

RUNNING HEAD: Targeting youth social avoidance with neurofeedback

**Training negative connectivity patterns between the dorsolateral prefrontal cortex and amygdala through fMRI-based neurofeedback to target adolescent socially-avoidant behaviour**

STEPHEN LISK<sup>1</sup>, KATHRIN COHEN KADOSH<sup>2,3</sup>, CATHARINA ZICH<sup>2,4</sup>, SIMONE P W HALLER<sup>2</sup>, & JENNIFER Y F LAU<sup>1</sup>

<sup>1</sup>*Department of Psychology, Institute of Psychiatry Psychology & Neuroscience, King's College London, London SE5 8AF, UK*

<sup>2</sup>*Department of Experimental Psychology, University of Oxford, Oxford, OX1 3UD, UK*

<sup>3</sup>*School of Psychology, University of Surrey, Guildford GU2 7XH, UK*

<sup>4</sup>*Department of Psychology, University of Oldenburg, 26111 Oldenburg, Germany*

Correspondence should be sent to: Dr Jennifer Lau<sup>1</sup>

---

<sup>1</sup> Psychology Department, Denmark Hill, London SE5 8AF

## **Abstract**

Social anxiety is prevalent in adolescence. Given its role in maintaining fears, reducing social avoidance through cognitive reappraisal may help attenuate social anxiety. We used fMRI-based neurofeedback (NF) to increase ‘adaptive’ patterns of negative connectivity between the dorsolateral prefrontal cortex (DLPFC) and the amygdala to change reappraisal ability, and alter social avoidance and approach behaviours in adolescents. Twenty-seven female participants aged 13-17 years with varying social anxiety levels completed a fMRI-based NF training task where they practiced cognitive reappraisal strategies, whilst receiving real-time feedback of DLPFC-amygdala connectivity. All participants completed measures of cognitive reappraisal and social approach-avoidance behaviour before and after NF training. Avoidance of happy faces was associated with greater social anxiety pre-training. Participants who were unable to acquire a more negative pattern of connectivity through NF training displayed significantly greater avoidance of happy faces at post-training compared to pre-training. These ‘maladaptive’ participants also reported significant decreases in re-appraisal ability from pre to post-training. In contrast, those who were *able* to acquire a more ‘adaptive’ connectivity pattern did not show these changes in social avoidance and re-appraisal. Future research could consider using strategies to improve the capacity of NF training to boost youth social-approach behaviour.

**Keywords:** social anxiety, neurofeedback, fMRI, cognitive appraisal, adolescence

**Abbreviations:** NF (neurofeedback), fc-NF (functional connectivity-based neurofeedback), ER (emotion regulation), AAT (approach avoidance task), DLPFC (dorsolateral prefrontal cortex)

Social anxiety is common in young people (Beesdo et al, 2009). Normative social fears and concerns arise across adolescence (Sumter, Bokhorst & Westenberg, 2009) and clinically-impairing social anxiety is often also first diagnosed during this transitional period (Beesdo, Knappe & Pine, 2009). Socially-anxious young people often avoid feared social situations (Miers, Blöte, Heyne & Westenberg, 2014). Indeed, self-reported questionnaires and experimental tasks such as the approach-avoidance task (AAT; Rinck & Becker, 2007) indicate greater social withdrawal and avoidant behaviours in socially-anxious adults (Heuer, Rinck & Becker, 2007) and adolescents (Klein, Becker & Rinck, 2011). Behavioural avoidance, while reducing short-term feelings of anxiety to negative social evaluation, can be a maladaptive long-term coping mechanism (Aldao, Nolen-Hoeksema & Schweizer, 2010), detrimental to academic, personal and social development (Rao et al., 2007). This is because it prevents any natural extinction of fears that may occur through exposure, instead allowing individuals to attribute the non-occurrence of feared outcomes to safety behaviours and restricting the opportunity to challenge irrational interpretations of the event (McManus, Sacadura & Clark, 2008). Managing social anxiety in adolescence thus involves reducing avoidance of social stimuli (Silverman et al., 1999).

Boosting cognitive reappraisal ability to reduce social avoidance in adolescents is a viable strategy (Lisk, Pile, Haller, Kumari & Lau, 2018). Throughout development, individuals learn to appraise social information, and by adolescence, individual differences in the endorsement of negative versus benign/positive interpretations of ambiguous events is known to consistently influence (social) anxiety symptoms (Haller et al., 2016; Stuijtzand et al., 2018) and behavioural tendencies of avoidance (Hofmann, Heering, Sawyer & Asnaani, 2009; Garnefski, Legerstee, Kraaij, Kommer & Teerds, 2002; Garnefski, Van Den Kommer et al., 2002). While cognitive techniques developed to shift emotional appraisals, from focusing on negative explanations and outcomes to alternative benign/positive ones) can reduce social avoidance and/or increase social approach (Narr & Teachman, 2017), implementing these techniques in youth through training has elicited more mixed findings (Cristea, Mogoșe, David & Cuijpers, 2015) (Krebs et al., 2018), underscoring a need for alternative methods to boost their effects.

One approach is to target the neural substrates of cognitive reappraisal. Extensive human work suggests that while the amygdala plays a key role in fear and salience processing (Adolphs, 2002), areas of the prefrontal cortex (PFC) are proposed to have a top-down, regulatory role in relation to amygdala activation by deploying cognitive strategies such as reappraisal (Ochsner & Gross, 2005). In particular, negative patterns of correlated activity between (increased) lateral PFC and (decreased) amygdala activity have been demonstrated in psychiatrically-healthy individuals during tasks involving emotion regulation (ER), including cognitive reappraisal (Ochsner, Bunge, Gross & Gabrieli, 2002). In contrast, weaker patterns of negative correlations between these regions associate with various psychiatric disorders, including anxiety in adults and social anxiety in adolescents (Prater, Hosanagar, Klumpp, Angstadt & Luan Phan, 2013) (although different patterns of perturbations in amygdala-PFC connectivity have sometimes been found in anxious adolescents; Gold et al., 2016). Regardless, boosting stronger patterns of “negative connectivity” between these regions that resemble those of psychiatrically-healthy adults (rather than anxious adults) could benefit cognitive reappraisal; indeed this this circuitry is altered following reappraisal training (e.g., Månsson et al., 2013; Goldin et al., 2013; Goldin et al., 2014; Young et al., 2017). Establishing these neurocognitive patterns in adolescence could be more optimal, as brain networks responsible for emotion regulation (ER) in particular top-down cognitive strategies are going through a vital period of development during adolescence (Paus, 2005). More specifically, there appears to be a turning point in adolescence where the nature of connectivity between regions of the PFC and amygdala changes from positive connectivity to the desired negative connectivity as children mature beyond 10 years of age (Gee et al., 2013). These developmental changes could provide a window of flexibility for learning external strategies to cultivate adaptive patterns of connectivity to impact cognitive reappraisal associated with adaptive ER (Haller, Cohen Kadosh, Scerif & Lau, 2015; Ahmed, Bittencourt-Hewitt & Sebastian, 2015).

In this study, we used a novel brain training approach, real-time fMRI-based neurofeedback (NF), to reinforce more adaptive patterns of connectivity between the DLPFC and amygdala. NF utilizes the latest developments of real-time data analysis (Johnston, Boehm, Healy, Goebel & Linden, 2010) enabling participants to monitor the relevant activity and connectivity of specific brain

areas to learn to self-regulate their brain responses and associated ER strategies (Koush et al., 2015). The suitability of this approach to modulate the underlying networks in the developing brain has already been demonstrated with data showing that children and adolescents could be taught to regulate activity in ER regions through NF (Cohen Kadosh et al., 2016). Moreover, self-regulation effects were not limited to the NF target region, but affected the overall ER network. In a second study, researchers used functional connectivity-based NF (fc-NF) to directly modulate ER network connectivity in girls aged 14-17 years (Zich et al., 2016). Participants were trained to modulate the functional coupling of the PFC and the amygdala towards a more negative connectivity pattern, which resembles the connectivity pattern found in the mature adaptive/healthy brain (Gee et al., 2013; Prater et al., 2013), with individual differences in responsiveness to NF.

No research has investigated how training adaptive connectivity patterns influence anxiety-relevant behaviours outside the scanner, such as social avoidance. The current study investigates whether the provision of feedback on patterns of connectivity between the amygdala and DLPFC can affect socially-avoidant behaviours in adolescents. Given individual differences in the degree to which these co-activation patterns can respond to fc-NF (i.e. become more negative), the primary hypothesis is that, amongst those who are responsive to training, there should be a significant reduction in socially-avoidant behaviours, and significant improvement in cognitive reappraisal ability. However, given prior findings of weaker (negative) functional connectivity of the amygdala and DLPFC in socially-anxious individuals (Prater et al., 2013), we also tested whether there would be associations between social anxiety symptoms and socially-avoidant behaviour with these co-activation patterns at baseline. To index socially-avoidant behaviour, we used the AAT (Rinck & Becker, 2007). The AAT requires the participant to engage in fast approach and avoid actions to social stimuli (emotional faces) via a joystick. Reaction times across different task conditions index the degree to which individuals avoid socially-aversive stimuli and approach socially-appetitive stimuli (Phaf, Mohr, Rotteveel & Wicherts, 2014) with socially-anxious individuals showing a greater tendency to avoid emotional faces than non-socially-anxious individuals (Heuer et al., 2007; Roelofs et al., 2010). As well as probing individual differences in social anxiety (Rinck et al., 2009), the AAT has also been used in studies to measure changes in social approach/avoidance following

neuroendocrine challenges (Enter, Spinhoven & Roelofs, 2014) and psychological manipulations (Voncken, Rinck, Deckers & Lange, 2011), as well as more specifically, interventions inducing different interpretational styles (Lange et al., 2010).

## Methods

### Participants and procedures

This study was registered as preclinical trial #NCT02463136. Forty-six female participants (mean age=15.09 years; SD=1.18 years; range: 13-17 years) were recruited from local schools in Oxfordshire. Nineteen of the initial 46 participants received other NF implementations (positive or non-weighted negative) as described in Zich et al., 2020. Therefore this report focuses on the 27 participants (mean age=15.22 years; SD=1.22 years; range: 13-17 years) who received real-time feedback of negative patterns of functional connectivity between the amygdala and DLPFC. All participants had normal or corrected-to-normal vision and no self-reported history of neurological and psychiatric disorders. This study was approved by the Oxford University Research Ethics Committee. Informed consent and assent were obtained from the primary caregiver or young person themselves. Participants completed self-reported questionnaires on social anxiety and emotion regulation, before completing the AAT. They were then prepared for fMRI scanning and provided with instructions for the in-scanner tasks. To identify key emotion regulation regions of the brain to be used in the NF task, a reappraisal task (“localiser”) was conducted inside the scanner. Following this, the NF training was completed. The participant then left the scanner and repeated the AAT, followed by a full debrief. Participants received gift vouchers for taking part.

### Measures

***Social anxiety (pre-training only):*** Participants rated social anxiety symptoms using the 22-item Social Anxiety Scale for Adolescents (SAS-A; La Greca & Lopez, 1998), a 22-item self-report measure. Each item is (e.g. ‘I worry about doing something new in front of others’) is rated on a 5-point Likert-scale from 0 (not at all) to 5 (all the time). Internal consistency was  $\alpha = 0.81$ .

**Cognitive reappraisal (pre and post-training):** Two items from the Cognitive Emotion Regulation Questionnaire (CERQ; Garnefski & Kraaij, 2006) assessed cognitive reappraisal (“I think I can learn something from the situation”, “I think that I can become a stronger person as a result of what has happened”). Items were rated on a 5-point Likert scale ranging from 1 ((almost) never) to 5 ((almost) always).

**Approach Avoidance Task (pre and post-training):** The AAT (Rinck & Becker, 2007) tests automatic behavioural avoidance tendencies to emotional faces (Figure 1). Participants were asked to react to a single picture on the centre of the screen (an adult face with either a happy or angry expression, and a gaze of either straight or averted left/right), by pulling or pushing a joystick (with their dominant hand) in the instructed direction, as quickly and accurately as possible. Upon movement of the joystick, the picture grew or shrunk in size (depending on push or pull) creating the impression of movement towards (approach) or away (avoidance). When the joystick reached 30° in the intended direction the picture disappeared and the joystick had to be returned to the centre position and the ‘fire’ button pressed for the next trial to begin. The task consisted of two blocks of 64 trials (each block preceded by 18 practice trials). In the congruent block, participants were instructed to pull happy faces toward them and push angry faces away, whereas in the incongruent block, participants pushed away happy faces and pulled angry faces towards them. The block order was counterbalanced across participants. Reaction times (RTs) were recorded. For data analysis, time between stimulus onset and the maximum joystick displacement (30°) was used (Radke, Roelofs & De Bruijn, 2013). Trials in which participants moved the joystick to maximum joystick displacement in the incorrect direction were recorded as errors.

## **fMRI tasks**

**Image acquisition:** fMRI data acquisition was performed using a 3T Siemens MAGNETOM Prisma MRI scanner (Siemens AG, Erlangen, Germany) using a standard 32-channel head matrix coil. Prior to the functional tasks a high-resolution structural volume was obtained from each subject using a T1-weighted magnetization-prepared rapid-acquisition gradient echo (MPRAGE) sequence (TR=1900 ms, TE=3.97 ms, FoV=192x192 mm<sup>2</sup>, flip angle=8°, slice thickness=1 mm, sagittal).

Functional measures comprised the localizer task and four NF training runs. The localizer comprised 570, and each NF run 310, 2D multiband gradient echo planer imaging volumes (Todd et al., 2016) (2.0x2.0x2.0 mm voxels, 0.57 mm gap, TR=933 ms, TE=33.40 ms, FoV=192x192 mm<sup>2</sup>, flip angle=64°, 72 slices, Multi-band factor=6, fat saturation, transverse slices with phase encoding in the A>>P direction). To avoid saturation effects, 10 additional volumes were acquired but not analysed at the beginning of the localizer task and each NF run.

***Localiser Task:*** A reappraisal task (Figure 2; Haller, Raeder, Scerif, Kadosh & Lau, 2016) was used to train the individual with cognitive reappraisal strategies and prompt activation of specific brain regions involved in emotion regulation for use in the NF task. During each trial, participants were shown a picture of a social scene from the perspective of a female adolescent approaching the scene, depicted from the back. Each scene connoted themes of negative peer evaluation (appraisal event), which the participant was instructed to appraise freely. Following appraisal, the participant was presented with a positively valenced explanation of the scene. Participants were then shown the scene again and asked to attempt to re-interpret it in the direction of the explanation (reappraisal). This was repeated for 30 trials, with an inter-trial interval displaying a fixation cross for 0.93 seconds. The task lasted a total of 9 minutes. Based on the brain activity maps yielded from this task, regions involved in emotion regulation were selected for participant feedback during NF training.

### ***Neurofeedback Training***

NF training comprised four runs, each lasting 4.8 minutes (Figure 3). Each run started with a fixation cross in the centre of the screen for 18.66 seconds, which the participant was instructed to focus their attention on. During each run the participant received seven blocks of NF (each lasting 18.66 seconds), during which they saw continuous feedback of brain activity via a simple picture of a ‘10-point thermometer’. The number of segments filled in provided the participant with real-time indication of the functional-connectivity between the target regions first defined using the localiser task; this was the negative partial correlation between DLPFC and amygdala relative to an unrelated brain region (a white matter region of the left corticospinal tract). Participants were given the following instructions: “You will see a thermometer with a green rim on the screen. The red bars

show how much the regions that are important for emotions are active. Your job is to get these regions as active as possible! So, try to get this thermometer up as much as possible. Similar to the task before, try to control your thoughts towards a positive feeling. When the thermometer does not have a green rim, the thermometer is not working. However, even if the thermometer is not working, your task will be the same and we are still measuring how much your brain is active. The two different thermometers will alternate.” Participants also received seven no-NF blocks in each run, during which participants were asked to continue with the same strategies they were using during the NF blocks, but that the thermometer would be frozen at point six. NF and no-NF blocks were presented in alternating order. To allow participants to differentiate between NF and no-NF blocks, the thermometer in the NF runs was presented with a green frame around it, whereas during the no-NF blocks this frame was missing.

### **Statistical analysis plan**

*AAT Data cleaning and extraction:* As per previous research using the AAT (Heuer et al., 2007; Marsh, Ambady & Kleck, 2005; Roelofs et al., 2010), RT outliers were filtered using lower and upper cut-offs of 150ms and 1500ms, respectively. Following this, a cut-off of three standard deviations from the mean was used to remove outliers. Incorrect responses were also removed. 91.4% of responses remained, for which medians were calculated per cell (defined by Emotion, Gaze and Action). No differences between conditions were observed for error rates. As per previous studies (Enter, Spinhoven & Roelofs, 2014; Heuer et al., 2007; Roelofs et al., 2010), effect-scores were calculated as an index of approach/avoidance tendencies. These were calculated by subtracting the mean RTs for pull movements from the mean RTs for push movements for each emotion per individual. Therefore, negative effect-scores indicate stronger avoidance tendencies and positive effect-scores indicate stronger approach tendencies for each emotion. As there were no significant differences between straight and averted gaze conditions for each Emotion (all  $t$ 's < .23, and all  $p$ 's > .82), we combined mean RTs for each gaze direction across Action-Emotion combinations.

***fMRI functional connectivity data processing:*** Full MRI data processing and analysis are described by Zich and colleagues (2020). In brief, [functional images obtained during the localiser task and the NF runs were processed in real-time using Turbo-BrainVoyager 3.2 \(Brain Innovation, Maastricht, The Netherlands\). After correcting for head motion and smoothed, three a priori defined ROIs \(voxel size = 12 mm<sup>3</sup>, 6 x 6 x 6\) based on the localiser task were manually placed. ROIs were based on a group average activation ascertained from an independent contrast \(and therefore did not vary across individuals\). GLM t-statistics of the brain activity during the localizer task, i.e. the sum of the three contrasts: appraisal > fixation, reappraisal > fixation, reappraisal > appraisal \(threshold  \$t = 3\$ \), was projected onto the processed structural scan. The local maximum of the t-statistics within the left dorsolateral and medial PFC constituted the centre of the PFC ROI. Similarly, the local maximum of the t-statistics within the left amygdala constituted the centre of the amygdala ROI. A ROI in the left corticospinal tract \(CST\) served as control ROI. During the NF task, PFC-amygdala fc was calculated in real-time. PFC-amygdala fc was defined as the partial correlation between PFC and amygdala activity, while controlling for CST 'activity'. Partial correlations were based on a moving window, which was updated with every incoming volume. The length of the correlation window was 20 volumes. Calculations were performed using a custom-made plugin for Turbo-BrainVoyager, which also provided a direct TCP/IP based link between the real-time analysis software and the stimulus application BrainStim.](#)

Initial functional connectivity for each participant was defined as the average partial correlation of the first two blocks of the first NF run (as described by Zich and colleagues (in press)). To determine which participants were able to acquire an adaptive pattern of connectivity in response to NF training versus those who were not, we created a training direction variable. This was quantified as the slope of the linear regression for average functional connectivity (i.e. the partial correlation between DLPFC and amygdala relative to a white matter region of the left corticospinal tract), from runs 1 to 4 for each participant. Using this slope variable, we divided participants into those who had a gradient of increasingly negative connectivity across runs (average slope = -.023, SD = .019, range: -0.058 to -0.000058) as those who were able to acquire an *adaptive* pattern (n=13) and participants who had a gradient of increasingly positive connectivity across runs (average

slope=.031, SD=.025, range: 0.0091 to 0.079) as those who acquired a *maladaptive* pattern (n=14). As one participant from the ‘adaptive’ group had a slope value close to zero (-0.000058), this could reflect difficulties in attaining the desired pattern of connectivity – which in turn could dilute any changes in socially-avoidant behaviour in this group. Removal of this participant from our key analysis, however, did not alter the pattern of results.

**Statistical analysis:** To assess the validity of the AAT, we tested whether approach-avoidance tendencies toward each emotion (Angry/Happy), pre-NF, were in the predicted direction. To do this, we conducted a two-way repeated-measures analysis of variance (ANOVA) with Action (Push/Pull) and Emotion (Angry/ Happy) as the within subject variables. Next, we investigated whether these behaviours were associated with social anxiety by calculating correlations between AAT effect scores for each emotion and SAS-A scores at pre-training. To test the hypothesis that there would be associations between social anxiety symptoms and socially-avoidant behaviour with less negative amygdala-DLPFC co-activation patterns, we calculated the correlation between initial functional connectivity (FC) and social anxiety scores and AAT effect scores pre-NF. To evaluate if there was a significant reduction in socially-avoidant behaviours, the effects of training direction group on changes in AAT effect scores were analysed in a 2x2x2 mixed ANOVA, with Time (pre, post) and Emotion (happy, angry) as within-subjects variables and training direction group (adaptive, maladaptive) as the between-subject variable. To examine changes in positive reappraisal from pre to post-NF, depending on training direction group, reappraisal scores at pre and post-NF were analysed with a mixed ANOVA of group direction variable (adaptive, maladaptive) and time (pre, post). In all analyses, we controlled for age.

Of note, all analysis conducted on pre-NF variables were done with 45 participants, as one participant was excluded from analysis due to RTs on the AAT deviating over three standard deviations from the group mean. Initial FC data was unable to be collected for 6 participants, therefore all analysis conducted with initial FC was performed with 40 participants. Any analysis assessing changes as a function of NF training were conducted using the 27 participants who received real-time feedback of negative patterns of functional connectivity between the amygdala and DLPFC. The excluded AAT participant was within this group, thus all analysis of NF effects on AAT

performance was conducted with 26 participants. In all analyses, alpha was set at .05, and effect sizes are reported as partial eta-squared ( $\eta_p^2$ ) or Cohen's *d*. Bonferroni adjustment controlled for type 1 error in analyses where multiple ANOVAs were conducted, with adjusted *p*-values reported.

## Results

### AAT performance and correlations with social anxiety

The mean SAS-A score across participants was 50.11 (SD=14.18). Figure 4 shows the mean RTs for each emotion-action combination for pre- and post-NF measures.

The 2x2 ANOVA with Emotion (Angry and Happy) and Action (Push and Pull) showed no significant main effect of Emotion ( $F(1,44)=.102, p=.752, \eta_p^2=.00$ ) but a significant main effect of Action ( $F(1,44)=5.92, p=.019, \eta_p^2=.12$ ), and a significant interaction between them ( $F(1,44)=6.67, p=.013, \eta_p^2=.13$ ). Tests of simple main effects showed that RT means were not significantly different between Happy-Push and Happy-Pull trials ( $F(1,44)=1.4, p=.243, \eta_p^2=.03$ ) but were significantly faster for Angry-Push than Angry-Pull trials ( $F(1,44)=10.93, p=.002, \eta_p^2=.20$ ). Moreover, RT means were faster for Happy-Pull than Angry-Pull trials ( $F(1,44)=6.18, p=.017, \eta_p^2=.12$ ), and also significantly faster for Angry-Push than Happy-Push trials ( $F(1,44)=4.45, p=.039, \eta_p^2=.09$ ).

Angry effect scores were not significantly associated with social anxiety ( $r(45)=-.07, p=.669$ ), but happy effect scores negatively correlated with symptoms ( $r(45)=-.33, p=.029$ ) such that individuals with higher social anxiety avoided happy faces more.

### Correlations between social anxiety and socially-avoidant behaviours, and initial FC

There was a non-significant negative correlation between initial FC and AAT happy effect scores ( $r(40)=-.30, p=.064$ ), but a significant positive correlation between initial FC and AAT angry effect scores emerged ( $r(40)=.32, p=.047$ ), suggesting that initial FC is more positive in individuals who approach angry faces more. The correlation between SAS-A score and initial FC was not significant ( $r(40)=.19, p=.246$ ). These results held true for the smaller sample that received NF training. Age (in months) and initial FC were not significantly correlated ( $r(40)=-1.12, p=.46$ ).

### **Pre-to-post NF changes in socially-avoidant behaviour depending on neurofeedback training ability**

The 2x2x2 mixed ANOVA, with age as a covariate, showed a significant 3-way interaction between time, emotion and training direction group ( $F(1,23)=8.83$ ,  $p=.007$ ,  $\eta^2=.28$ ). The 2x2x2 mixed ANOVA showed a significant 3-way interaction between time, emotion and training direction group ( $F(1,23)=8.83$ ,  $p=.007$ ,  $\eta^2=.28$ ). To decompose this interaction a Time-by-Emotion ANOVA was conducted for each group. For the maladaptive group, neither the main effects of Time or Emotion were significant, however the Time-by-Emotion interaction was significant ( $F(1,12)=5.84$ ,  $p=.034$ ,  $\eta^2=.35$ ). To explore this two-way interaction, we assessed the main effect of time for each emotion in the maladaptive group: Happy effect scores significantly decreased from pre ( $M=49.46$ ,  $SD=70.87$ ) to post ( $M=3.04$ ,  $SD=106.62$ ) NF training ( $F(1,11)=5.78$ ,  $p=.035$ ,  $\eta^2=.34$ ), suggesting that avoidance of happy faces increased. Of note, decomposing the Happy effect scores into raw scores for Happy push and Happy pull (Supplementary Table 1) showed that this change was driven by reduced reaction times in pushing away Happy faces ( $t(12)=3.51$ ,  $p=.004$ ) with no significant change in pulling Happy faces towards ( $t(12)=0.99$ ,  $p=.340$ ). Angry effect scores were not significantly different pre ( $M=-54.96$ ,  $SD=64.28$ ) to post ( $M=-20.69$ ,  $SD=107.13$ ) NF training ( $F(1,11)=.57$ ,  $p=.47$ ,  $\eta^2=.05$ ). For the adaptive direction group, there were no main effects of Time ( $F(1,11)=0.02$ ,  $p=.88$ ,  $\eta^2=.00$ ), Emotion ( $F(1,11)=0.25$ ,  $p=.63$ ,  $\eta^2=.02$ ) or an interaction effect, ( $F(1,11)=.73$ ,  $p=.41$ ,  $\eta^2=.06$ ).

### **Pre-to-post NF changes in self-report positive reappraisal depending on neurofeedback training ability**

There was a main effect of Time ( $F(1,24)=4.76$ ,  $p=.01$ ,  $\eta^2=.24$ ). Although there was no significant interaction with Training direction ( $F(1,24)=2.29$ ,  $p=.14$ ,  $\eta^2=.09$ ), due to a priori expectations around changes in reappraisal, we nonetheless carried out separate paired sample t-tests for each group on the pre and post-NF variables. The reduction in positive appraisal from pre ( $M=6.50$ ,  $SD=2.10$ ) to post ( $M=5.71$ ,  $SD=2.37$ ) in the maladaptive group was significant after

correction for multiple comparisons ( $t(13)=2.80$ ,  $p=.015$ ,  $d=.53$ ) but not amongst the adaptive group (pre:  $M=6.92$ ,  $SD=2.10$ ; post:  $M=6.70$ ,  $SD=1.97$ ;  $t(12)=0.61$ ,  $p=.553$ ,  $d=.24$ ).

### Discussion

The current study tested whether real-time fMRI neurofeedback (NF) could target neural correlates of emotion regulation (ER) and alter socially avoidant behaviour in unselected adolescents. We assessed whether any improvement in cognitive reappraisal abilities, targeted through NF of functional connectivity between the DLPFC and amygdala, would also be observed. Amongst individuals unable to acquire an adaptive pattern of connectivity (an increasingly negative connectivity between the DLPFC and amygdala) through NF training, there was an increasing tendency to avoid happy faces from pre to post-NF training. These same participants showed a significant decrease in self-report positive appraisal ability following NF training. These changes were absent in the group able to acquire an adaptive pattern of connectivity. As expected, all individuals pulled happy faces faster than angry faces and pushed angry faces faster than happy faces, but social approach-avoidance tendencies differed amongst individuals with social anxiety: Those with higher social anxiety showed greater avoidance of happy faces than those with lower levels. Initial positive functional connectivity between the amygdala and DLPFC was associated with the tendency to approach angry faces (but not with social anxiety).

The findings show that among individuals who did not respond to fc-NF training, socially avoidant behaviour (of happy faces becomes more pronounced with a more exaggerated reduction in reappraisal ability (relative to baseline). These results fall in line with studies showing that an absence of the “adaptive” connectivity is correlated with social anxiety in adults (Kim et al., 2011). They also illustrate a potentially aversive consequence of failing to engage with the NF technique. Whilst our results did not show improvement of avoidant behaviour in the “adaptive” group, they do demonstrate the absence of deteriorations. It may be the case that with further development of this relatively exploratory training technique, possibly administered over multiple sessions to identify and consolidate effective strategies, we could see positive changes in cognitive reappraisal, reductions in withdrawal behaviour and increases in approach behaviour among the ‘adaptive’ group. However, for the ‘maladaptive’ group, multiple sessions could exaggerate the aversive effect associated with an

inability to acquire training. This suggests future research should first focus on the salient factors differentiating responders and non-responders in order to ensure training approaches can be adapted appropriately.

Perhaps due to the non-clinical nature of the sample, social anxiety correlated not with avoidance of angry faces (Roelofs et al., 2010) but happy faces. Evolutionary-based avoidance tendency for angry faces may exist in all individuals regardless of anxiety (Marsh et al., 2005), however, due to the nature of social anxiety - where individuals are excessively concerned about negative social evaluation - differences in avoidant behaviour may only become apparent to happy faces, due to the distinct lack of threat interpretation to these stimuli by healthy compared with anxious individuals. Previous studies have reported that anxious individuals avoid happy faces, at automatic and controlled levels of processing (Heuer et al., 2007; Mansell, Clark, Ehlers & Chen, 1999). Finally, when investigating how initial connectivity may impact existing approach-avoidance behaviours, we found a greater likelihood to approach angry faces in those individuals with a 'maladaptive' pattern of connectivity between DLPFC and amygdala. This unexpected finding may signal a maladaptive tendency that could be linked to other behavioural problems such as aggression, though without these measures, we were unable to assess this.

There are some study limitations. Firstly, although we measured reappraisal before and after NF training, we did not know whether participants were deploying reappraisal ability during NF training. Moreover, our measure of reappraisal was based on only two items from the CERQ, a measure that typically measures emotion regulation strategies as traits and may therefore be less amenable to change following a single training session. Future research should measure state aspects of reappraisal to give a more accurate picture over whether changes in cognitive reappraisal ability were targeted during NF training, and mediated changes in social avoidance. Second, the modest sample size in each group may have limited the significance of some statistical comparisons. We also used an all-female sample. This was to limit homogeneity in the sample given many age and gender effects in adolescence but restricts generalizability of the findings to all adolescents. Similarly, data from unselected adolescents cannot inform whether the same findings characterise clinically-anxious

patients. Finally, the degree to which AAT reflects real-life behavioural avoidance of social situations is questionable.

### **Conclusions**

This study demonstrated that in our non-clinical sample of female adolescents, socially avoidant behaviour of positive faces was associated with greater social anxiety. When using neurofeedback training to alter cognitive reappraisal, we found that those who were unable to acquire a more adaptive pattern of connectivity showed increased avoidance of happy faces and decrease in positive appraisal ability. These results suggest that NF training can have a differential effect on cognitive reappraisal ability and subsequent social approach-avoidance tendencies, however, at present this effect is not in the desired direction. Further research is required to understand factors differentiating individuals who show difficulties responding to training from those who do respond. Using our initial findings on the proportion of young people who do and do not respond to training and the effect sizes of within-group changes in social avoidance in each group, future studies can conduct larger and powered studies to identify the optimal parameters for positive outcomes.

Acknowledgements

*Funding:* This work was supported by the European Commission FP7 Braintrain grant (602186) and by a studentship to Stephen Lisk from the Medical Research Council (MR/K50130X/1).

## References

- Adolphs, R. (2002) 'Recognizing emotion from facial expressions: psychological and neurological mechanisms', *Behav Cogn Neurosci Rev*, 1(1), pp. 21-62.
- Ahmed, S. P., Bittencourt-Hewitt, A. and Sebastian, C. L. (2015) 'Neurocognitive bases of emotion regulation development in adolescence', *Dev Cogn Neurosci*, 15, pp. 11-25.
- Aldao, A., Nolen-Hoeksema, S. and Schweizer, S. (2010) 'Emotion-regulation strategies across psychopathology: A meta-analytic review', *Clin Psychol Rev*, 30(2), pp. 217-37.
- Beesdo, K., Knappe, S. and Pine, D. S. (2009) 'Anxiety and anxiety disorders in children and adolescents: developmental issues and implications for DSM-V', *Psychiatr Clin North Am*, 32(3), pp. 483-524.
- Cristea, I. A., Mogoase, C., David, D. and Cuijpers, P. (2015) 'Practitioner Review: Cognitive bias modification for mental health problems in children and adolescents: a meta-analysis', *J Child Psychol Psychiatry*, 56(7), pp. 723-34.
- Enter, D., Spinhoven, P. and Roelofs, K. (2014) 'Alleviating social avoidance: effects of single dose testosterone administration on approach-avoidance action', *Horm Behav*, 65(4), pp. 351-4.
- Garnefski, N. and Kraaij, V. (2006) 'Cognitive emotion regulation questionnaire - development of a short 18-item version (CERQ-short)', *Personality and Individual Differences*, 41(6), pp. 1045-1053.
- Garnefski, N., Legerstee, J., Kraaij, V., Van Den Kommer, T. and Teerds, J. (2002) 'Cognitive coping strategies and symptoms of depression and anxiety: a comparison between adolescents and adults', *Journal of Adolescence*, 25(6), pp. 603-611.
- Garnefski, N., Van den Kommer, T., Kraaij, V., Teerds, J., Legerstee, J. and Onstein, E. (2002) 'The relationship between cognitive emotion regulation strategies and emotional

problems: Comparison between a clinical and a non-clinical sample', *European Journal of Personality*, 16(5), pp. 403-420.

Gee, D. G., Humphreys, K. L., Flannery, J., Goff, B., Telzer, E. H., Shapiro, M., Hare, T. A., Bookheimer, S. Y. and Tottenham, N. (2013) 'A developmental shift from positive to negative connectivity in human amygdala-prefrontal circuitry', *J Neurosci*, 33(10), pp. 4584-93.

Gold, A. L., Shechner, T., Farber, M. J., Spiro, C. N., Leibenluft, E., Pine, D. S. and Britton, J. C. (2016) 'Amygdala-Cortical Connectivity: Associations with Anxiety, Development, and Threat', *Depress Anxiety*, 33(10), pp. 917-926.

Goldin, P.R., Ziv, M., Jazaieri, H., Hahn, K., Heimberg, R., Gross, J.J., 2013. Impact of cognitive behavioral therapy for social anxiety disorder on the neural dynamics of cognitive reappraisal of negative self-beliefs: randomized clinical trial. *JAMA Psychiatry* 70, 1048-1056.

Goldin, P.R., Ziv, M., Jazaieri, H., Weeks, J., Heimberg, R.G., Gross, J.J., 2014. Impact of cognitive-behavioral therapy for social anxiety disorder on the neural bases of emotional reactivity to and regulation of social evaluation. *Behav. Res. Ther.* 62, 97-106.

Haller, S. P. W., Kadosh, K. C., Scerif, G. and Lau, J. Y. F. (2015) 'Social anxiety disorder in adolescence: How developmental cognitive neuroscience findings may shape understanding and interventions for psychopathology', *Developmental Cognitive Neuroscience*, 13, pp. 11-20.

Haller, S. P. W., Raeder, S. M., Scerif, G., Kadosh, K. C. and Lau, J. Y. F. (2016) 'Measuring online interpretations and attributions of social situations: Links with adolescent social anxiety', *Journal of Behavior Therapy and Experimental Psychiatry*, 50, pp. 250-256.

Heuer, K., Rinck, M. and Becker, E. S. (2007) 'Avoidance of emotional facial expressions in social anxiety: The Approach-Avoidance Task', *Behav Res Ther*, 45(12), pp. 2990-3001.

Hofmann, S. G., Heering, S., Sawyer, A. T. and Asnaani, A. (2009) 'How to handle anxiety: The effects of reappraisal, acceptance, and suppression strategies on anxious arousal',

*Behaviour Research and Therapy*, 47(5), pp. 389-394.

Johnston, S. J., Boehm, S. G., Healy, D., Goebel, R. and Linden, D. E. (2010)

'Neurofeedback: A promising tool for the self-regulation of emotion networks', *Neuroimage*, 49(1), pp. 1066-72.

Kadosh, K. C., Luo, Q., de Burca, C., Sokunbi, M. O., Feng, J. F., Linden, D. E. J. and Lau, J. Y. F. (2016) 'Using real-time fMRI to influence effective connectivity in the developing emotion regulation network', *Neuroimage*, 125, pp. 616-626.

Kim, M. J., Loucks, R. A., Palmer, A. L., Brown, A. C., Solomon, K. M., Marchante, A. N. and Whalen, P. J. (2011) 'The structural and functional connectivity of the amygdala: from normal emotion to pathological anxiety', *Behav Brain Res*, 223(2), pp. 403-10.

Klein, A. M., Becker, E. S. and Rinck, M. (2011) 'Approach and Avoidance Tendencies in Spider Fearful Children: The Approach-Avoidance Task', *J Child Fam Stud*, 20(2), pp. 224-231.

Koush, Y., Meskaldji, D. E., Pichon, S., Rey, G., Rieger, S. W., Linden, D. E. J., Van de Ville, D., Vuilleumier, P. and Scharnowski, F. (2017) 'Learning Control Over Emotion Networks Through Connectivity-Based Neurofeedback', *Cerebral Cortex*, 27(2), pp. 1193-1202.

Krebs, G., Pile, V., Grant, S., Degli Esposti, M., Montgomery, P. and Lau, J. Y. F. (2018) 'Research Review: Cognitive bias modification of interpretations in youth and its effect on anxiety: a meta-analysis', *J Child Psychol Psychiatry*, 59(8), pp. 831-844.

La Greca, A. M. and Lopez, N. (1998) 'Social anxiety among adolescents: Linkages with peer relations and friendships', *Journal of Abnormal Child Psychology*, 26(2), pp. 83-94.

- Lisk, S. C., Pile, V., Haller, S. P. W., Kumari, V. and Lau, J. Y. F. (2018) 'Multisession Cognitive Bias Modification Targeting Multiple Biases in Adolescents with Elevated Social Anxiety', *Cognit Ther Res*, 42(5), pp. 581-597.
- Mansell, W., Clark, D. M., Ehlers, A. and Chen, Y. P. (1999) 'Social anxiety and attention away from emotional faces', *Cognition & Emotion*, 13(6), pp. 673-690.
- Månsson, K. N., Carlbring, P., Frick, A., Engman, J., Olsson, C. J., Bodlund, O., ... & Andersson, G. (2013). Altered neural correlates of affective processing after internet-delivered cognitive behavior therapy for social anxiety disorder. *Psychiatry Research: Neuroimaging*, 214(3), 229-237.
- Marsh, A. A., Ambady, N. and Kleck, R. E. (2005) 'The effects of fear and anger facial expressions on approach- and avoidance-related behaviors', *Emotion*, 5(1), pp. 119-124.
- Mcmanus, F., Sacadura, C. and Clark, D. M. (2008) 'Why social anxiety persists: An experimental investigation of the role of safety behaviours as a maintaining factor', *Journal of Behavior Therapy and Experimental Psychiatry*, 39(2), pp. 147-161.
- Miers, A. C., Blote, A. W., Heyne, D. A. and Westenberg, P. M. (2014) 'Developmental pathways of social avoidance across adolescence: the role of social anxiety and negative cognition', *J Anxiety Disord*, 28(8), pp. 787-94.
- Ochsner, K. N., Bunge, S. A., Gross, J. J. and Gabrieli, J. D. (2002) 'Rethinking feelings: an fMRI study of the cognitive regulation of emotion', *J Cogn Neurosci*, 14(8), pp. 1215-29.
- Ochsner, K. N. and Gross, J. J. (2005) 'The cognitive control of emotion', *Trends in Cognitive Sciences*, 9(5), pp. 242-249.
- Paus, T. (2005) 'Mapping brain maturation and cognitive development during adolescence', *Trends Cogn Sci*, 9(2), pp. 60-8.

Phaf, R. H., Mohr, S. E., Rotteveel, M. and Wicherts, J. M. (2014) 'Approach, avoidance, and affect: a meta-analysis of approach-avoidance tendencies in manual reaction time tasks', *Front Psychol*, 5, p. 378.

Prater, K. E., Hosanagar, A., Klumpp, H., Angstadt, M. and Phan, K. L. (2013) 'Aberrant amygdala-frontal cortex connectivity during perception of fearful faces and at rest in generalized social anxiety disorder', *Depress Anxiety*, 30(3), pp. 234-41.

Radke, S., Roelofs, K. and de Bruijn, E. R. A. (2013) 'Acting on Anger: Social Anxiety Modulates Approach-Avoidance Tendencies After Oxytocin Administration', *Psychological Science*, 24(8), pp. 1573-1578.

Rao, P. A., Beidel, D. C., Turner, S. M., Ammerman, R. T., Crosby, L. E. and Sallee, F. R. (2007) 'Social anxiety disorder in childhood and adolescence: descriptive psychopathology', *Behav Res Ther*, 45(6), pp. 1181-91.

Rinck, M. and Becker, E. S. (2007) 'Approach and avoidance in fear of spiders', *J Behav Ther Exp Psychiatry*, 38(2), pp. 105-20.

Roelofs, K., Putman, P., Schouten, S., Lange, W. G., Volman, I. and Rinck, M. (2010) 'Gaze direction differentially affects avoidance tendencies to happy and angry faces in socially anxious individuals', *Behav Res Ther*, 48(4), pp. 290-4.

Silverman, W. K., Kurtines, W. M., Ginsburg, G. S., Weems, C. F., Lumpkin, P. W. and Carmichael, D. H. (1999) 'Treating anxiety disorders in children with group cognitive-behavioral therapy: a randomized clinical trial', *J Consult Clin Psychol*, 67(6), pp. 995-1003.

Sumter, S. R., Bokhorst, C. L. and Westenberg, P. M. (2009) 'Social fears during adolescence: Is there an increase in distress and avoidance?', *Journal of Anxiety Disorders*, 23(7), pp. 897-903.

Young, K. S., Burklund, L. J., Torre, J. B., Saxbe, D., Lieberman, M. D., & Craske, M. G. (2017). Treatment for social anxiety disorder alters functional connectivity in emotion

regulation neural circuitry. *Psychiatry Research: Neuroimaging*, 261, 44-51.

Zich, C., Haller, S. P., Luehrs, M., Lisk, S., Lau, J. Y., Kadosh, K. C. (2018) 'Modulatory effects of dynamic fMRI-based neurofeedback on emotion regulation networks during adolescence', *bioRxiv*, 347971.

Figure legends

Figure 1. Graphic representation of the Approach Avoidance Task with angry faces.

Figure 2. Graphic representation of the reappraisal task used to localise ER areas of the brain.

Figure 3. Graphic representation of Neurofeedback training.

Figure 4. Mean reaction times (with standard error bars) as a function of Timepoint (Pre, Post), Picture Type (Happy, Angry), and Response Direction (Pull, Push).