specifics of our setup, e.g. using VR that adds depth and space information, unlike photographs.

Comparison A vs AD interpretation. As shown in Table 4, *dwell time, fixation count, first fixation duration, second fixation duration and second fixation time* were significantly different between conditions averted and averted-then-directed, with lower value for the second fixation time and higher values for the other metrics in the averted-then-directed condition. **These results confirm the stare-in-the-crowd effect in VR**: indeed in an averted-then-directed condition, once started the first fixation on the active agent, its gaze remains a directed gaze, therefore significant differences are consistent with the ones observed for these metrics on averted vs directed comparison. Moreover, *first fixation time* was not significantly different between the two conditions, which is coherent since in both conditions the active virtual agent starts with an averted gaze.

In addition, in comparison with Crehan et al.'s results [15], we found similar results, except the fact that they did not observe a higher level of first fixation duration in the dynamic condition. Similarly, our differences may come from the specifics of our setup, e.g. using VR that adds depth and space information, unlike photographs.

Comparison D vs AD interpretation. As shown in Table 5, *dwell time, fixation count, first fixation duration, second fixation duration and second fixation time* were not significantly different between the two conditions. These results are coherent since in the averted-then-directed condition, once the first fixation on the active agent had started, its gaze remained directed, i.e., with a gaze similar to the directed condition. Finally, for the *first fixation time* metric, there was no significant difference; see the paragraph **Active agent's position effect** for the discussion of this result. In addition, all our results are coherent with Crehan et al.'s ones [15].

Active agent's position effect. In addition to these results that

average all the data by gaze condition, our metrics can also be computed based on the averages obtained by each user across the trials that share both the same viewing conditions and the same position of the "active agent" in the crowd and therefore in the user's field of view – 2 repetitions in total for each condition. Due to the variability of the number of fixations across conditions and users, dwell time and fixation count metrics were preferred here over fixation time and duration metrics, since the former ones can always be computed even when no fixations occurred on the expected agent during the trials – in that case, missing values would be reported for the other metrics when computing averages. For these nine position conditions, we only compared the averted and directed gaze conditions here, as they were the most representative ones for the evaluation of our hypothesis H1. For our two metrics, the normality assumption could not be verified for all our dependent paired samples, thus Student's t-tests or Wilcoxon signed rank tests were ran depending on the case. Due to our multiple comparisons, we conducted a Bonferroni correction that changed our target significance level from α=0.05 to α=0.00555. Table 6 shows the results of these comparisons for the dwell time on the left, and for the fixation count on the right. In the tables, a symbol * indicates a p-value $<$0.00555, ** a p-value $<$0.00111, and *** a p-value $<$0.000111.

Table 6: Metrics comparison for each position A vs. D - dwell time and fixation count

	Dwell time metric	Fixation count metric
Paired samples for t-test	p-value	p-value
Left-close A - Left-close D	0.00203 *	0.00078 **
Left-middle A - Left-middle D	0.71318	0.43556
Left-far A - Left-far D	0.00902	0.05810
Centre-close A - Centre-close D	$<$0.00001 ***	$<$0.00001 ***
Centre-middle A - Centre-middle D	$<$0.00001 ***	$<$0.00001 ***
Centre-far A - Centre-far D	0.00137 *	0.00399 *
Right-close A - Right-close D	0.00001 ***	$<$0.00001 ***
Right-middle A - Right-middle D	0.00006 ***	0.00074 **
Right-far A - Right-far D	0.00008 ***	0.00021 **

These results show an effect of the active agent's position on the dwell time and fixation count results when comparing averted and directed conditions. For seven out nine positions a significant difference was found between these two conditions for this metric, revealing the presence of a stare-in-the-crowd effect; in contrast, for the middle and far left positions, no significant difference was found. Nonetheless, this result is in line with previous studies that discussed the real existence of a stare-in-the-crowd effect across any stimuli positions [14] and any position in the user's field of view [37,38]. In addition, we found that this absence of significant difference between averted and directed condition was due to a larger time spent on the middle/far left field of view on the averted gaze conditions rather than to a lower one on the directed condition, compared to the results obtained on other positions. This could be explained by a leftward bias of humans during a visual exploration on a scene, as described in the literature [8,18,36]. Finally, this difference on the left may also have been caused by our experimental stimuli. Indeed, in our experiment the averted gazes of the virtual crowd were always towards a distractor – our virtual speaker – positioned at the left of the user, meaning that the majority of the virtual crowd was looking in that direction. Yet, in their study about the stare-in-the-crowd effect, Palanica et Itier [38] found a congruency effect of the averted gazes on the user's gaze behaviour, in the sense that active agents whose positions were in the direction signaled by averted gazes were detected faster. Similarly, Sun et al. [49] also found an effect of the perceived direction of the gaze of the virtual crowd on users' gaze behaviour, where users tend to look towards the same direction that they perceive when the majority of the crowd is looking towards one particular direction – in our case to the left.

We also wanted to test if the active agent's position could have affected other metrics than dwell time and fixation count. We found an effect for the first fixation time on the trials where the active agent was in the centre – without distinction of depth i.e. 6 trials in total for each gaze condition (3 positions by left/central/right zone * 2 repetitions for each user). For these data samples, the normality assumption could not be verified for all our dependent paired samples, thus Student's t-tests or Wilcoxon signed rank tests were ran depending on the case. Due to our multiple comparisons, we conducted a Bonferroni correction that changed our target significance level from α=0.05 to α=0.016. Table 7 shows the results of these first fixation time comparisons, with one column for each gaze comparison studied, one line for each position zone – left/central/right, and one final line with the p-value previously obtained with the global data without position distinction. In this table, a symbol * indicates a p-value $<$0.0160, ** a p-value $<$0.0033, and *** a p-value $<$0.0003.

Table 7: First fixation time metric - comparisons by pair of gaze conditions and across position zones

Gaze comp.	A vs D	A vs DA	D vs DA	A vs AD	D vs AD
Position	p-value	p-value	p-value	p-value	p-value
Left	0.0745	0.0792	0.8816	0.0293	0.4587
Central	0.0027 **	0.0027 **	0.8340	0.3765	0.012 *
Right	0.9018	0.2103	0.2421	0.7535	0.6181
All	0.5312	0.3290	0.6906	0.8712	0.1670

These comparison results give new insights on the *first fixation time* metric, and allow for new interpretations about the effect of gaze conditions on it. First the data where all positions are gathered show no significant differences between any gaze conditions, as well as the results considering only left or right positions. However, data related to central positions reveal different results with: 1) the presence of significant differences for the comparisons between averted and directed gaze conditions (A vs. D), averted and directed-then-averted ones (A vs. DA), and directed and averted-then-directed ones (D vs. AD), and 2) the absence of significant differences for the other comparisons. Such results are interesting because they are the ones that were expected according to the stare-in-the-crowd effect: indeed, before the first fixation, the three comparisons present in 1) are equivalent to an averted vs. directed gaze comparison, whereas for the two comparisons of 2), gazes are the same ones in both conditions for these two comparisons (two averted, or two directed). These results **confirmed the presence of a stare-in-the-crowd effect in VR**, here regarding the results for the first fixation time metric for active agents in **central positions**.

We may have found this effect only in the central position because of visual differences between VR and photographs. Photographs resolution allows for high-quality display of a crowd in a narrow field of view, about 30° for a user looking at a computer screen. In contrast, in our VR setup the total field of view was larger for the user (the 100° of the FOVE headset), but, because of resolution issues and the scale 1:1 for the agents used to provide immersion in VR, more space was required for each agent. Therefore, it could explain why previous results are equivalent to our central part results.

5.2 Gaze behaviours and social anxiety

To investigate whether users with a higher level of social anxiety were less likely to gaze towards agents who are gazing at them, we computed correlations between the final score on the social anxiety questionnaire, i.e., the Liebowitz Social Anxiety Scale, and our gaze metric data. This final social anxiety score can range from 0 to 144, with low score depicting absence of social anxiety and high score depicting a significant presence of social anxiety. We conducted a Shapiro-Wilk test to determine if our variables were normally

distributed or not. As some of them were not normally distributed and to be able to compare the correlation coefficients between themselves, we conducted Spearman's rank-order correlation on our data, between the final social anxiety scores and the gaze metrics results.

As expected, we found some negative correlations between social anxiety and metrics of the eye-tracking data. In particular, dwell time for directed (D) and dynamic conditions (DA, AD) showed significant negative correlations ($D: r_s = -0.42, p = 0.022, AD: r_s = -0.57, p = 0.001, DA: r_s = -0.37, p = 0.047$), indicating that the more socially anxious the user was, less time he or she spent observing the agent whose gaze was directed towards them. The correlation was particularly high in the AD condition (*getting caught staring*). Other metrics were not correlated with social anxiety, except for the averted condition first fixation duration ($A: r_s = -0.40, p = 0.028$) and the averted-then-directed condition fixation count ($AD: r_s = -0.49, p = 0.006$).

6 General Discussion

Our study evaluated VR users' gaze behaviours depending on different gaze conditions that were applied to a virtual crowd, and therefore aimed to test the stare-in-the-crowd effect in VR. Our H1 hypothesis was that the stare-in-the-crowd effect would be preserved in VR, and H2 hypothesis that we would observe a negative correlation between the time spent towards the agents who are staring at the user and the user's level of social anxiety.

In terms of verifying H1, we compared our results with the one obtained by Crehan et al. [15] using similar metrics, and found similar effects, confirming the stare-in-the-crowd effect in VR. Some differences with the previous study were found also, but we were able to find explanations for this (see section 5.1). One major difference was that we used a VR environment that could have affected the gaze behaviour simply due to the field of view being different to the view of the people looking at photographs. It appears to be important how the user is positioned in VR as well, since our results showed that for the middle and far left field of view, some aspects of the stare-in-the-crowd effect were not present (related to dwell time metric) and in the left and right field of view for other aspects (here through first fixation metric, by finding expected results only for active agents in central positions.) An important difference was also between dynamic conditions of both studies. In our study, we found less gaze fixation behaviour in the dynamic conditions than in the directed static one, oppositely to the findings of the previous study. This could be explained by users expecting changes in the behaviour of virtual agents in VR, since agents were slightly animated (blinking), whereas photographic stimuli may not have had the same anticipation effect. Our results are potentially more accurately transferable to physical reality than previous results that were collected by using photographs as stimuli.

Regarding H2, our results show that social anxiety is negatively correlated with dwell time for all conditions that include directed gaze. Therefore, on average, the higher the social anxiety, the less time users spent looking at the agents when their gaze was directed towards them, which is in line with the gaze behaviour of socially anxious individuals [6]. Particularly interesting is the result that the averted-then-directed condition ("being caught staring") had the strongest correlation compared to other conditions, meaning that socially anxious individuals were particularly sensitive to agents who looked at them after the user saw them. Other metrics (fixation time, etc.) were not correlated, meaning that perhaps the additive effect of dwell time metric was stronger. However, we did get the negative correlation with fixation count for the averted-then-directed condition again, but also for the averted condition, with the first fixation duration. The latter could indicate that users with higher social anxiety may avoid to look at characters at the very beginning of the trial for fear of meeting their gaze. Some users reported their fear of the virtual agents in our post-experiment questionnaire and reported avoiding agents who were staring at them: *"actually, older people are super scary", "embarrassed by the stare of the avatars towards me, I run away from them rather quickly", "some avatars felt creepier than others, their gaze felt heavier when they were looking for afar, and more normal or natural when they were actually just in front of me"*. Importantly, we were able to demonstrate a stare-in-the-crowd effect in our study, indicating that the amount of socially anxious individuals in our sample of users was not high.

There are limitations to our study. Firstly, our sample of participants was not balanced in terms of gender, which may have affected our data. While our sample was not balanced in gender, we made sure that we had a balanced representation of both genders in the stimuli sample. We also cannot generalise our results to more natural social situations. While we designed the agents to be as realistic in appearance as possible, better models and animations could be used to make the results more transferable to interactions in the physical world. In addition, other scenarios than the one where virtual audience is listening to a speaker, could be considered. Moreover, in this study we took behavioural measures using an eye-tracking system and an indirect measure with the social anxiety questionnaire, however we could also have used some subjective measures such as presence and social presence [3, 47]. Another limitation is that we did not check specifically for cybersickness. Nonetheless we ensured a sufficient framerate in the FOVE headset and our VR users were seated and had limited movements, therefore adverse effects of cybersickness were limited. We also found the importance of where the user is positioned in VR as this affects the stare-in-the-crowd effect. Future studies are needed to better understand the stare-in-the-crowd effect at different observing positions and also in times when the user is allowed to move through the environment.

7 Conclusions and Future Work

This work addressed the well-known stare-in-the-crowd effect, which predicates the existence of a search asymmetry between directed and averted gaze towards the observer, with faster detection and longer fixation towards directed gaze. In other words, it represents the tendency of humans in noticing and observing, more frequently and for longer time, gazes oriented toward them (directed gaze) than gazes directed elsewhere (averted gaze). The existence of the stare-in-the-crowd effect has been already proven using photographic stimuli, but never in VR.

Our results confirmed the stare-in-the-crowd effect in VR alongside the evidence that this effect is milder with people reporting higher social anxiety levels. With this, we showed that gaze can indeed change the focus of attention of a user, and potentially trigger the interaction with an agent. Such results are very encouraging, since they can improve our understanding of social interactions in VR applications and help design more engaging experiences with agents. For example, our gaze conditions could be used to initiate the interaction with the user in a virtual crowd. We also demonstrated a simple dynamic gaze condition that signals complex social behaviour, e.g., directed-then-averted gaze could potentially be interpreted as a sign of embarrassment of the agent. These subtle gaze conditions could be explored further to create more believable social interactions in VR.

In the future, we plan to explore the stare-in-the-crowd and other related effects in more complex scenarios, e.g., including more dynamic and heterogeneous virtual agents, changing their number, giving the user different tasks. Moreover, we will expand our analysis to also consider further social and behavioural aspects of our human users, so as to see how they relate to the gazing times. Finally, we want to expand the subjects pool for achieving a better balance in terms of gender and age, which will also enable us to analyse the effects of such characteristics on the results.

Acknowledgments

The authors wish to thank the participants of the experiment. This work received funding from the European Union under grant agreement No 856879 (H2020 ICT-25 RIA PRESENT project), as well as the Brittany Region, project SAD-2019-REACTIVE, No 19023.

REFERENCES

[1] K. Allmendinger. Social presence in synchronous virtual learning situations: The role of nonverbal signals displayed by avatars. *Educational Psychology Review*, 22(1):41–56, 2010.

[2] J. N. Bailenson, A. C. Beall, J. Loomis, J. Blascovich, and M. Turk. Transformed social interaction, augmented gaze, and social influence in immersive virtual environments. *Human communication research*, 31(4):511–537, 2005.

[3] J. N. Bailenson, J. Blascovich, A. C. Beall, and J. M. Loomis. Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. *Presence: Teleoperators & Virtual Environments*, 10(6):583–598, 2001.

[4] J. N. Bailenson, J. Blascovich, A. C. Beall, and J. M. Loomis. Interpersonal distance in immersive virtual environments. *Personality and social psychology bulletin*, 29(7):819–833, 2003.

[5] J. N. Bailenson and N. Yee. Digital chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments. *Psychological science*, 16(10):814–819, 2005.

[6] S. R. Baker and R. J. Edelmann. Is social phobia related to lack of social skills? duration of skill-related behaviours and ratings of behavioural adequacy. *British Journal of Clinical Psychology*, 41(3):243–257, 2002.

[7] G. Bente, S. Rüggenberg, N. C. Krämer, and F. Eschenburg. Avatar-mediated networking: Increasing social presence and interpersonal trust in net-based collaborations. *Human communication research*, 34(2):287–318, 2008.

[8] A. Bourgeois, E. Badier, N. Baron, F. Carruzzo, and P. Vuilleumier. Influence of reward learning on visual attention and eye movements in a naturalistic environment: A virtual reality study. *Plos one*, 13(12):e0207990, 2018.

[9] M. A. Bühler and A. Lamontagne. Circumvention of pedestrians while walking in virtual and physical environments. *IEEE transactions on neural systems and rehabilitation engineering*, 26(9):1813–1822, 2018.

[10] J. K. Burgoon and A. E. Bacue. *Nonverbal communication skills.* Lawrence Erlbaum Associates Publishers, Mahwah, New Jersey, USA, 2003.

[11] M. Chollet and S. Scherer. Perception of virtual audiences. *IEEE computer graphics and applications*, 37(4):50–59, 2017.

[12] D. M. Clark. A cognitive model. *Social phobia: Diagnosis, assessment, and treatment*, pp. 69–73, 1995.

[13] C. Colombatto, B. van Buren, and B. J. Scholl. Gazing without eyes: A "stare-in-the-crowd" effect induced by simple geometric shapes. *Perception*, 49(7):782–792, 2020.

[14] R. M. Cooper, A. S. Law, and S. R. Langton. Looking back at the stare-in-the-crowd effect: Staring eyes do not capture attention in visual search. *Journal of vision*, 13(6):10–10, 2013.

[15] E. T. Crehan and R. R. Althoff. Measuring the stare-in-the-crowd effect: a new paradigm to study social perception. *Behavior research methods*, 47(4):994–1003, 2015.

[16] E. T. Crehan and R. R. Althoff. Me looking at you, looking at me: The stare-in-the-crowd effect and autism spectrum disorder. *Journal of Psychiatric Research*, 2021.

[17] H. Doi and K. Ueda. Searching for a perceived stare in the crowd. *Perception*, 36(5):773–780, 2007.

[18] T. Foulsham, A. Gray, E. Nasiopoulos, and A. Kingstone. Leftward biases in picture scanning and line bisection: A gaze-contingent window study. *Vision Research*, 78:14–25, 2013. doi: 10.1016/j.visres.2012.12.001

[19] D. Framorando, N. George, D. Kerzel, and N. Burra. Straight gaze facilitates face processing but does not cause involuntary attentional capture. *Visual cognition*, 24(7-8):381–391, 2016.

[20] M. Garau, M. Slater, V. Vinayagamoorthy, A. Brogni, A. Steed, and M. A. Sasse. The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 529–536, 2003.

[21] Y. Glémarec, J.-L. Lugrin, A.-G. Bosser, A. Collins Jackson, C. Buche, and M. E. Latoschik. Indifferent or enthusiastic? virtual audiences animation and perception in virtual reality. *Frontiers in Virtual Reality*, 2:72, 2021.

[22] M. Gonzalez-Franco, E. Ofek, Y. Pan, A. Antley, A. Steed, B. Spanlang, A. Maselli, D. Banakou, N. Pelechano Gómez, S. Orts-Escolano, et al. The rocketbox library and the utility of freely available rigged avatars. *Frontiers in virtual reality*, 1(561558):1–23, 2020.

[23] J. Harrigan, R. Rosenthal, K. R. Scherer, and K. Scherer. *New handbook of methods in nonverbal behavior research*. Oxford University Press, 2008.

[24] E. M. Hodge, M. Tabrizi, M. A. Farwell, and K. L. Wuensch. Virtual reality classrooms: Strategies for creating a social presence. *International Journal of Social Sciences*, 2(2):105–109, 2008.

[25] K. Horley, L. M. Williams, C. Gonsalvez, and E. Gordon. Social phobics do not see eye to eye:: A visual scanpath study of emotional expression processing. *Journal of anxiety disorders*, 17(1):33–44, 2003.

[26] T. Iachini, Y. Coello, F. Frassinetti, V. P. Senese, F. Galante, and G. Ruggiero. Peripersonal and interpersonal space in virtual and real environments: Effects of gender and age. *Journal of Environmental Psychology*, 45:154–164, 2016.

[27] S. Kahlon, P. Lindner, and T. Nordgreen. Virtual reality exposure therapy for adolescents with fear of public speaking: a non-randomized feasibility and pilot study. *Child and adolescent psychiatry and mental health*, 13(1):47, 2019.

[28] B. Lange and P. Pauli. Social anxiety changes the way we move—a social approach-avoidance task in a virtual reality cave system. *PloS One*, 14(12):e0226805, 2019.

[29] J. Li, Y. Kong, T. Röggla, F. De Simone, S. Ananthanarayan, H. De Ridder, A. El Ali, and P. Cesar. Measuring and understanding photo sharing experiences in social virtual reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2019.

[30] M. R. Liebowitz. Social phobia. *Modern problems of pharmacopsychiatry*, 1987.

[31] B. R. Manor and E. Gordon. Defining the temporal threshold for ocular fixation in free-viewing visuocognitive tasks. *Journal of Neuroscience Methods*, 128(1):85–93, 2003.

[32] L. Marschner, S. Pannasch, J. Schulz, and S.-T. Graupner. Social communication with virtual agents: The effects of body and gaze direction on attention and emotional responding in human observers. *International Journal of Psychophysiology*, 97(2):85–92, 2015.

[33] Y. Mohammad and T. Nishida. Reactive gaze control for natural human-robot interactions. In *2008 IEEE Conference on Robotics, Automation and Mechatronics*, pp. 47–54. IEEE, 2008.

[34] S. Narang, A. Best, T. Randhavane, A. Shapiro, and D. Manocha. Pedvr: Simulating gaze-based interactions between a real user and virtual crowds. In *Proceedings of the 22nd ACM conference on virtual reality software and technology*, pp. 91–100, 2016.

[35] L. Nummenmaa, J. Hyönä, and J. K. Hietanen. I'll walk this way: Eyes reveal the direction of locomotion and make passersby look and go the other way. *Psychological science*, 20(12):1454–1458, 2009.

[36] J. P. Ossandón, S. Onat, and P. König. Spatial biases in viewing behavior. *Journal of Vision*, 14(2):20–20, 02 2014. doi: 10.1167/14.2.20

[37] A. Palanica and R. Itier. Measuring the stare-in-the-crowd effect using eye-tracking: Effects of task demands. *Journal of Vision*, 11(11):1327–1327, 2011.

[38] A. Palanica and R. J. Itier. Searching for a perceived gaze direction using eye tracking. *Journal of Vision*, 11(2):19–19, 2011.

[39] X. Pan and A. F. d. C. Hamilton. Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology*, 109(3):395–417, 2018.

[40] C. Pelachaud. Modelling multimodal expression of emotion in a virtual agent. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3539–3548, 2009.

[41] A. Pittig, J. J. Arch, C. W. Lam, and M. G. Craske. Heart rate and heart rate variability in panic, social anxiety, obsessive–compulsive, and generalized anxiety disorders at baseline and in response to relaxation and hyperventilation. *International journal of psychophysiology*, 87(1):19–27, 2013.

[42] N. Ramamoorthy, K. Plaisted-Grant, and G. Davis. Fractionating the stare-in-the-crowd effect: Two distinct, obligatory biases in search for gaze. *Journal of Experimental Psychology: Human Perception and Performance*, 45(8):1015, 2019.

[43] D. Roth, P. Kullmann, G. Bente, D. Gall, and M. E. Latoschik. Effects of hybrid and synthetic social gaze in avatar-mediated interactions. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 103–108. IEEE, 2018.

[44] K. Ruhland, C. E. Peters, S. Andrist, J. B. Badler, N. I. Badler, M. Gleicher, B. Mutlu, and R. McDonnell. A review of eye gaze in virtual agents, social robotics and hci: Behaviour generation, user interaction and perception. *Computer Graphics Forum*, 34(6):299–326, 2015. doi: 10.1111/cgf.12603

[45] R. M. Schuetzler, G. M. Grimes, and J. S. Giboney. An investigation of conversational agent relevance, presence, and engagement. *Americas conference on information systems 2018 proceedings*, 2018.

[46] L. Schulze, J. S. Lobmaier, M. Arnold, and B. Renneberg. All eyes on me?! social anxiety and self-directed perception of eye gaze. *Cognition and Emotion*, 27(7):1305–1313, 2013. PMID: 23438447. doi: 10.1080/02699931.2013.773881

[47] M. Slater, M. Usoh, and A. Steed. Depth of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 3(2):130–144, 1994.

[48] H. J. Smith and M. Neff. Communication behavior in embodied virtual reality. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1–12, 2018.

[49] Z. Sun, W. Yu, J. Zhou, and M. Shen. Perceiving crowd attention: Gaze following in human crowds with conflicting cues. *Attention, Perception, & Psychophysics*, 79(4):1039–1049, 2017.

[50] A. M. Von der Pütten, N. C. Krämer, J. Gratch, and S.-H. Kang. "it doesn't matter what you are!" explaining social effects of agents and avatars. *Computers in Human Behavior*, 2010.

[51] M. Von Grünau and C. Anston. The detection of gaze direction: A stare-in-the-crowd effect. *Perception*, 24(11):1297–1313, 1995.

[52] T. F. Wechsler, M. Pfaller, R. E. van Eickels, L. H. Schulz, and A. Mühlberger. Look at the audience? a randomized controlled study of shifting attention from self-focus to nonsocial vs. social external stimuli during virtual reality exposure to public speaking in social anxiety. *Frontiers in Psychiatry*, 12, 2021.

[53] G. Westheimer. Eye movement responses to a horizontally moving visual stimulus. *AMA Archives of Ophthalmology*, 52(6):932–941, 12 1954.

[54] M. J. Wieser, P. Pauli, G. W. Alpers, and A. Mühlberger. Is eye to eye contact really threatening and avoided in social anxiety?—an eye-tracking and psychophysiology study. *Journal of anxiety disorders*, 23(1):93–103, 2009.

[55] M. J. Wieser, P. Pauli, M. Grosseibl, I. Molzow, and A. Mühlberger. Virtual social interactions in social anxiety—the impact of sex, gaze, and interpersonal distance. *Cyberpsychology, Behavior, and Social Networking*, 13(5):547–554, 2010.