



Published in final edited form as:

Cogn Behav Ther. 2022 September ; 51(5): 371–387. doi:10.1080/16506073.2022.2053882.

Attention Guidance Augmentation of Virtual Reality Exposure Therapy for Social Anxiety Disorder: A Pilot Randomized Controlled Trial

Mikael Rubin^{1,*}, Karl Muller², Mary M. Hayhoe², Michael J. Telch¹

¹Department of Psychology, The University of Texas at Austin, TX, USA

²Center for Perceptual Systems, The University of Texas at Austin, TX, USA

Abstract

Biased attention to social threat has been implicated in social anxiety disorder. Modifying visual attention during exposure therapy offers a direct test of this mechanism. We developed and tested a brief virtual reality exposure therapy (VRET) protocol using 360°-video and eye tracking. Participants (N = 21) were randomized to either standard VRET or VRET + attention guidance training (AGT). Multilevel Bayesian models were used to test (1) whether there was an effect of condition over time and (2) whether post-treatment changes in gaze patterns mediated the effect of condition at follow-up. There was a large overall effect of the intervention on symptoms of social anxiety, as well as an effect of the AGT augmentation on changes in visual attention to audience members. There was weak evidence against an effect of condition on fear of public speaking and weak evidence supporting a mediation effect, however these estimates were strongly influenced by model priors. Taken together, our findings suggest that attention can be modified *within and during* VRET and that modification of visual gaze avoidance may be casually linked to reductions in social anxiety. Replication with a larger sample size is needed.

Keywords

virtual reality; social anxiety; exposure therapy; attention

Social anxiety disorder (SAD) is characterized by elevated fear or anxiety about one or more social situations in which the individual may encounter scrutiny by others (American Psychiatric Association, 2013). SAD a highly prevalent common psychological concern with a lifetime prevalence of 12% (Ruscio et al., 2008), and confers significant impairment in multiple spheres of functioning (Aderka et al., 2012). SAD is also associated with elevated risk of comorbid depression and substance abuse (Stein & Stein, 2008).

There is increasing emphasis on the importance of understanding mechanisms underpinning psychological disorders (Moreno-Peral et al., 2020). One structured approach to addressing this in research has been through the Research Domain Criteria (RDoC), which emphasizes

*Corresponding Author: mikaelrubin@utexas.edu, 108 E. Dean Keeton, Austin, Texas, 78712-0187.

The authors report no conflicts of interest.

the importance of investigating specific mechanisms across taxonomic levels, from genes to behaviors (Kozak & Cuthbert, 2016). Attentional processes figure prominently in most contemporary models of SAD (Clark & Wells, 1995; Wong & Rapee, 2016) and have been the focus of considerable empirical work (Bögels & Mansell, 2004). Several distinct attentional profiles have been implicated. These include attentional *hypervigilance* to social evaluative cues (Rapee & Heimberg, 1997), attentional avoidance (Cisler & Koster, 2010), self-focused attention (Clark & Wells, 1995), attentional switching between internal and external social-evaluative threat cues (Rapee & Heimberg, 1997) and attentional hyperscanning (Chen et al., 2015).

Both theoretical models of Social Anxiety Disorder (Clark & Wells, 1995; Wong & Rapee, 2016) and some prior research (Kim et al., 2018; Rubin et al., 2020) support the hypothesis that avoidance of social threat may serve as a maintaining factor of the disorder. Following from the large literature on attention bias and social anxiety supporting its role as a critical feature of the disorder, researchers have worked to target attentional mechanisms of social anxiety by modifying attention directly using computerized tasks. Results from Attention Bias Modification (ABM) have been promising with meta-analysis indicating that it has moderate efficacy for reducing symptoms of social anxiety (Heeren, Mogoase, et al., 2015). However, several critical issues have emerged. First, ABM does not often outperform the attention control condition (where the probe distribution is 50/50), making it unclear whether the attention bias being targeted is really responsible for the treatment effects. Second, there is some evidence that within individuals with social anxiety there is no consistent attentional bias (Kruijt et al., 2019). Third, response to ABM in terms of clinical symptoms does not seem to be reflected in a change in attention bias. Together, these concerns suggest that while ABM may be effective, the reason why is unclear, making it important to consider additional ways to test attentional change in the treatment of social anxiety. Directly testing attentional change in the context of psychotherapy is challenging. For instance, it can be difficult, if not impossible, for a clinician to accurately assess the degree to which a patient is actually looking at audience members during an *in vivo* public speaking exposure, to assess the degree of attentional avoidance. However, this difficulty can be addressed with virtual environments that incorporate eye tracking. A virtual reality context makes it practicable to directly test whether attentional avoidance is causally implicated as a mechanism maintaining SAD.

To investigate attentional avoidance as a potential mechanism for SAD we developed a virtual reality environment where participants were required to deliver a speech to a real (that is, not digital avatars) audience who had been pre-recorded in a 360°-video. Participants were randomized to receive this public speaking exposure with or without visual attention guidance training. The attention guidance training involved instructing participants to look directly at the audience members during repeated presentations of the speech to the virtual audience. Visual gaze data were obtained for both groups using an Oculus Rift headset capable of displaying 360°-video conference room environment (See Figure 1). The SMI eye tracker upgrade in the Oculus Rift Headset allowed the clinician to validate attention allocation during each speech and provide feedback to participants. There is substantial evidence that virtual reality exposure therapy (VRET) for SAD is a highly effective intervention modality (Carl et al., 2019; Chesham et al., 2018). 360°-video is more

realistic than digital computer-generated avatars and while most VRET studies have used digital avatars, considering the focus on attention in the current research, the use of more realistic stimuli was important. Recent research has demonstrated the use of 360°-video for VRET to be effective (Reeves et al., 2021).

In line with the primary aims of the study, we asked whether individuals who received VRET augmented with attention guidance would show greater reduction in symptoms of social anxiety following the intervention compared to VRET alone. Second, we asked if the effects of the attention guidance augmentation were mediated by changes in gaze following the intervention. Specifically, we expected that individuals who received attention guidance would look more at uninterested (socially threatening) audience members posttreatment, and that this would at least partially account for the effects of the intervention on symptoms of social anxiety at the one-week follow-up. Thus, we tested not only whether the augmentation enhanced VRET, but also whether increased visual attention to audience members was a mechanism underpinning changes in symptoms. The current pilot RCT aimed to establish whether a larger study is warranted to test the efficacy of attentional guidance as a component for VRET.

Methods

Participants

Individuals ($N=21$) from the Austin community and from a large subject pool at the University of Texas who were diagnosed with SAD and displayed marked fear of public speaking were enrolled in the study (see Table 1 for the demographic summary – there were no meaningful differences between groups at baseline). The Institutional Review Board at the University of Texas at Austin approved all study procedures. Inclusion Criteria for the study were: (1) Age 18–65; (2) Fluent in English; (3) Personal Report of Public Speaking Anxiety > 98; (4) Leibowitz Social Anxiety Scale > 30; (5) Peak fear > 50 on the behavioral approach task during the baseline public speaking challenge; (6) Meets DSM-5 Criteria for Social Anxiety Disorder. Exclusion Criteria for the study were: (1) Currently receiving CBT for Social Anxiety Disorder; (2) Significant visual impairment precluding the use of virtual reality equipment; (3) Unstable dose of psychotropic medications within 3 weeks prior to baseline assessment; (4) Current alcohol or substance use disorders; (5) Current, or history of bipolar disorder; current, or history of psychosis; (6) Serious suicidal risk, as determined by clinical interview.

Study Design

The study investigated an attention augmentation strategy in a 2-arm randomized controlled trial. Adults with SAD were enrolled in a 2-week (3 visit) VRET protocol. Participants were randomly assigned to one of two conditions: (a) Virtual reality exposure therapy plus attention guidance training (VRET + AGT, $n = 11$) or (b) Standard virtual reality exposure therapy (VRET, $n = 10$). Symptoms of social anxiety were assessed at baseline, posttreatment and one-week posttreatment. Enrollment began 03/13/2019 and data collection ended 03/11/2020. The trial was registered as “Efficacy of an Attention Guidance VR

Intervention for Social Anxiety Disorder”, trial number: [NCT03683823](https://clinicaltrials.gov/ct2/show/study/NCT03683823) and can be accessed through clinicaltrials.gov. See Figure 2 for the flow diagram.

Screening procedures

Potential participants first provided written informed consent, then completed an online pre-screen consisting of demographics information, treatment history, the Leibowitz Social Anxiety Scale (LSAS) and the Personal Report of Public Speaking Anxiety (PRPSA). Participants that endorsed clinically elevated symptoms of social anxiety (LSAS score of 30 or greater) and endorsed moderate levels of public speaking anxiety (PRPSA score of 98 or greater) were invited to the in-person assessment. Participants reviewed the informed consent process, treatment procedures, and potential risks and benefits of participation with a staff member. Participants then completed self-report questionnaires and a diagnostic assessment conducted by a doctoral student in clinical psychology. Participants were then invited to the VR-lab to complete a public speaking challenge. The public speaking challenge involved an orientation to the virtual reality environment, 5-minutes to prepare a 3-minute speech on a topic they selected (from a list), then giving the 3-minute speech while standing (participants were given strict instructions not to walk around in order to ensure safety). Participants reported subjective units of distress before and after the speech. Following the public speaking challenge, eligible participants were randomized (using a balanced block-randomization procedure conducted in R using the *blockrand* package (Snow & Snow, 2013) implemented by M.R.; block size was 4) to one of the two arms and started treatment immediately. The clinician that conducted the assessment also allocated the participant (they were not blind to the condition prior to allocation) and conducted the intervention. Neither the clinician nor participant were blind to the condition of the intervention following allocation or during the follow-up assessments.

Intervention

Virtual Reality Exposure Therapy (VRET).—Participants received a brief standardized VRET protocol for social anxiety, which consisted of two 45-minute sessions delivered over a one-week period by a graduate student clinician supervised by MJT. Treatment consisted of (1) psychoeducation and (2) public speaking exposures. Threat appraisals were collected prior to and following each exposure to assess anticipated fear, peak fear, and post fear. During the first session of the treatment participants received brief psychoeducation regarding SAD and a treatment rationale emphasizing that confrontation of feared and/or avoided situations is critical. Each exposure session consisted of completing six 3-minute speeches delivered while standing. Participants were able to select one of several prompts at the beginning of each session. Participants had 5-minutes to prepare a speech based on the prompt and gave all six speeches during a given day on the same prompt. Between speeches participants had a 1-minute break. Following the six exposures, participants briefly processed the session with the clinician.

Virtual Reality Exposure Therapy + Attention Guidance Training (VRET + AGT).—Participants completed the same protocol as the standard VRET condition with three differences. (1) The treatment rationale included information about the importance of engaging in actions that directly counteract the naturalistic behavioral tendencies associated

with anxiety and specifically the importance of looking directly at audience members. (2) Before each speech the participant was directed to address a specific audience member throughout the speech, focusing specifically on their face. For each of the six speeches, the participant focused on a different (neutral, interested, or uninterested) audience member's face. (3) After each speech the clinician used an automated program to assess the percentage of time the participant was directly looking at their "target" audience member. The clinician provided the specific percentage along with encouragement to focus on the "target" audience member during each speech. Two video examples (Video 1 – standard VRET and Video 2 – VRET + AG) are provided to illustrate the difference between the two conditions in terms of gaze by study participants.

Posttreatment Assessments

Posttreatment and follow-up assessments were the same self-report measures as pre-treatment and included another public speaking challenge. Follow-up assessment clinicians were not blinded to intervention allocation. Participants completed the posttreatment assessment immediately following the completion of the intervention and the follow-up assessment one-week after the posttreatment assessment.

Measures

Personal Report of Public Speaking Anxiety (PRPSA)—The PRPSA (McCroskey, 1970) is a 34-item instrument that is designed to assess public speaking anxiety.

Liebowitz Social Anxiety Scale Self Report Version (LSAS-SR)—The LSAS self-report scale (Liebowitz, 1987) is a 48-item measure of fear and avoidance concerning social interactions and performance situations.

Speech Anxiety Thoughts Inventory (SATI)—The SATI (Cho et al., 2004) is a two factor (prediction of poor performance and fear of negative evaluation by audience) instrument, measuring maladaptive cognitions associated with speech anxiety.

Subjective Units of Distress (SUDs)—Participants rated their anticipated fear (before each speech) and their peak fear and their end fear (after each speech) from 0, no fear to 100, extreme fear.

Structured Clinical Interview for DSM-5 (SCID-5)—The SCID-5 (First et al., 2015) is a semi-structured clinician administered interview that is the gold-standard for determining mental health diagnoses for DSM-5. Selected portions of the SCID-5 were administered by a graduate clinician to assess for social anxiety disorder, alcohol and substance use disorders, psychosis, and bipolar disorder.

Columbia Suicidality Severity Rating Scale (C-SSRS)—The C-SSRS (Posner et al., 2011) is a semi-structured clinician-administered measure to assess suicidality. The C-SSRS was administered by a graduate clinician.

Concurrent treatment.: Psychotropic medication and current utilization of psychotherapy was assessed on the online prescreen.

Demographics.: Participants were asked to provide demographic information including sex, age, race/ethnicity, visual impairment, language history and use, etc. on the internet prescreen.

Materials

360°-video virtual reality environments.—The 360°-video virtual reality (VR) environments consist of an audience of six individuals sitting in chairs around a conference table or in a lecture hall (Figure 1). The pre-treatment public speaking challenge and treatment context were the conference room. The post-treatment and follow-up public speaking challenge context were the auditorium. There are two groups of audience members – public speaking challenge audience members and treatment audience members. All videos featured the six audience members acting as if they are listening to a speech with varying levels of interest. Audience members were coached to behave interested (nodding, smiling), neutral (no facial expressions), or uninterested (looking away, looking at phone). Audience members played different roles in each video. The actors in the video were researchers (undergraduate, post-bac, and graduate) in psychology at the University of Texas at Austin. The video was filmed with a Samsung Gear 360 camera, mounted on a tripod.

There were only two videos filmed for each of the treatment days. Due to logistical constraints the same video was used for all six exposures on a given day. However, no participant observed that the same 360°-video was used multiple times.

Virtual Reality Headset and Eye Tracker.—Participants wore the Oculus Rift DKII virtual reality headset with built-in position tracking. The Oculus was upgraded with an SMI eyetracker to provide high-resolution eye tracking at a sampling rate of 75 Hz. A HiBall motion-tracking system (3rdTech) was used to track head movements. However, because the video was filmed from a fixed viewpoint, only the rotations (and not the translations) were used to update the image in the HMD. Participants completed a brief calibration procedure prior to beginning the speech. Videos of the eye tracking and the video-display (i.e., what the participants saw) were recorded at each video-frame and saved as a .MOV file. These .MOV files were used to later verify the automated eye-gaze analyses.

Gaze Processing

Eye movement data were collected pre-treatment, at each exposure trial, at the posttreatment assessment and at the 1-week follow-up assessment. The methods used for processing gaze data were the same as those previously used (Rubin et al., 2020): Vizard 4 (WorldViz) was used to display the 360°-video and collect the eye tracking data. OpenPose (Cao et al., 2021) was used to detect audience members within the 360°-video and dynamic regions of interest (ROIs) that encompassed each audience member were generated using custom MATLAB code (because the audience members moved, the ROIs could not be static). ROIs encompassed the face, hand, and torsos of each audience member with a small (~3° to ~6°) buffer to encompass fixations very close to audience members. Each fixation was assigned

an ROI (i.e., audience member) based on the closest OpenPose keypoint – however, if the fixation was not on an ROI it was assigned as a background fixation. Fixations were detected using a well-established in-house algorithm (Kit et al., 2014; Li et al., 2016; Tong et al., 2017). A fixation was identified if the eye was relatively stable (less than 50°/s and longer than 85 milliseconds). Fixations that were close together (within 1° and less than 80ms apart) were combined. If there was missing gaze data (i.e., track loss) a single fixation on an ROI was still counted as long as the fixations were close together (as above).

Data Analysis Plan

The primary aims of this study were to 1) examine whether an attention guidance augmentation enhanced VRET compared to VRET alone and 2) test whether changes in gaze behavior following the intervention mediated the effects VRET. To test our primary hypotheses regarding the influence of the intervention on fear of public speaking (PRPSA), we conducted Bayesian multilevel models using the *brms* package version 2.15 (Bürkner, 2018). For aim 1 we examined the interaction between assessment and group predicting the outcome post-treatment and at 1-week-followup. For aim 2, we examined the indirect (i.e., mediating) effect of proportion of fixations to uninterested (socially threatening audience members) at the post-treatment assessment, on the relationship between intervention group and the post-treatment assessment of PRPSA at the 1-week-followup (see Supplementary Figure 1). To facilitate interpretation of the mediation analysis, we partially standardized the model coefficients after completing the analyses using unstandardized variables following recommendations in the literature regarding indirect effect sizes when X is dichotomous (Hayes & Rockwood, 2017; Preacher & Kelley, 2011). In all models, we included average proportion of fixations to audience members during intervention sessions as a covariate to control for variation in treatment adherence. We completed the same analyses to evaluate our secondary outcome of general social anxiety symptoms measured with the LSAS. As integrity checks on the efficacy of the attention augmentation condition we tested whether there were group differences for average number of fixations on audience members during the intervention trials, as well as whether there were differences in proportion of fixations to uninterested audience members post-intervention and at 1-week follow-up.

We computed Bayes Factors (BFs) using the Savage-Dickey Density ratio (Wagenmakers et al., 2010) for all models where we set priors using the *hypothesis* function in *brms*. The Savage-Dickey Density ratio was calculated in the current context by dividing the posterior density by the prior density at zero (a null effect). For each result we report the beta estimates, 95% highest posterior density interval (HDI), and BFs of the model estimated with our original prior. We also provide the range of BFs as well as the sensitivity of the beta estimates based on our sensitivity analyses (see supplementary materials). BFs reflect the likelihood of an estimate in relation to the priors, whereas the 95% HDI indicates likelihood of the estimate falling within the posterior distribution. Discrepancy between evidence from the 95% HDI and the BF reflect an issue with ascertaining a null effect and may indicate a Type I conflict - a value is outside a credible interval, but the BF supports the null, or a Type II conflict - a value is within the credible interval, but the BF rejects the null (Lovric, 2019). Note that BFs < 3 or > .33 reflect only anecdotal evidence, suggesting only a small degree of confidence in the estimates in relation to prior evidence. We also computed an approximate

effect size for the main outcomes of the multilevel models following recent guidelines from Kurz (2021).

Prior Estimates.—We largely followed the WAMBS (When to worry and how to Avoid the Misuse of Bayesian Statistics) checklist (Depaoli & van de Schoot, 2017). This checklist provides a step-by-step approach to ensuring that a model estimation procedure is acceptable and that the influence of the priors is well delineated. We tested the sensitivity of the priors by using less informative (smaller effects) parameter estimates as well as uninformative default (flat) priors centered on zero to determine the influence of different priors on the posterior estimates.

Power analysis.—We did not conduct a power analysis that reflected the sample size for the current pilot study. We had initially conducted a power analysis through a simulation study prior to COVID-19, which indicated a sample size of 60 would be sufficient to detect a meaningful effect for both aims 1 and 2. However, due to COVID-19 enrollment ended before we could meet our recruitment goals. Given that research was necessarily stopped, we decided to rely on the strengths of the Bayesian approach highlighted above to investigate whether it would be worthwhile to conduct a more extensive RCT with a larger sample.

Data and syntax are available at <https://osf.io/un92m/>. Supplemental materials include additional information regarding choice of priors, analytic methods, and results for tests of baseline group differences, analyses with the Speech Anxiety Thoughts Inventory, and estimation bias based on use of a range of different priors.

Results

Intervention Integrity Checks

Table 2 summarizes the proportion of fixations averaged across participants to Uninterested, Interested, and Neutral audience members by group across time points. We primarily evaluated the role of proportion of fixations to uninterested (socially threatening) audience members, but also explored proportion of fixations to interested and neutral audience members. There were meaningful differences in the proportion of gaze allocated to audience members during treatment in the standard-exposure compared to attention-guidance conditions $b = 0.27$, 95% highest posterior density interval (HDI) [0.11, 0.41], Bayes Factor (BF) = 70.14, with greater gaze allocated to audience members in the attention-guidance condition during the intervention trials. There were no meaningful differences in the overall change in proportion of fixations to uninterested (socially threatening) audience members at post-treatment $b = 0.14$, 95% HDI [-0.16, 0.44], BF = 1.02 or at one-week follow-up $b = 0.23$, 95% HDI [-0.08, 0.53], BF = 2.00. There was also no main effect of group $b = 0.09$, 95% HDI [-0.30, 0.48], BF = 1.02. There was a meaningful group (VRET vs. VRET + attention guidance) \times assessment interaction at post-treatment $b = 0.16$, 95% HDI [0.07, 0.26], BF = 2.52, and at one-week follow-up $b = 0.30$, 95% HDI [0.20, 0.39], BF = 0.91.

Effects of Virtual Reality Exposure Therapy

Means and standard deviations for the primary and secondary outcomes are presented in Table 2, below. Figure 3 illustrates the primary findings for the intervention outcomes across assessments, relating to aim 1. Results for the Speech Anxiety Thoughts Inventory were in line with the other measures and are reported in the supplementary materials.

There was a large main effect of time on fear of public speaking, with a meaningful reduction in fear of public speaking post-treatment $b = -17.37$, 95% highest posterior density interval (HDI) $[-21.64, -9.63]$, Bayes Factor (BF) = 510.80, $d_{\text{GMA-raw}} = -1.11$, 95% HDI $[-1.58, -0.63]$ and at 1-week follow-up $b = -13.16$, 95% HDI $[-21.28, -9.01]$, BF = 26.14, $d_{\text{GMA-raw}} = -1.68$, 95% HDI $[-2.68, -0.70]$. We found moderate evidence against a main effect of group (standard exposure vs. attention augmentation) on fear of public speaking $b = -5.55$, 95% HDI $[-13.88, 2.53]$, BF = 0.28. Moreover, we did not find an effect of group at post-treatment $b = 6.14$, 95% HDI $[-1.57, 13.95]$, BF = 0.38, or at one-week follow-up $b = 3.19$, 95% HDI $[-10.16, 16.57]$, BF = 0.37.

We found a moderate main effect of time for general symptoms of social anxiety at post-intervention $b = -11.72$, 95% HDI $[-18.61, -4.58]$, Bayes Factor (BF) = 6.45, $d_{\text{GMA-raw}} = -0.60$ 95% HDI $[-1.02, -0.15]$ and large effect at 1-week follow-up $b = -21.99$, 95% HDI $[-29.08, -14.70]$, BF = 615.78, $d_{\text{GMA-raw}} = -2.07$ 95% HDI $[-2.94, -1.16]$. We did not find a main effect of group on general social anxiety symptoms $b = -4.86$, 95% HDI $[-17.08, 11.48]$, BF = 0.84; similarly, we found no effect of group at post-treatment $b = -2.05$, 95% HDI $[-13.48, 18.18]$, BF = 0.24 or one-week follow-up $b = -3.20$, 95% HDI $[-15.27, 22.15]$, BF = 0.44.

Mediating Effects of Gaze Behavior on Intervention Outcomes

We found anecdotal evidence (based on the BFs) that greater proportion of fixations to Uninterested audience members at the post-treatment assessment, mediated the effect of group (standard exposure vs. attention guidance augmentation) on fear of public speaking (partially standardized indirect effect = -0.218 , 95% HDI $[-0.605, 0.026]$, BF = 2.85) or general symptoms of social anxiety (partially standardized indirect effect = -0.097 , 95% HDI $[-0.197, -0.024]$, BF = 1.58) at the one-week follow-up. Taken together our results suggest that there is very weak evidence supporting the role of attention change on symptoms of fear of public speaking and general social anxiety, but further research is needed with larger samples. In Figure 4 we present the mediation models and in Supplemental Figure 1, we highlighted the influence of priors on our estimation of the indirect effect – the large degree of bias further emphasizes the importance of obtaining a larger sample to evaluate the stability of the estimates of these effects.

Discussion

This pilot study tested attentional avoidance as a potential change mechanism for social anxiety *during* 12 repeated public speaking exposure trials across two sessions. Our first aim was to examine whether an attention guidance augmentation would enhance the efficacy of a virtual reality exposure intervention for social anxiety disorder. There was a large reduction

in fear of public speaking, general symptoms of social anxiety, and cognitions related to public speaking anxiety across groups. There was anecdotal (based on the BFs) evidence in favor of the null (that there was no difference between the two intervention groups). Given the small sample size in this pilot study, the bias of the estimates based on the priors and the consistently weak support for the null across priors, further research with larger samples may be warranted. However, preliminary evidence does not support the presence of an effect of the attention guidance augmentation on fear of public speaking or general symptoms of social anxiety.

Our second aim was to test whether the influence of intervention group on fear of public speaking and general social anxiety symptoms was mediated by changes in gaze behavior following the intervention. There was strong evidence that our intervention engaged the target mechanism - the exposure augmentation led to a meaningful change in attention allocation, with a substantially greater proportion of gaze toward uninterested (socially threatening) audience members compared to the standard exposure group following the intervention. However, evidence regarding decreased avoidance as a potential mechanism maintaining social anxiety was slight. It is important to acknowledge that the sample size of the current study may have limited the possibility of detecting this indirect effect. In particular, if the effect of reducing gaze avoidance on symptoms of social anxiety is smaller than anticipated, then conducting a study with a larger sample is especially important.

There are other considerations beyond sample size that may have influenced the effect of the attention augmentation. We conducted a brief 2-session protocol, because the reduced efficacy compared with a full-length (e.g., 8-week) protocol makes it more feasible to test the influence of potential mechanisms (see for instance Niles et al., 2015). Given that the two-session VRET intervention was highly effective, it is possible that the influence of the intervention augmentation was masked. Longer follow-ups may have been useful in determining whether the augmentation provided additional benefits in social anxiety symptom reduction. It is also possible that the intervention and assessment periods were too few and/or too close together to detect the effect of changes in gaze behavior on social anxiety symptoms. Since gaze and attention are tightly linked to learning, it is possible that individuals with social anxiety who typically avoid social information needed more time to adjust their priors about appraisals of social information before symptom change could emerge. With only a one-week follow-up, there may not have been sufficient opportunities to acquire evidence in the real-world that greater gaze towards others in social situations is acceptable. Also, although the way we measured and targeted attention was straightforward and based on previous work (Kim et al., 2018; Rubin et al., 2020), attention is a dynamic and complex process, and it is possible that alternative ways to evaluate gaze behavior may yield further insights as to their role in social anxiety treatment. Additionally, our data indicated that there was reduced avoidance of all audience members in the attention augmentation group. This may suggest that our experimental intervention (which focused on all audience members) was not sufficiently focused on explicit social threat (i.e., uninterested audience members alone). Finally, we screened participants based on their self-reported fear following the pre-treatment public speaking challenge. This was to ensure responsivity to the 360°-video stimuli during VRET. However, this may also have facilitated some of the treatment efficacy observed across groups as individuals that found

the public speaking more challenging were more likely to benefit from the intervention. It may be worthwhile to consider including anyone meeting criteria for SAD (or even sub-clinical levels of social anxiety) in future research.

The use of eye tracking in VR is growing but may not be accessible to all researchers. Use of head orientation as a proxy for gaze location has been suggested as one way to overcome this limitation. In the current context the divergence between gaze and head orientation was sufficiently large to make such comparisons extremely problematic. However, this may have been a feature of the 360°-videos used. Researchers may consider development of 360°-video and other VR stimuli in such a way so as to make use of head orientation an acceptable proxy.

Several limitations are important to note. First, while there is evidence that non-interactive public speaking challenges with 360°-video can be an effective treatment modality (Reeves et al., 2021), it is likely that interactions are an important ingredient to the efficacy of social exposures broadly. Second, we did not include at baseline a non-socially anxious group to confirm the presence of biased attention among individuals with social anxiety. Moreover, we did not assess presence (the feeling that you are really ‘there’) during assessments or treatment, which may have played a role in limiting the efficacy of the intervention for some. Additionally, for the follow-up assessment, clinicians were not blind to the treatment condition.

Our findings highlight the utility of a Bayesian approach as we were able to conduct a meaningful analysis despite a small sample size and interpret our ambiguous findings in a way that serves to inform future research. It can be useful to identify support for the null at early stages in testing potential treatment mechanisms (Else et al., 2020), since even with small samples strong support for the null can curtail avenues of research that are unlikely to yield meaningful results. That we found no strong support in either direction, can be taken as evidence that future research with larger sample sizes is warranted to clarify the role of attention as an augmentation strategy for VRET.

Taken together our findings offer further validation that VRET for social anxiety disorder is a highly effective treatment. Additionally, we showed that attentional processes can be directly altered *during* exposure therapy. Despite ambiguous findings regarding the causal influence of attentional change on social anxiety symptoms, this study represents a useful first step towards the integration of attention modification directly into therapy.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding:

This work was supported by the National Institute of Health (NIH) [F31 MH118784]

References

- Aderka IM, Hofmann SG, Nickerson A, Hermesh H, Gilboa-Schechtman E, & Marom S (2012). Functional impairment in social anxiety disorder. *Journal of Anxiety Disorders*. 10.1016/j.janxdis.2012.01.003
- American Psychiatric Association. (2013). DSM-5 Diagnostic Classification. In *Diagnostic and Statistical Manual of Mental Disorders*. 10.1176/appi.books.9780890425596.x00diagnosticclassification
- Bögels SM, & Mansell W (2004). Attention processes in the maintenance and treatment of social phobia: Hypervigilance, avoidance and self-focused attention. *Clinical Psychology Review*. 10.1016/j.cpr.2004.06.005
- Bürkner PC (2018). Advanced Bayesian multilevel modeling with the R package brms. *R Journal*. 10.32614/rj-2018-017
- Cao Z, Hidalgo G, Simon T, Wei SE, & Sheikh Y (2021). OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 10.1109/TPAMI.2019.2929257
- Carl E, Stein AT, Levihn-Coon A, Pogue JR, Rothbaum B, Emmelkamp P, Asmundson GJG, Carlbring P, & Powers MB (2019). Virtual reality exposure therapy for anxiety and related disorders: A meta-analysis of randomized controlled trials. *Journal of Anxiety Disorders*. 10.1016/j.janxdis.2018.08.003
- Chen NTM, Thomas LM, Clarke PJF, Hickie IB, & Guastella AJ (2015). Hyperscanning and avoidance in social anxiety disorder: The visual scanpath during public speaking. *Psychiatry Research*. 10.1016/j.psychres.2014.11.025
- Chesham RK, Malouff JM, & Schutte NS (2018). Meta-Analysis of the Efficacy of Virtual Reality Exposure Therapy for Social Anxiety. *Behaviour Change*. 10.1017/bec.2018.15
- Cho Y, Smits JAJ, & Telch MJ (2004). The Speech Anxiety Thoughts Inventory: Scale development and preliminary psychometric data. *Behaviour Research and Therapy*. 10.1016/S0005-7967(03)00067-6
- Clark D. a., & Wells A (1995). A cognitive model of social phobia. *Social Phobia: Diagnosis, Assessment, and Treatment*.
- Depaoli S, & van de Schoot R (2017). Improving transparency and replication in Bayesian statistics: The WAMBS-checklist. *Psychological Methods*. 10.1037/met0000065
- Elsley JWB, Filmer AI, Galvin HR, Kurath JD, Vossoughi L, Thomander LS, Zavodnik M, & Kindt M (2020). Reconsolidation-based treatment for fear of public speaking: a systematic pilot study using propranolol. *Translational Psychiatry*. 10.1038/s41398-020-0857-z
- First MB, Williams JBW, Karg RS, & Spitzer RL (2015). Structured clinical interview for DSM-5 research version. American Psychiatric Association, Washington D.C.
- Hayes AF, & Rockwood NJ (2017). Regression-based statistical mediation and moderation analysis in clinical research: Observations, recommendations, and implementation. *Behaviour Research and Therapy*. 10.1016/j.brat.2016.11.001
- Kim H, Shin JE, Hong YJ, Shin, Bin Y, Shin YS, Han K, Kim JJ, & Choi SH (2018). Aversive eye gaze during a speech in virtual environment in patients with social anxiety disorder. *Australian and New Zealand Journal of Psychiatry*. 10.1177/0004867417714335
- Kit D, Katz L, Sullivan B, Snyder K, Ballard D, & Hayhoe M (2014). Eye movements, visual search and scene memory, in an immersive virtual environment. *PLoS ONE*. 10.1371/journal.pone.0094362
- Kozak MJ, & Cuthbert BN (2016). The NIMH Research Domain Criteria Initiative: Background, Issues, and Pragmatics. *Psychophysiology*. 10.1111/psyp.12518
- Li CL, Aivar MP, Kit DM, Tong MH, & Hayhoe MM (2016). Memory and visual search in naturalistic 2D and 3D environments. *Journal of Vision*. 10.1167/16.8.9
- Liebowitz MR (1987). Social phobia. *Modern Problems of Pharmacopsychiatry*.
- Lovric MM (2019). Conflicts in bayesian statistics between inference based on credible intervals and bayes factors. *Journal of Modern Applied Statistical Methods*. 10.22237/JMASM/1556670540

- McCroskey JC (1970). Measures of communication-bound anxiety. *Speech Monographs*. 10.1080/03637757009375677
- Moreno-Peral P, Bellón JÁ, Huibers MJH, Mestre JM, García-López LJ, Taubner S, Rodríguez-Morejón A, Bolinski F, Sales CMD, & Conejo-Cerón S (2020). Mediators in psychological and psychoeducational interventions for the prevention of depression and anxiety. A systematic review. In *Clinical Psychology Review*. 10.1016/j.cpr.2020.101813
- Posner K, Brown GK, Stanley B, Brent DA, Yershova KV, Oquendo MA, Currier GW, Melvin GA, Greenhill L, Shen S, & Mann JJ (2011). The Columbia-suicide severity rating scale: Initial validity and internal consistency findings from three multisite studies with adolescents and adults. *American Journal of Psychiatry*. 10.1176/appi.ajp.2011.10111704
- Preacher KJ, & Kelley K (2011). Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychological Methods*. 10.1037/a0022658
- Rapee RM, & Heimberg RG (1997). A cognitive-behavioral model of anxiety in social phobia. *Behaviour Research and Therapy*. 10.1016/S0005-7967(97)00022-3
- Reeves R, Elliott A, Curran D, Dyer K, & Hanna D (2021). 360° Video virtual reality exposure therapy for public speaking anxiety: A randomized controlled trial. *Journal of Anxiety Disorders*. 10.1016/j.janxdis.2021.102451
- Rubin M, Minns S, Muller K, Tong MH, Hayhoe MM, & Telch MJ (2020). Avoidance of social threat: Evidence from eye movements during a public speaking challenge using 360°-video. *Behaviour Research and Therapy*, 103706.
- Ruscio AM, Brown TA, Chiu WT, Sareen J, Stein MB, & Kessler RC (2008). Social fears and social phobia in the USA: Results from the National Comorbidity Survey Replication. *Psychological Medicine*. 10.1017/S0033291707001699
- Stein MB, & Stein DJ (2008). Social anxiety disorder. In *The Lancet*. 10.1016/S0140-6736(08)60488-2
- Tong MH, Zohar O, & Hayhoe MM (2017). Control of gaze while walking: Task structure, reward, and uncertainty. *Journal of Vision*. 10.1167/17.1.28
- Wagenmakers EJ, Lodewyckx T, Kuriyal H, & Grasman R (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage-Dickey method. *Cognitive Psychology*. 10.1016/j.cogpsych.2009.12.001
- Wong QJJ, & Rapee RM (2016). The aetiology and maintenance of social anxiety disorder: A synthesis of complimentary theoretical models and formulation of a new integrated model. In *Journal of Affective Disorders*. 10.1016/j.jad.2016.05.069

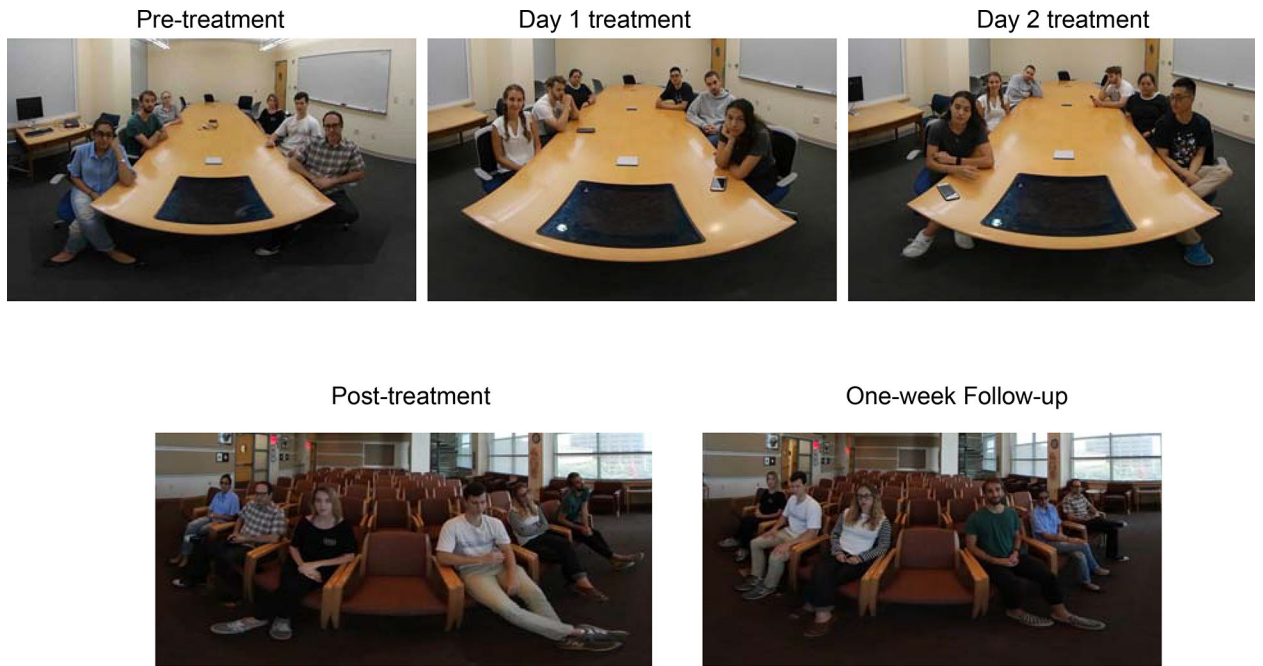


Figure 1. Stills of 360°-video stimuli used for each part of the study. All stills are cropped. Audience members in Pre-treatment, Post-treatment, and One-week Follow-up are the same; audience members in Day 1 and Day 2 are the same.

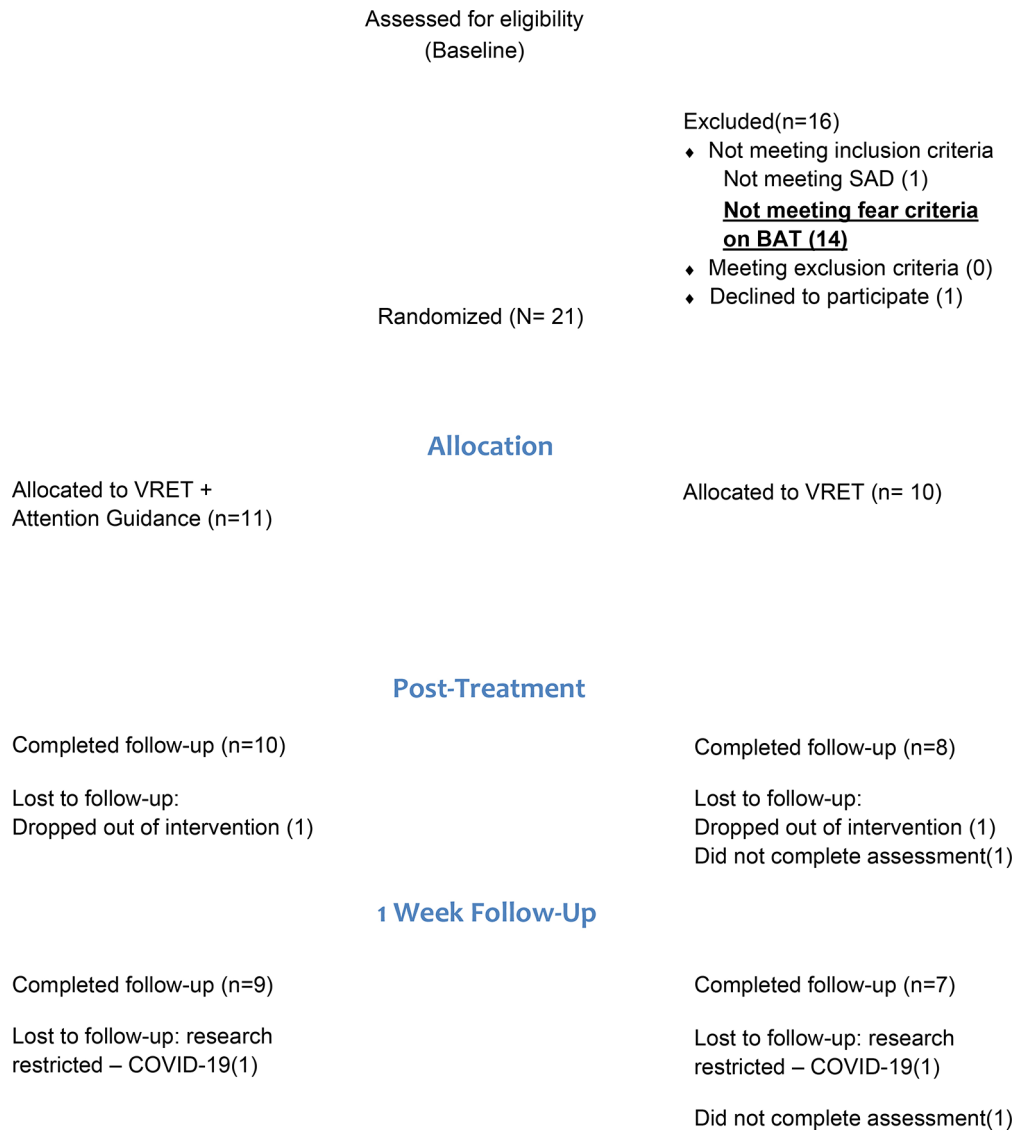


Figure 2.
Flow diagram of participant enrollment.

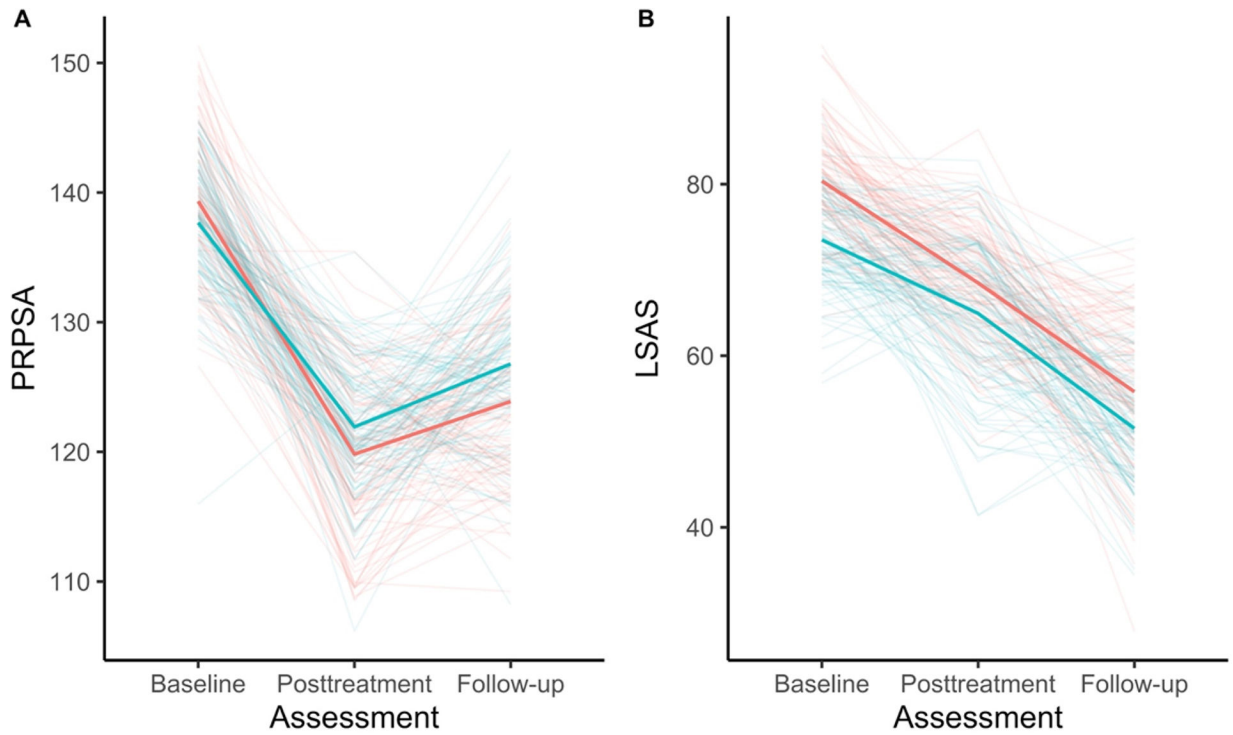


Figure 3.

This figure depicts the effects of the intervention on (A) the primary outcome (fear of public speaking) and (B) the secondary outcome (general social anxiety symptoms). Solid lines reflect the median effect for each intervention group. We included 100 draws of the posterior distribution for each group, which are lightly shaded. There is anecdotal evidence to support no differences (the null hypothesis) at posttreatment and follow-up.

Note. PRPSA = Personal Report of Public Speaking Apprehension; LSAS = Leibowitz Social Anxiety Scale.

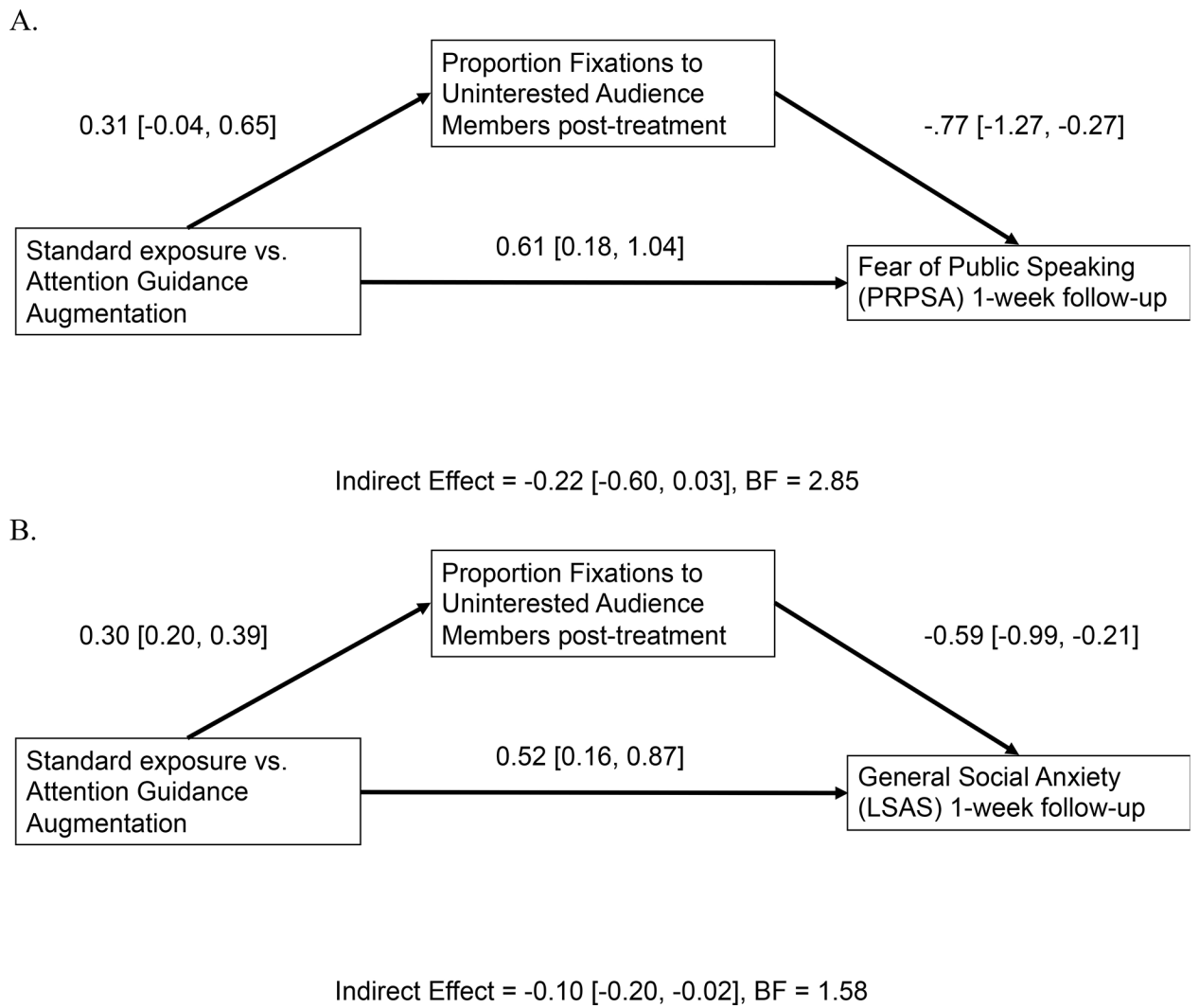


Figure 4. Multilevel mediation models evaluating the indirect effect of changes in fixations to uninterested audience members on symptoms of (A) fear of public speaking and (B) general social anxiety symptoms. Model estimates are partially standardized.

Table 1.

Participant Demographics

	Standard Exposure (n = 10)	Exposure Augmentation (n = 11)
	M (SD)	M (SD)
Age	19.20 (1.23)	25.90 (15.42)
PRPSA	139.40 (14.30)	140.27 (16.85)
LSAS	79.70 (25.64)	75.27 (22.32)
SATI	83.40 (18.19)	86.55 (17.04)
Pre-Fear	61.50 (16.28)	65.36 (17.31)
Peak Fear	66.20 (14.09)	70.64 (17.08)
Post Fear	50.60 (29.67)	62.18 (29.25)
	N (%)	N (%)
Female	5 (50)	8 (73%)
Hispanic/Latinx	5 (50)	6 (55)
Race		
American Indian or Alaska Native	1 (10)	0
Asian	3 (30)	1 (9)
Black or African American	0	1 (9)
White	6 (60)	8 (73)
Current Tx	0 (0)	2 (18)

Note. There were no meaningful differences between groups at baseline. Comparisons are provided in the supplemental materials. One participant in the exposure augmentation group declined to provide demographic information. PRPSA = Personal Report of Public Speaking Apprehension; LSAS = Leibowitz Social Anxiety Scale; SATI = Speech anxiety Thoughts Inventory; Current Tx = currently receiving psychotherapy (excluding CBT for social anxiety disorder).

Table 2.

Symptoms of Social Anxiety and Proportion of Fixations Across Intervention Group and Assessment

		Standard Exposure		Exposure Augmentation	
<i>Measure (summed score)</i>		M	SD	M	SD
PRPSA	Baseline	139.40	14.30	140.27	16.85
	Post-Treatment	119.63	19.54	123.50	21.21
	Follow-up	123.71	17.01	123.56	18.98
LSAS	Baseline	79.70	25.64	75.27	22.32
	Post-Treatment	65.88	27.27	64.8	21.29
	Follow-up	55.57	15.08	51.89	28.18
SATI	Baseline	83.40	18.19	86.55	17.04
	Post-Treatment	70.12	19.69	73.8	22.11
	Follow-up	60.43	18.30	64.22	20.45
<i>Subjective Units of Distress</i>		M	SD	M	SD
Pre-Fear	Baseline	61.50	16.28	65.36	17.31
	Post-Treatment	38.38	20.51	33.10	21.12
	Follow-up	26.29	11.70	48.78	26.48
Peak Fear	Baseline	66.20	14.09	70.64	17.08
	Post-Treatment	32.25	20.31	21.40	18.85
	Follow-up	24.57	10.50	37.22	18.34
Post-Fear	Baseline	50.6	29.67	62.18	29.25
	Post-Treatment	15.13	14.88	17.00	15.72
	Follow-up	11.57	6.43	24.67	9.54
<i>Audience Members (proportion)</i>		M	SD	M	SD
Uninterested	Baseline	10.42	6.74	11.69	7.91
	Post-Treatment	6.69	6.53	22.42	9.59
	Follow-up	15.38	11.65	18.35	10.69
Neutral	Baseline	11.73	7.35	10.96	5.42
	Post-Treatment	14.49	6.80	26.74	11.44
	Follow-up	5.70	5.22	19.68	12.83
Interested	Baseline	8.31	5.70	8.40	6.07
	Post-Treatment	3.29	3.49	10.17	6.43
	Follow-up	12.23	8.39	20.89	9.69

Note. PRPSA = Personal Report of Public Speaking Apprehension; LSAS = Leibowitz Social Anxiety Scale; SATI = Speech Anxiety Thoughts Inventory