Title: Empathy to emotional voices and the use of real-time fMRI to enhance activation of the anterior insula

Author names and affiliations:

Dana Kanel[1], Salim Al-Wasity[1,2], Kristian Stefanov[1] and Frank E. Pollick[1*]

[1]School of Psychology, University of Glasgow, Glasgow G12 8QB, UK

[2]School of Engineering, University of Glasgow, Glasgow G12 8QB, UK

***Corresponding author:**

Frank Pollick

School of Psychology,

University of Glasgow,

Glasgow

G12 8QB

Highlights

- Participants in the NF group only learned to up-regulate their AI activity in response to auditory stimuli
- Negative relationship found between individuals' empathic traits and up-regulation abilities
- Participants did better at increasing AI activity whilst listening to positive, compared to negatively-valenced, auditory stimuli.

Abstract

The right anterior insula (AI), known to have a key role in the processing and understanding of social emotions, is activated during tasks that involve the act of empathising. Neurofeedback provides individuals with a visualisation of their own brain activity, enabling them to regulate and modify this activity. Following previous research investigating the ability of individuals to up-regulate right AI activity levels through neurofeedback, we investigated whether this could be similarly accomplished during an empathy task involving auditory stimuli of human positive and negative emotional expressions. Twenty participants, ten with feedback from right anterior insula and ten with feedback from a sham brain region, participated in two sessions that included sixteen neurofeedback runs and four transfer runs. Results showed that for the second session participants in the right AI neurofeedback group demonstrated better ability to up-regulate their right AI compared to the control group who received sham feedback. Examination of the relationship between individual participants' empathic traits and their ability to up-regulate right AI activity showed that participants low on empathic traits produced a greater increase in activation of right AI by the end of training. Moreover, the response to positively valenced audio stimuli was greater than for negatively valenced stimuli. These results have implications for therapeutic training of empathy in populations with limited empathic response.

Keywords: neurofeedback, fMRI, rt-fMRI, up-regulation, anterior insula, empathy.

1. Introduction

The underlying neural mechanisms of emotion have been the focus of much research, with neuroimaging revealing the anterior insula (AI) to be the most consistently activated region in studies of emotion (Kober *et al.*, 2008). Social emotions, specifically empathy - the ability to identify other people's emotions and respond to these appropriately with one's own emotions - have also been found to activate the AI region (Lamm & Singer, 2010). Further support for a link between empathy and AI activity comes from autism spectrum disorders (ASD), which frequently involve abnormalities in social and communication development. The right AI has been shown to be hypoactive in autistic individuals during social processing tasks (Di Martino *et al.*, 2009), suggesting a dysfunctional right AI in autistic people produces difficulties in social awareness.

Neurofeedback using real time fMRI (rt-fMRI) is a technique that aims to allow voluntary control of brain function through monitoring metabolic activity in the brain (as denoted by the blood oxygenation-dependent level (BOLD) signal) and visually relaying it back to the participant in real-time (Ruiz *et al.*, 2013). The resulting readout is used by participants to either up- or down-regulate activity levels in a specific brain region using cognitive strategies. If these learned techniques can be used by participants outside the scanner, then they have the potential to manifest beneficial behavioural changes in these individuals (Paret *et al.*, 2014). Real time fMRI has been examined as an intervention in several conditions, including attention deficit hyperactivity disorder (ADHD), depression and phobias (Zilverstand *et al.*, 2017, Linden *et al.*, 2012, Zilverstand *et al.*, 2015, respectively).

Real time neurofeedback has also been utilised in emotion research as a method of emotion regulation. The ability of humans to empathise with others has been trialled as a mechanism to help participants gain control over brain activation patterns. In particular, one study explored the affective aspect of empathy, and demonstrated participants' abilities at increasing BOLD responses in key regions implicated in these traits (Moll *et al.*, 2014). Similarly, the ability of participants to self-regulate their amygdala BOLD activity was investigated by Zotev and colleagues (2011), by contemplating positive autobiographical memories. The researchers found that BOLD signal was

significantly increased by the end of the training, and as well as this, these effects were seen in transfer runs later on, which lacked any neurofeedback. More recently, one research group found rt-fMRI neurofeedback not only allowed participants to up-regulate their amygdala activity, but also helped most participants in the experimental group to meet conventional criteria for remission by the end of the study via a decrease in the depressive symptoms displayed (Young *et al.*, 2014; Young *et al.*, 2017).

Previous research has found rt-fMRI neurofeedback techniques to be successful at up-regulating activity levels of the AI in healthy individuals (Caria *et al.*, 2007, Lawrence *et al.*, 2014). Participants are asked to employ various cognitive strategies to increase their AI BOLD signal, such as using emotional imagery (Berman *et al.*, 2013), or responding to aversive visual stimuli (Veit *et al.*, 2012). A further study found that individuals who managed to increase their AI activity went on to assign more ratings that are negative to aversive pictures post-training. These ratings were in direct correlation with AI activation, demonstrating a behavioural effect of increased emotional engagement (Caria *et al.*, 2010). Further, participants that managed to up-regulate AI activity went on to exhibit stronger empathic responses to painful stimuli, a behavioural effect that was also apparent two days after the training (Yao *et al.*, 2016). One limitation of previous research involves the methods used to elicit these empathic responses. It could be said that simply asking participants to recall emotional memories may create methodological issues, as it is impossible to measure, quantify, and compare emotional memories between participants. Further, the use of visual stimuli to elicit emotional responses, in between visually displaying the NFB signal, poses a problem due to the fact that these two things are not done at the same time. By supplying an alternative stimuli modality, this problem could be overcome, and both stimuli and NFB signal could be administered at the same time. Moreover, the right insula has been shown to respond more strongly than left insula to emotional stimuli of crying and laughing (Sander & Scheich, 2005) and thus the use of such auditory stimuli could reveal new understanding of insula function.

The aim of the present research was to determine whether participants can, using rt-fMRI neurofeedback, learn to up-regulate and enhance right-AI activity levels through empathising in response to auditory stimuli. Although this area has previously been explored using visual inputs, to

our knowledge, no study has made use of auditory stimuli. Furthermore, we seek to determine whether an individual's intrinsic empathy levels affect their ability to enhance activity in their right-AI and whether right-AI activity can be up-regulated in the absence of a visual readout. Finally, we are interested in determining whether the type of auditory stimuli heard in the experiment, either positively- or negatively-valenced, affected participants' up-regulating abilities. It is important to note that we are not exploring a case of pure neurofeedback as participants will be learning to up-regulate in the presence of an audio signal and an empathy task and thus up-regulation can be considered an enhancement of the untrained response.

2. Materials and Methods

2.1 Participants

Twenty healthy participants were recruited and all successfully completed the experiment. Sixteen were right-handed and four were left-handed, as determined by the Edinburgh Handedness Inventory (Oldfield, 1971). Participants were all proficient in written and spoken English; 17 of them had acquired an Undergraduate degree, two a Masters degree and 1 a PhD. Ten participants were assigned to the neurofeedback (NFB) group and received neurofeedback from the right-AI (mean age 24.9±3.07, 6 females). Participants were not told which group they were assigned to and did not have experience with neurofeedback. Ten participants were assigned to the control group (mean age 24.3±3.53, 7 females), which underwent an identical experiment to the NFB group, but were instead shown 'sham' feedback from a distant, unrelated brain region. All participants either verified that their visual acuity was sufficient to resolve images and text presented on the screen without any correction, or they were provided with sufficient correction. Similarly, the volume on the headphones were adjusted so that participants would comfortably hear the audio stimuli.

Participants' empathy scores were collected using the Interpersonal Reactivity Index (IRI; Davis, 1980), a 28-item questionnaire measuring empathy trait levels. As the IRI provides a trait measure of empathy, we collected this data only at the beginning of the experiment. The IRI also includes subscales on perspective taking, fantasy, emotional concern and personal distress. Both groups were

matched, and did not significantly differ, in age ($t(18) = 0.3324$, $p = 0.7435$) or IRI scores ($t(18) = 0.1754$, $p = 0.8627$). Two left-handed participants were assigned to each group to control for brain lateralisation effects.

This study was approved by the ethics committee of the College of Science and Engineering, University of Glasgow. All participants provided their informed consent for the experiment. The study conformed with the World Medical Association Declaration of Helsinki. Participants were compensated £6/hour for their time. No additional financial incentives or reward were used in association with training performance.

2.2 Imaging parameters and rt-fMRI Neurofeedback platform

The present study was performed at the University of Glasgow Centre for Cognitive Neuroimaging (CCNi), using a 3T Siemens Tim Trio MRI scanner with a 32-channel head coil. A T1 weighted structural scan was acquired at the beginning of each session (TR=2300ms, TE=2.96ms, 192 sagittal slices, 1 mm^3 isotropic voxels and image resolution 256×256). T2*-weighted functional scans were obtained using an Echo Planar Imaging (EPI) sequence (TR=2000ms, TE=30ms, whole brain coverage with 32 axial slices, with 0.3 mm gap, 3 mm^3 isotropic voxel).

The neurofeedback system was comprised of Turbo-BrainVoyager v3.2 (TBV) (Brain Innovation, Maastricht, The Netherlands) and a custom script running on MATLAB (Mathworks Inc., Natick, MA, USA) (Goebel *et al.*, 2006). The script was designed to play sound clips and to display the feedback signal, represented as a thermometer, with a fluctuating red bar indicating increasing and decreasing levels of activity in the target region-of-interest (ROI). An LCD projector displayed the thermometer onto a rear projection screen that was viewable through a mirror mounted on the head coil.

2.3 Experimental procedure

Before testing, all participants were given an information sheet detailing the study and were asked to fill out two questionnaires: the Edinburgh Handedness Inventory (Oldfield, 1971) and the

Interpersonal Reactivity Index (Davis, 1980). Participants were also required to fill out a consent form before starting the first scanning session, and given instructions as to what should be done in the scanner. In particular, they were told that during the baseline blocks, they should try and keep their eyes on the cross, and during the up-regulating blocks, their aim was to try and increase the red bar by listening to the sounds and empathising with the human noises.

The experiment consisted of two scanning sessions, each carried out on a separate day and lasting roughly 1 hour 10 minutes. The separation in time of the two sessions was limited to occur within the same week and whenever possible these two sessions occurred on consecutive days. Overall, most participants performed the two sessions on consecutive days (11 overall, 6 NFB and 5 sham). Some participants had a duration between the sessions that lasted 2 days (4 overall, 1 NFB & 3 sham), some 3 days (4 overall, 2 NFB & 2 sham), and one participant waited 4 days to complete the second session. Before each session participants underwent an MRI screening questionnaire. At the beginning of each session, participants underwent preliminary anatomical and functional localisers, before continuing onto the rt-fMRI neurofeedback training. Participants were assigned to the NFB or the control condition in an alternating fashion so that consecutive participants were assigned to different conditions.

2.3.1 Anatomical Scan & Localisation of the AI, sham and reference regions

The target region of the right anterior insula, or ROI_{target}, used in the NFB group, was a 4 x 4 voxel square, spread over 3 slices (making 48 voxels for each ROI_{target}), and was selected using the central sulcus of the insula as an anatomical landmark to separate anterior and posterior regions (Naidich et al., 2004). The selection was done manually using the capability of Turbo-Brainvoyager to allow overlay of functional and anatomical data and alignment was adjusted for different brain sizes to best cover the anterior insula. The sham brain region (ROI_{sham}) used in the control group was taken from a single axial slice of the functional scan and included the anterior lobe of the cerebellum, as well as parts of the midbrain. The reference brain region (ROI_{ref},) used to control for nonspecific global variation of the signal was taken from a single axial slice of the functional scan and included bilateral motor regions. The sham and reference regions were chosen based on being distant from the

anterior insula and not being closely implicated with either co-activation with the anterior insula or processes associated with empathy.

The ROIs were defined in each session of each participant (in the participant's native space) and were then saved for the following neurofeedback training runs in order to acquire the neurofeedback signal. The ROIs were later normalised into Talaraich space for further analysis, and their statistical threshold was modified based on the individual variation. To confirm that there was no difference in ROI selection between the two sessions we performed offline analysis of the ROI centroids for sessions 1 and 2 using paired, two-tailed, t-tests to verify that there was no statistical difference in the centroid locations between session for either group.

2.3.2 rt-fMRI Neurofeedback

The two scanning sessions each included ten runs, each lasting 340 seconds (5.66 minutes). The first scanning session included 10 neurofeedback runs, while the second consisted of 6 neurofeedback runs and 4 'transfer' runs to measure performance after training was complete. We chose to examine transfer only at the end of training, rather than at the end of each session because we wished to measure the full effect that training could have on transfer. During a neurofeedback run participants performed eight up-regulation blocks where they were asked to empathise with presented human vocalisations and attempt to increase the height of a red bar situated inside a thermometer. The transfer runs were identical to the neurofeedback runs except that participants did not receive any feedback when asked to empathise with the human vocalisations. Each run began with a 20 second fixation and after each of the eight blocks of 24 seconds was a baseline block lasting 16 seconds where participants looked at a fixation cross and counted back from 100 (Figure 1). For those in the NFB group, activation levels (or BOLD signals) in participants' AI (ROI_{target}) directly influenced the height of the red bar. For those in the control group, the height was influenced by activation in the ROI_{sham}, which ran through the primary fissure of the cerebellum.

During the up-regulating blocks, participants were presented with one of eight different audio clips, each consisting of a human nonspeech vocalisation, expressing either a negative or a positive emotion, such as laughing or crying. Recordings were taken from the International Affective Digital Sounds (IADS; Bradley & Lang, 1999). Half of the audio clips included positively valenced ('happy')

expressions, such as laughing while the other half depicted negatively valenced ('sad') emotions, such as crying. These included the following eight sound files (with index number): BoyLaugh(220), MaleLaugh(221), Laughing(226) (group of people laughing), Giggling(230) (woman laughing), BabyCry(261), ManSobbing(293), CoupleSobbing(295), WomenCrying(296). Each sound lasted exactly 6 seconds. In a single run each sound file was played repetitively for an entire block and the assignment of audio file to block was randomised for each run, meaning each participant heard all 8 sound clips, but in a random order. The eight sound clips were chosen for their valence and arousal ratings. All had relatively high arousal ratings, scoring above 5 out of a total of 9 points in an affective rating of sounds experiment (Bradley & Lang, 2007). To aid in up-regulating their right AI activity, participants were instructed to try and empathise with the voices they heard. The audio recordings were randomly selected, and presented in a different order for each participant, to reduce bias.

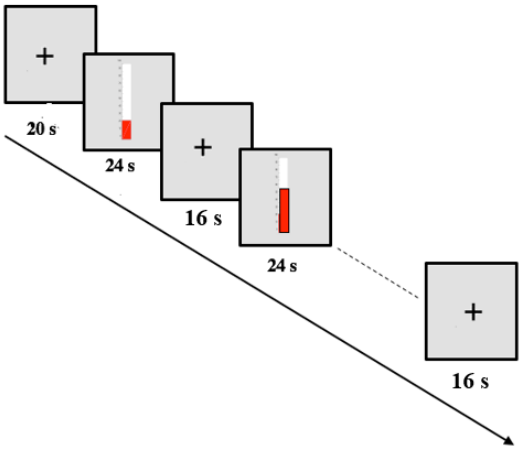

Figure 1. rt-fMRI Neurofeedback training run. Each run lasted 340 seconds and comprised eight neurofeedback blocks alternating with 9 baseline (rest) blocks.

2.4 Real-time neurofeedback display

Real-time data analysis and neurofeedback signal presentation were achieved using TBV and MATLAB. Here, the fMRI data were transmitted from the scanner to the TBV-equipped analysis computer where functional data were pre-processed. This transmission occurred in real-time and additionally featured linear de-trending and 3D motion correction. Further, images were smoothed spatially using an 8-mm Gaussian kernel.

The continually changing feedback signal was displayed as a red column, with a height that was constantly updated at each TR (repetition time; 1 TR = 2000ms), based on the following equation for the NFB group:

$$Column\ height\ (t) = \left(\frac{ROI_{target}(t) - ROI_{target\_base}}{ROI_{target\_base}}\right) - \left(\frac{ROI_{ref}(t) - ROI_{ref\_base}}{ROI_{ref\_base}}\right)$$

$ROI_{target}(t)$ and $ROI_{ref}\ (t)$ denote the averaged ROI BOLD signals of ROI_{target} and ROI_{ref} during the neurofeedback block at time t. ROI_{target_base} and ROI_{ref_base} refer to the average BOLD signals of the last three volumes in each fixation block of ROI_{target} and ROI_{ref}, respectively. For the control group the same definitions were used but the ROI region changed such that $ROI_{target}(t)$ was replaced with $ROI_{sham}(t)$ and ROI_{target_base} was replaced with ROI_{sham_base}. The first half of this equation serves to calculate the average BOLD response in ROI_{target} (the right-AI) and the second half the average BOLD response from ROI_{ref}, a background region, which was used to cancel global effects and average out any unspecific activation.

2.5 Off-line data analysis

BrainVoyager QX 2.8.4 (Brain Innovation, Maastricht, The Netherlands) was used to pre-process raw data offline. To account for T1 equilibration effects, the first two volumes of each run were excluded. Subsequent pre-processing of the functional scanning images included 3D motion correction with Trilinear/Sinc interpolation, slice scan-time correction with cubic-spline interpolation, high-pass filtering with a 2 cycle cut-off and linear trend removal. Functional images were aligned to the first functional volume after the anatomical scan, which was in turn co-registered to the high-resolution anatomical images, before being spatially normalized onto a Talairach template to allow for group analysis (Talairach & Tournoux, 1988).

First level analyses involved a general linear model (GLM) to analyse each participant individually, with one predictor – 'feedback' for the neurofeedback runs. This was achieved using a block-design function convolved with a standard hemodynamic response delay in addition to six head motion parameters added as nuisance predictors (Van Dijk *et al.*, 2012).

2.5.1 Region of interest analysis

Hypothesis-driven ROI analyses were performed using each subject's ROI used during neurofeedback training. Derived beta values were used to represent the extent of up-regulation of BOLD signal in the right-AI, and were estimated using a ROI-GLM that separately analysed each neurofeedback run. A 2-way mixed effects analysis of variance (ANOVA) was run to compare beta values, with the within-group factor being the NFB runs (16 runs), and the between-group factor being Group (NFB vs. control). In addition, an ANOVA was performed for each session separately, with the within-group factor being either session 1 NFB runs (10 runs) or session 2 NFB runs (6 runs); and the between-group factor being Group (NFB vs. control). A similar ANOVA was run for the transfer runs, with the within-group factor being the four transfer runs (runs 17 to 20), and between-group factor being Group. Follow-up, paired samples t-tests were conducted to compare beta values in the first NFB run (run 1) and the last NFB run (run 16) to identify learning effects.

2.5.2 Whole-brain analysis

Group data were evaluated based on a second level random effect analysis general linear model (RFX-GLM). The obtained statistical maps were corrected for multiple comparisons using cluster-level thresholding (Goebel *et al.*, 2006). In this method, the uncorrected voxel-level threshold maps were submitted to a whole-brain correction criterion based on the estimate of the map's spatial smoothness and on an iterative procedure (Monte Carlo simulation) for estimating cluster-level false-positive rates. After 5000 iterations, the minimum cluster-size that produced a cluster-level false positive rate (alpha) of 0.1% was applied to threshold the statistical maps.

A whole brain RFX-GLM analyses was performed: first, all 16 NFB runs of both groups separately, comparing the NFB blocks to the baseline ($p<0.001$ uncorrected with cluster-level thresholding of 356 mm^3 and 254 mm^3 for the NFB and the control groups respectively), second, each session (10 runs for session 1 and 6 runs for session 2) of each group separately ($p<0.001$ uncorrected with cluster-level thresholding of 108 mm^3 for the NFB and the control groups). Finally, a t-test was

run between the two cluster maps to create contrast maps, highlighting any significant differences between the two groups (thresholding at $p < 0.001$).

2.5.3 Brain-behaviour analysis

A simple linear regression was run to identify a relationship between empathy scores and NFB participants' improvements at up-regulating abilities. Improvement in up-regulating ability was calculated by finding the difference in beta values between the first and last neurofeedback run (i.e. run 16 – run 1). Regression analyses were performed for both the NFB and control groups and using a Fisher r-to-z transform the correlation coefficient of two groups were compared. In addition, for a finer grain analysis of IRI scores, we examined the correlation for all four IRI subscales (perspective taking, fantasy, empathic concern, personal distress). Finally, simple linear regressions were run between total IRI scores and NFB participants' beta values for the first and last neurofeedback training runs of session 1 (run 1, run 10) and for session 2 (run 11, run 16).

2.5.4 Valence analysis

We examined whether there was any difference between the effectiveness of positive and negative valence audio clips for neurofeedback training. As the pattern of sounds was completely randomised, and so different auditory stimuli were heard in each run and for each participant, personalised protocol files were created for each participant. These protocol files were linked to participants' single GLM, before second-level analyses were run to determine participants' new beta values. Then, for subsequent analysis, these beta values were split into two bins: those derived from blocks with positively-valenced ('happy') sounds, and those derived from blocks with negatively-valenced ('sad') sounds.

3. Results

3.1 ROI analysis

Each participant completed sixteen NFB runs followed by four transfer runs spread over two sessions. Participants of the NFB group were trained to increase the brain activity measured from their right-AI regions. The average beta values in the right AI estimated during each run of the NFB and equivalent results for the control group are shown in Figure 2. To check that the ROI overlap between sessions was similar for the two groups we calculated overlap proportion and found 47% overlap for the NFB group and 43% overlap for the control group. After using a Shapiro-Wilk test ($p>0.1$) to confirm the distributions were normal we performed a two-tailed t-test, which did not reveal a difference ($p=0.66$).

We performed separate ANOVA analyses on the beta values examining the effect of Group (NFB, control), Run and their interaction for both sessions together (16 runs), Session 1 (10 runs), Session 2 (6 runs) and the four runs of Transfer (Table 1). For session 2, the ANOVA of beta values indicated significant main effects for Group, Run and their interaction. However, for session 1 and both sessions together no significant main effects or interaction were found. The ANOVA of the transfer runs revealed a significant main effect of Group, however no significant effects were found for either Run or interaction between Group and Run.

Subsequent paired t-tests revealed a significant increase in the right AI activity between run 1 to run 16 ($t(9) = -1.946$, $p = 0.041$) in the NFB group. Conversely, no significant difference was found in the control group between the beta values of the first and last runs ($t(9) = 1.397$, $p = 0.098$).

Table 1. The 2-way ANOVA results for NFB sessions.

	Runs	**Group**	**Group x Run**
Both sessions (16 runs)	$F(15,270) = 0.95$, $p = 0.5$	$F(1,18) = 2.26$, $p = 0.15$	$F(15,270) = 1.4$, $p = 0.14$
Session-1 (10 runs)	$F(9,162) = 0.74$, $p = 0.67$	$F(1,18) = 0.14$, $p = 0.7$	$F(9,162) = 0.54$, $p = 0.84$
Session-2 (6 runs)	**$F(5,90) = 2.39$, $p = 0.04$**	**$F(1,18) = 6.24$, $p = 0.02$**	**$F(5,90) = 2.68$, $p = 0.03$**
Transfer runs (4 runs)	$F(3,54) = 0.87$, $p = 0.46$	**$F(1,18) = 6.78$, $p = .018$**	$F(3,54) = 0.92$, $p = 0.43$

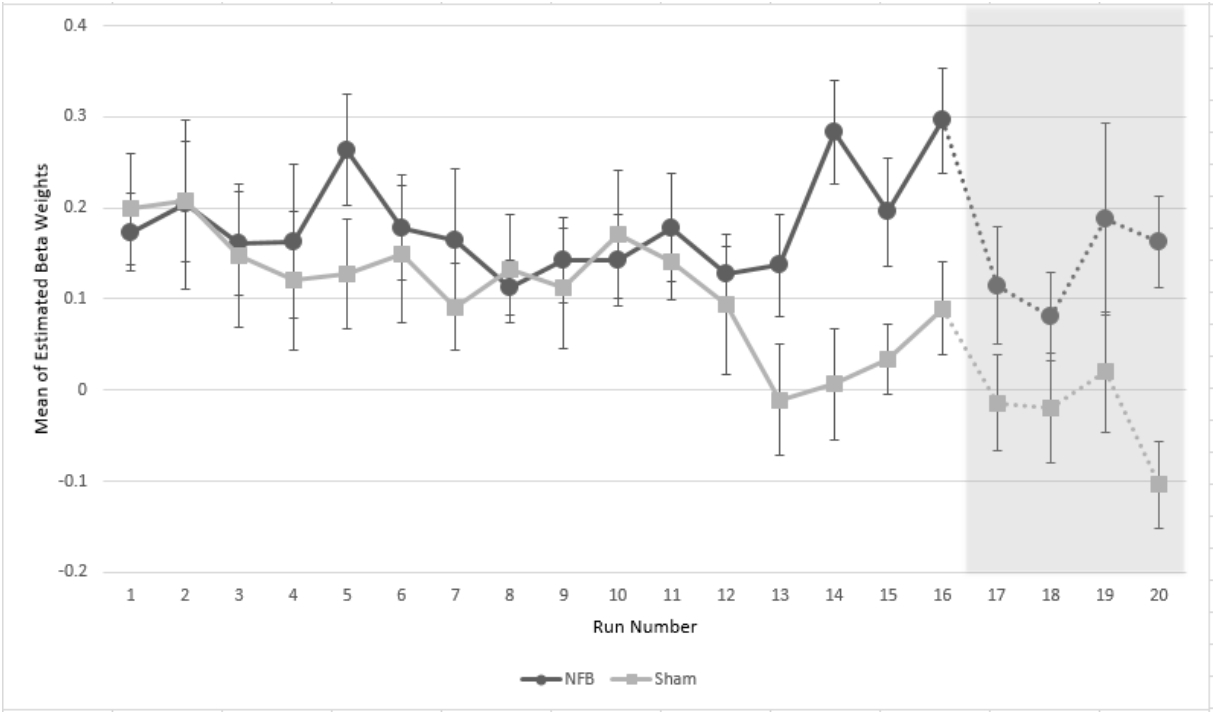

Figure 2. Mean beta values of each neurofeedback run for the two groups. Abilities of increasing right AI activity are shown to improve in the NFB group, especially around the 14th run. The shaded area represents the four transfer runs that were performed at the end of the second session. Error bars indicate the standard error of the mean.

Further investigation into the paired data points between run 1 and run 16 show that, in the NFB group, a majority (70%) of participants had increased beta values in run 16, compared to run 1, indicating mostly successful trials. This was not, however, seen in the control group, where 70% of participants showed decreased beta values in run 16, compared to run 1.

3.2. Whole-brain analyses

A whole brain RFX-GLM analysis was run across all 16 NFB runs for both NFB and control groups separately, as illustrated in Table 2 and Figure 3. For the NFB group, significant activations were found in the right Superior Temporal Gyrus, and bilateral Lentiform Nucleus. For the control group, significant activation levels were found only in right Superior Temporal Gyrus.

Table 2. Table displaying coordinates of the peaks of cluster activation in the NFB and control groups, produced using RFX-GLM analysis, with *t* and *p* values displaying the significance of activation.

(Note: x,y,z are given in Talairach coordinates, LH= Left hemisphere. RH= right hemisphere. BA= Brodmann area).

	Cortical Area	*x*	*y*	*z*	*t*	*p*-value	*Size*
NFB	RH, Superior Temporal Gyrus, BA 41	57	-19	4	8.733942	0.000011	866
	RH, Lentiform Nucleus, Putamen	15	8	4	7.640125	0.000032	2183
	LH, Lentiform Nucleus, Putamen	-21	2	4	6.938818	0.000068	1459
control	RH, Superior Temporal Gyrus, BA 41	54	-22	4	8.310386	0.000016	510

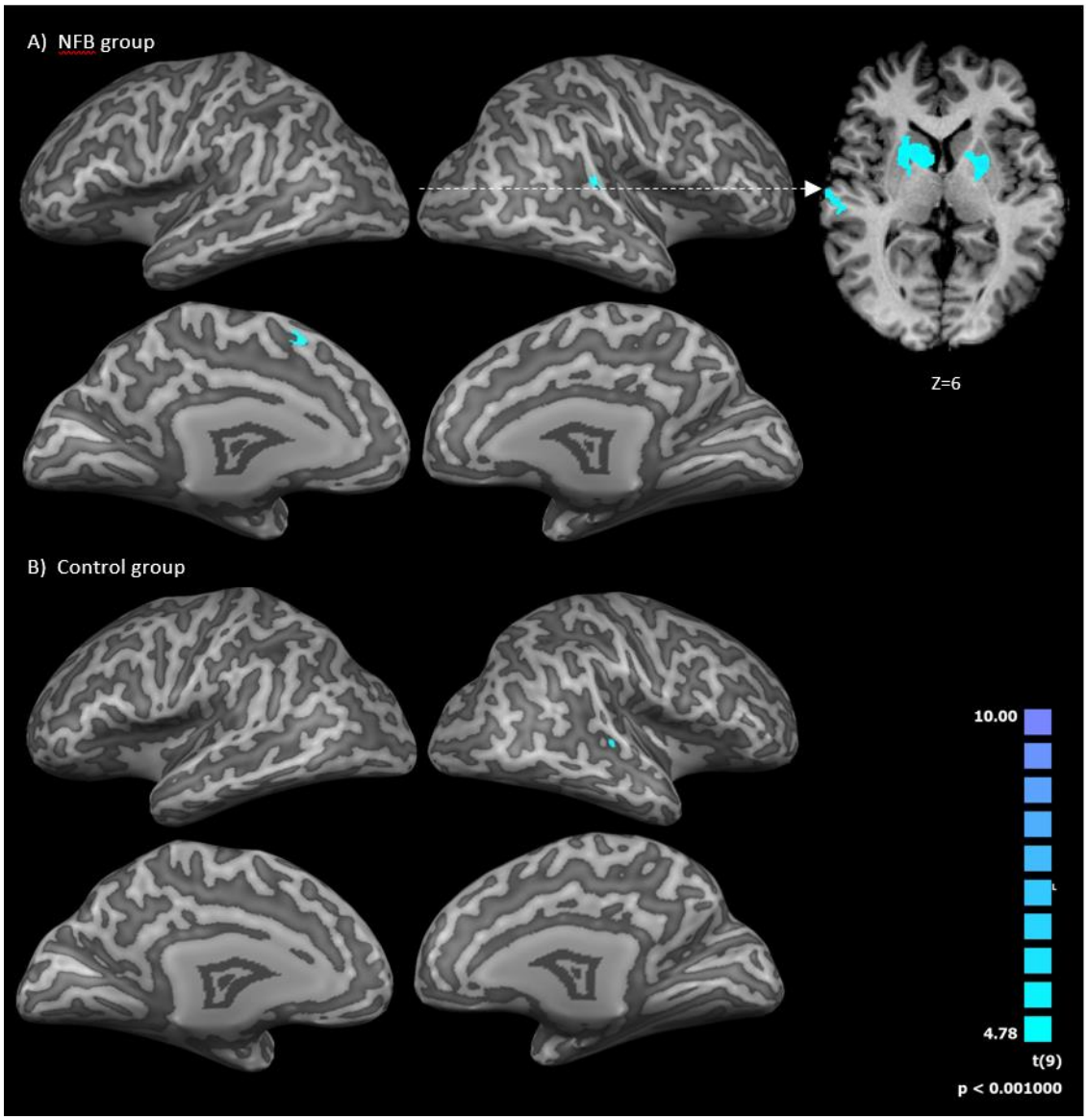

Figure 3. Results of RFX-GLM analysis for the A) NFB group and B) control group. These activations are significant at p<0.001 (cluster size> 356 mm^3 and > 254 mm^3 respectively).

Furthermore, the RFX-GLM was performed on each NFB session separately for both NFB and control groups, as illustrated in Table 3 and Figure 4. The NFB group produced increased activation in

the right anterior insula in the second session only. The control group did not produce increased AI activation in either session. Finally, a t-test examining differences between the NFB and control groups revealed no differences in activation between groups.

Table 3. Table displaying coordinates of the peaks of cluster activation in the NFB and control groups, produced using RFX-GLM analysis, with *t* and *p* values displaying the significance of activation. (Note: x,y,z are given in Talairach coordinates, LH= Left hemisphere. RH= right hemisphere. BA= Brodmann area).

	Session	***Cortical Area***	***x***	***y***	***z***	***t***	***p*-value**	***Size***
NFB	**1**	RH, Superior Temporal Gyrus, BA 41	57	-19	4	7.518297	0.000036	432
		RH, Lentiform Nucleus, Putamen	15	8	4	7.188156	0.000051	1116
		LH, Lentiform Nucleus, Putamen	-15	2	10	6.281009	0.000144	1016
	2	RH, Superior Temporal Gyrus, BA 42	63	-28	13	7.239437	0.000049	333
		RH, Superior Temporal Gyrus, BA 22	60	-13	1	7.90771	0.000024	242
		RH, Insula, BA 13 (anterior)	42	-16	10	7.263908	0.000047	343
		RH, Insula, BA 13 (posterior)	39	20	1	7.401721	0.000041	139
		RH, Superior Temporal Gyrus, BA 22	51	5	-2	7.459524	0.000039	250
		RH, Lentiform Nucleus, Putamen	21	5	10	5.877091	0.000236	170
		RH, Lentiform Nucleus, Putamen	18	8	-2	8.850809	0.00001	967
		RH, Cingulate Gyrus, BA 32	12	20	34	8.311233	0.000016	120
		RH, Midbrain, Subthalamic Nucleus	9	-13	-2	7.339516	0.000044	135
		RH, Midbrain, Red Nucleus	3	-22	-2	7.247826	0.000048	213
		LH, Superior Frontal Gyrus, BA 6	-3	11	49	7.799897	0.000027	286
		LH, Caudate, Caudate Head	-9	8	4	7.198783	0.000051	769
		LH, Middle Frontal Gyrus, BA 9	-36	29	37	8.255245	0.000017	110
		LH, Precentral Gyrus, BA 6	-39	2	34	6.11654	0.000176	119
		LH, Transverse Temporal Gyrus, BA 41	-54	-22	10	8.579368	0.000013	234
control	**1**	RH, Superior Temporal Gyrus, BA 41	54	-22	4	8.291884	0.000017	269
		RH, Superior Temporal Gyrus, BA 22	48	5	-2	9.511335	0.000005	240
		LH, Inferior Frontal Gyrus, BA 9	-33	8	28	8.903774	0.000009	113
	2	RH, Superior Temporal Gyrus, BA 41	54	-22	7	9.291552	0.000007	284

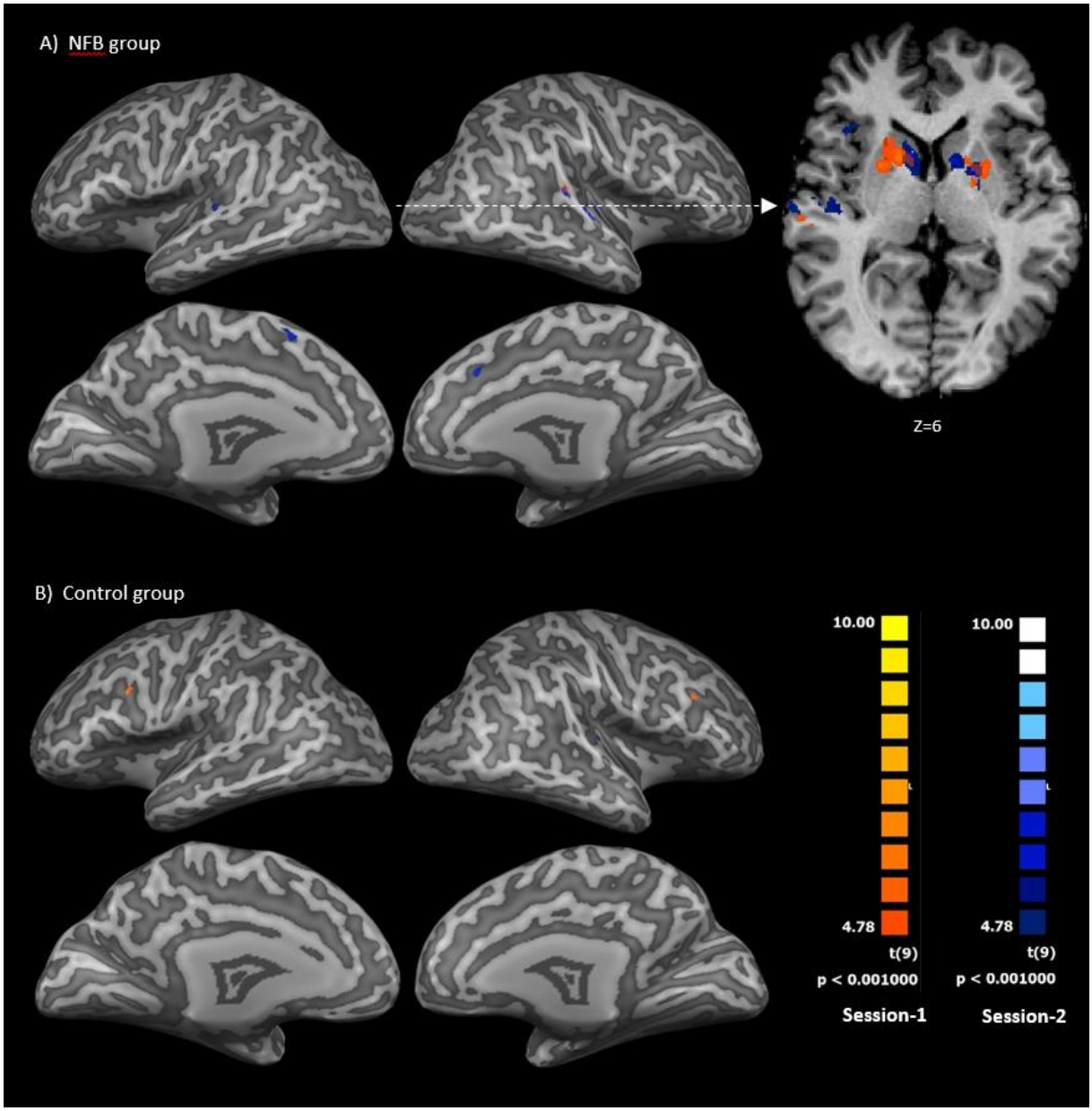

Figure 4. Results of RFX-GLM analysis for each session for the A) NFB group and B) control group. Activations shown in red/orange indicate higher activation in session-1, whereas activations shown in blue/white indicate higher activation in session-2. These activations are significant at $p<0.001$ (cluster size> 108 mm^3 for both groups).

3.3 Brain-Behaviour Association

Empathy levels, as assessed by the IRI, were compared with improvement levels of participants at up-regulation of activity levels in the right-AI. Normality of the distributions of IRI and all subscales was confirmed using a Shapiro-Wilk test ($p>0.1$). A simple linear regression was calculated to predict beta values based on participants' total IRI scores. For NFB participants a marginally significant regression equation was found ($F(1,8) = 5.280$, $p = 0.051$, with an r^2 of 0.398) (Figure 5), with a trend for participants with low IRI scores to show the largest increase in beta values. Results for the control group showed no significant effect of total IRI score on change in beta value ($F(1,8) = 0.669$, $p = 0.437$, with an r^2 of 0.077). Comparison of the correlation coefficients between the NFB ($r=-0.631$) and control ($r=0.278$) participants revealed a significant difference ($z=-1.92$, $p=0.027$, one tailed).

Analysis of all the IRI subscales for a relationship between IRI subscale score and changes in beta values in the right-AI was performed on both the NFB and control groups. Results for the NFB group showed a similar pattern to total IRI, with beta values decreasing with increasing IRI value, but none of these subscales reached significance (perspective taking ($F(1,8)=4.012$, $p=0.08$, $r^2=0.334$), fantasy ($F(1,8)=0.46$, $p=0.517$, $r^2=0.054$), emotional concern ($F(1,8)=2.026$, $p=0.192$, $r^2=0.202$), personal distress ($F(1,8)=1.057$, $p=0.334$, $r^2=0.117$)). Results of the analysis of control participants for the regression between IRI subscale and change in beta values showed that three of these subscales did not reach significance (perspective taking ($F(1,8)=0.018$, $p=0.897$, $r^2=0.002$), fantasy ($F(1,8)=0.197$, $p=0.669$, $r^2=0.024$), emotional concern ($F(1,8)=0.220$, $p=0.652$, $r^2=0.027$), though there was a significant effect for personal distress ($F(1,8)=6.735$, $p=0.032$, $r^2=0.457$)), with greater changes in beta values for individuals with higher subscale scores of personal distress. All reported p values not corrected for multiple comparisons.

Further analyses were performed to examine the relationship between IRI scores and participants' beta values at different runs. Results of simple linear regressions for runs 1, 10, 11 and 16, revealed correlation coefficients of -0.149, -0.368, -0.520 and -0.803 respectively, of which only run 16 was found to be statistically significant ($F(1,8) = 14.486$, $p = 0.005$). Reported p value not corrected for multiple comparisons.

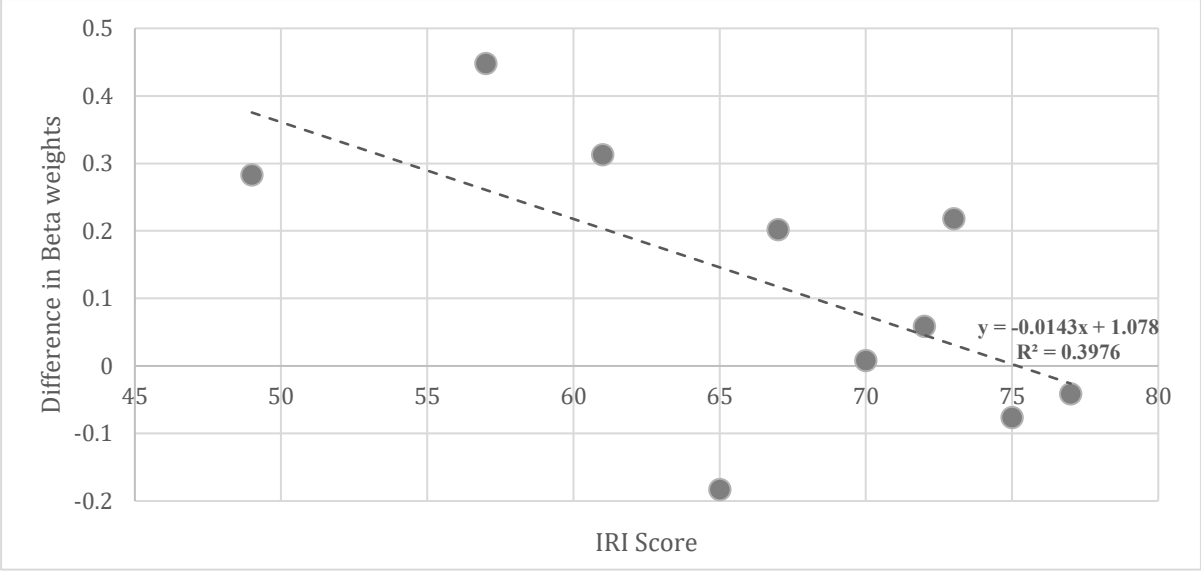

Figure 5. Plot of the difference in beta values between NFB participants' first and last neurofeedback runs (run 16 – run 1) versus total IRI score; a negative trend that was marginally significant (p=0.051) is seen.

3.4 Valence Analysis

Beta values were split into positively- and negatively-valenced blocks, indicating how well participants increased activity levels for each valence. These beta values were collected for each participant in the NFB group, and after performing a Shapiro-Wilk test to confirm normality ($p>0.1$), a t-test was conducted to compare the positive ('happy') and negative ('sad') conditions. There was a significant difference in the scores for the positive ('happy') ($M = 0.195$, $SD = 0.063$) and negative ('sad') conditions ($M = 0.178$, $SD = 0.059$); ($t(15) = 12.395$, $p < 0.0001$). As seen in Figure 6, in all but four of the neurofeedback runs, participants did better (increased right AI activity levels) whilst listening to positively-valenced auditory stimuli.

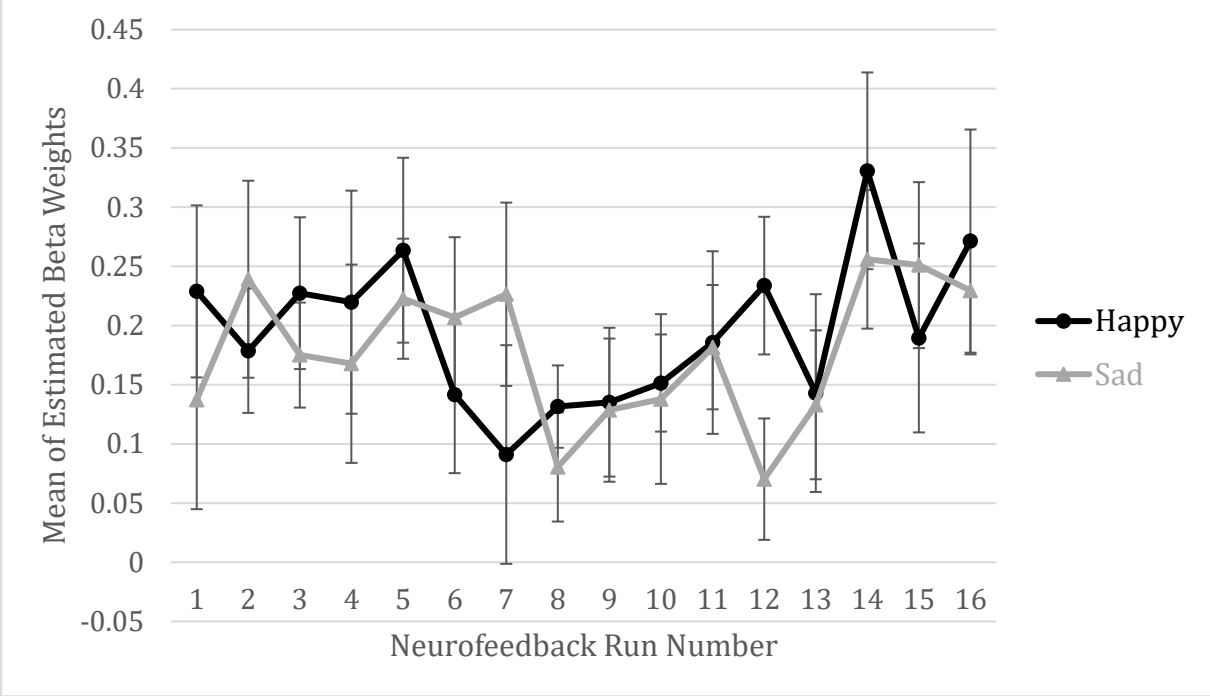

Figure 6. Mean beta values of each neurofeedback run, split between stimuli type. Ability to increase right AI activity are shown for both positively- and negatively-valenced stimuli. In all but four runs, positively-valenced sounds elicited a greater reaction than negatively-valenced ones. Error bars indicate the standard error of the mean.

4. Discussion

Participants in the NFB group showed their ability to up-regulate right-AI activity levels as evidenced by a significant increase in up-regulation during the second session. The same was not seen in the control group, who in fact displayed no significant change in BOLD activity during the up-regulating blocks, from the first to the last run. Over all runs, seven out of the ten NFB participants managed to improve their ability at increasing AI BOLD signal over the training period, while this was true for only three control participants. Previous studies have also reported a portion of neurofeedback trials being unsuccessful, with one paper indicating a failure rate of a quarter for all participants in the experimental group – a figure not hugely different to the one seen in the current study (Auer *et al.*, 2015). Whole-brain analyses revealed brain regions activated across during the up-regulation blocks. For both the NFB and control groups, activated regions included the bilateral auditory cortex (BA 41). For the NFB group only, the areas with a significant level of activation included bilateral putamen.

Examination of the relationship between behavioural and brain data revealed an association between empathy traits, as measured by the IRI, and ability to up-regulate. Although results showed no relationship between IRI scores and beta values of the first neurofeedback run, a significant negative correlation between IRI scores and beta values was found for the final neurofeedback run. In addition, IRI scores showed a marginally significant negative trend with the difference in beta values between the first and last run. Finally, results from valence analyses indicated happy sounds elicited greater responses in participants, or higher BOLD levels, than negatively-valenced sounds.

4.1 Regions-of-Interest Analyses

ANOVAs examining the effects of the within-subject factor of Run and between-subject factor of Group for session 1 showed main effects for Run, Group and their interaction, while for session 1 and both sessions together no significant effects were found. However, a planned comparison between beta values at run 16 versus run 1 showed an increase for only the NFB group. Taken together, the increase in the NFB group from run 1 to run 16 and a significant interaction between Group and Run as well as main effects for both Group and Runs in session 2 suggests that the second session was critical for individuals in the NFB group to learn to up-regulate activity in their right AI. The ability to learn to up-regulate right AI activity is in line with previous neurofeedback research that employed different up-regulation strategies (Caria *et al.*, 2010, Lawrence *et al.*, 2014). Interestingly, visual examination of the pattern of activation across runs reveals a slight decrease in activation levels for the NFB group from runs six through ten – towards the end of the first session. It is possible that, due to the cognitive effort required by this experiment, participants became fatigued by the time they reached the later runs, performing worse towards the end of the session.

Results from the transfer runs did not clearly indicate an effect of transfer. There was a statistically significant effect of Group reflecting better performance of the NFB group in up-regulating in the absence of a feedback signal. However, a true transfer effect could be said to manifest in a significantly greater activation than run 1. Although this was not the case in the current study, it could be suggested that as the BOLD levels in the first transfer run were similar to that of the first NFB run,

participants learnt these skills and were able to produce the same results, without the help of any visual feedback. In order to confirm this, future studies could implement a baseline transfer run before the NFB training. The ability to demonstrate transfer bears significance to real-world applications for possible future therapies (e.g. for autism, ADHD or depression) that incorporate neurofeedback with the objective of patients applying learned abilities to their everyday lives.

Finally, it has been suggested that sham-neurofeedback would have its own effects on participants' up-regulating abilities, compared to passively viewing an unmoving image. Due to this, it is of utmost importance that control groups are shown these placebo neurofeedback stimuli, to distinguish the benefits of genuine rt-fMRI neurofeedback, above the psychosocial influences that are an inevitable part of the study (Thibault *et al.*, 2017).

4.2 Whole-brain analyses

The brain networks predicted to be involved in this task were elements of auditory processing, emotional understanding and empathic processes. In both NFB and control groups, the auditory cortex (BA 41) was activated due to the presence of auditory input during the up-regulation blocks.

For the NFB group only, the cluster map revealed activation for the basal ganglia, known to have an important role in emotional processing (Lanciego *et al.*, 2012). Interestingly, support for the contribution of the basal ganglia to recognising emotional prosody emerged from studies of patients with lesions of the basal ganglia, who exhibited difficulties in recognising emotional tone of pre-recorded utterances (Weddell, 1994). This claim is reinforced by more recent functional neuroimaging studies, which provide evidence for the activation of the basal ganglia during processing of emotional vocal cues (Kotz *et al.*, 2003).

4.3 Empathy levels association

A notable finding of the current study is the link between participants' intrinsic empathy levels, as indicated by the IRI, and their ability to up-regulate right AI BOLD signal over the course of the feedback training. Although the relationship between IRI and change in up-regulation abilities was

only marginally significant with a *p* value of 0.51, there was a significant correlation between IRI and beta values at the final neurofeedback run that revealed less ability to up-regulate for individuals with higher IRI scores. Such a relationship was not apparent for the first neurofeedback run. Moreover, there was a significant difference in the correlation between the NFB and control group. These results point to individual differences in the effectiveness of this rt-fMRI training paradigm. This might be explained by the fact that individuals' empathic trait levels have been shown to have a positive association with their AI activation intensities, whilst witnessing the expression of emotions by others (Jabbi *et al.*, 2007). Therefore, it could be argued that individuals with a decreased empathic ability – those who display a hypoactive AI – would possess greater potential for increasing their activation levels than individuals whose AI activity is constitutively high. This therefore highlights the suitability of using such techniques in certain populations, such as individuals with autism, who are known to have low empathic traits, as well as a hypoactive AI (Di Martino *et al.*, 2009). Therapeutics directed at changing patients' inherent activation levels may thus be a useful supplement to behavioural therapies.

4.4 Valence analysis

The final stage of analysis examined the effects of the type of stimuli on up-regulation success of right AI activity levels. Previous research in this area has indicated that positively-valenced emotions, elicited by viewing pleased facial expressions, produced greater left than right AI activation (Jabbi *et al.*, 2007), whilst the right AI is thought to become activated predominantly by arousing, negative stimuli (Craig, 2009). One might assume, therefore, that it would be negative stimuli that would allow participants in this experiment to increase AI activity levels to the highest degree. It could be argued, however, that an area in the brain that does not usually become significantly activated in response to happy auditory inputs, i.e. the right anterior insula, would have the greatest potential to increase this activity. Future research that use emotional stimuli in neurofeedback should further examine this question of whether the emotion typically associated with the strongest response in a region has the best potential for training up-regulation. An additional distinction that can be made is that the current task involved listening to emotional sounds and the right hemisphere is known to be dominant for the

perception of prosody (Gandour et al., 2004). In addition, the right insula has been reported to respond stronger than the left insula to the sound of laughing and crying, with the response for laughing (nonsignificantly) appearing slightly greater than for crying (Sander & Scheich, 2005). This highlights that the use of voices rather than visually presented displays for neurofeedback provides a unique window into emotional processing in right-AI.

4.5 Limitations

One limitation of the current research was that it lacked any form of behavioural follow-up. This was mainly due to the nature of the IRI, as it was felt that this targeted a global concept of the participants' character traits, as opposed to the participants' current states. Further research could go on to investigate the differences in empathic levels before and after such training, using variables, specifically behavioural ones, that encapsulate this feature better. In addition, other neurofeedback studies have been designed with a follow-up over a longer period of time to explore sustained differences after training (Scheinost et al., 2013) and it would be interesting to examine whether any changes found would maintain over longer periods of time.

A further limitation of the current study involves the small sample size, which presents a problem in regards to the use of random effects analyses. These results should be replicated with a larger group of participants to support the current findings, as well as more systematically randomising group assignment, to avoid selection bias. In addition, our sample population was heterogeneous and thus the general mask we used to identify right anterior insula might not have accommodated for variability in the insula arising from sex and handedness differences. Given that sex and handedness were matched across our NFB and control groups this is not likely to have influenced our results. However, future research would benefit from a definition of the region of interest that reflects an individual's individual insula structure and functional organisation. This would be of particular relevance in consideration of any particular clinical group.

A final consideration is that although our pilot studies indicated that participants could up-regulate right AI using the current stimulation paradigm, it is possible that changes to this paradigm might

yield a higher success rate. For example, more precise targeting within the anterior insula could potentially isolate regions that are more specific for the processing of empathy. Similarly, changes in the timing or numbers of blocks within a run might be more effective. In particular, it is a possibility that participants became fatigued during the scan, which could have adversely influenced neurofeedback training.

4.6 Conclusion

The motivation for the current study stemmed from our view that an increase in the activation of the right AI during social tasks, specifically ones involving the ability to empathise, might ultimately lead to better functioning social interactions. This would be especially beneficial to those known to have an underactive AI, such as autistic individuals. Our results show that participants in the NFB group learned to up-regulate and enhance their brain activity whilst receiving real-time feedback from the AI, compared to the control group, who did not manage to up-regulate activity. This suggests that up-regulation of the AI can be achieved within two rt-fMRI neurofeedback sessions and reveals the importance of the second session. Further, a negative link was found between individuals' intrinsic empathic tendencies and their up-regulation learning capabilities to enhance right AI activity in response to emotional auditory stimuli, suggesting that individuals with lower empathic traits have greater potential for increasing right AI activity than those with pre-existing high empathic traits, a finding that may hold significance when considering the delivery of neurofeedback-based therapies to populations with low empathy levels, such as those with autistic spectrum disorders.

Acknowledgments

We thank the editor and reviewers for their many helpful comments and suggestions. SA acknowledges the support of the Higher Committee for Education Development in Iraq.

References

Auer, T., Schweizer, R., & Frahm, J. (2015). Training efficiency and transfer success in an extended real-time functional MRI neurofeedback training of the somatomotor cortex of healthy subjects. *Frontiers in human neuroscience*, *9*.

Berman, B. D., Horovitz, S. G., & Hallett, M. (2013). Modulation of functionally localized right insular cortex activity using real-time fMRI-based neurofeedback. *Frontiers in human neuroscience*, *7*, 638.

Bradley, M. M., & Lang, P. J. (1999). International affective digitized sounds (IADS): Stimuli, instruction manual and affective ratings (Tech. Rep. No. B-2). Gainesville, FL: The Center for Research in Psychophysiology, University of Florida

Bradley, M. M., & Lang, P. J. (2007). The International Affective Digitized Sounds (IADS-2): Affective ratings of sounds and instruction manual. *University of Florida, Gainesville, FL, Tech. Rep. B-3*.

Caria, A., Veit, R., Sitaram, R., Lotze, M., Weiskopf, N., Grodd, W., & Birbaumer, N. (2007). Regulation of anterior insular cortex activity using real-time fMRI. *Neuroimage*, *35*(3), 1238-1246.

Caria, A., Sitaram, R., Veit, R., Begliomini, C., & Birbaumer, N. (2010). Volitional control of anterior insula activity modulates the response to aversive stimuli. A real-time functional magnetic resonance imaging study. *Biological psychiatry*, ***68***(5), 425-432.

Craig, A. D. (2009). How do you feel--now? The anterior insula and human awareness. *Nature reviews neuroscience*, *10*(1).

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. JSAS Catalog of Selected Documents in Psychology, 10, 85.

Di Martino, A., Ross, K., Uddin, L. Q., Sklar, A. B., Castellanos, F. X., & Milham, M. P. (2009). Functional brain correlates of social and nonsocial processes in autism spectrum disorders: an activation likelihood estimation meta-analysis. *Biological psychiatry*, ***65***(1), 63-74.

Gandour, J., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y., ... & Lowe, M. (2004). Hemispheric roles in the perception of speech prosody. *Neuroimage*, *23*(1), 344-357.

Goebel, R., Esposito, F., & Formisano, E. (2006). Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Human brain mapping*, *27*(5), 392-401.

Jabbi, M., Swart, M., & Keysers, C. (2007). Empathy for positive and negative emotions in the gustatory cortex. *Neuroimage*, ***34***(4), 1744-1753.

Johnson, K. A., Hartwell, K., LeMatty, T., Borckardt, J., Morgan, P. S., Govindarajan, K., ... & George, M. S. (2012). Intermittent "Real‐time" fMRI feedback is superior to continuous presentation for a motor imagery task: a pilot study. *Journal of Neuroimaging*, ***22***(1), 58-66.

Kober, H., Barrett, L. F., Joseph, J., Bliss-Moreau, E., Lindquist, K., & Wager, T. D. (2008). Functional grouping and cortical–subcortical interactions in emotion: a meta-analysis of neuroimaging studies. *Neuroimage*, ***42***(2), 998-1031.

Kotz, S. A., Meyer, M., Alter, K., Besson, M., von Cramon, D. Y., & Friederici, A. D. (2003). On the lateralization of emotional prosody: an event-related functional MR investigation. *Brain and language*, ***86***(3), 366-376.

Lamm, C., & Singer, T. (2010). The role of anterior insular cortex in social emotions. *Brain Structure and Function*, ***214***(5-6), 579-591.

Lanciego, J. L., Luquin, N., & Obeso, J. A. (2012). Functional neuroanatomy of the basal ganglia. *Cold Spring Harbor perspectives in medicine*, ***2***(12), a009621.

Lawrence, E. J., Su, L., Barker, G. J., Medford, N., Dalton, J., Williams, S. C., ... & Brammer, M. (2014). Self-regulation of the anterior insula: Reinforcement learning using real-time fMRI neurofeedback. *Neuroimage*, ***88***, 113-124.

Linden, D. E., Habes, I., Johnston, S. J., Linden, S., Tatineni, R., Subramanian, L., ... & Goebel, R. (2012). Real-time self-regulation of emotion networks in patients with depression. *PloS one*, ***7***(6), e38115.

Moll, J., Weingartner, J. H., Bado, P., Basilio, R., Sato, J. R., Melo, B. R., ... & Zahn, R. (2014). Voluntary enhancement of neural signatures of affiliative emotion using FMRI neurofeedback. *PloS one*, *9*(5), e97343.

Naidich, T. P., Kang, E., Fatterpekar, G. M., Delman, B. N., Gultekin, S. H., Wolfe, D., ... & Yousry, T. A. (2004). The insula: anatomic study and MR imaging display at 1.5 T. *American Journal of Neuroradiology*, *25*(2), 222-232.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, ***9***(1), 97-113.

Paret, C., Kluetsch, R., Ruf, M., Demirakca, T., Hoesterey, S., Ende, G., & Schmahl, C. (2014). Down-regulation of amygdala activation with real-time fMRI neurofeedback in a healthy female sample. *Frontiers in behavioral neuroscience*, ***8***, 299.

Ruiz, S., Lee, S., Soekadar, S. R., Caria, A., Veit, R., Kircher, T., ... & Sitaram, R. (2013). Acquired self‐control of insula cortex modulates emotion recognition and brain network connectivity in schizophrenia. *Human brain mapping*, ***34***(1), 200-212.

Sander, K., & Scheich, H. (2005). Left auditory cortex and amygdala, but right insula dominance for human laughing and crying. *Journal of cognitive Neuroscience*, *17*(10), 1519-1531.

Scheinost, D., Stoica, T., Saksa, J., Papademetris, X., Constable, R. T., Pittenger, C., & Hampson, M. (2013). Orbitofrontal cortex neurofeedback produces lasting changes in contamination anxiety and resting-state connectivity. *Translational psychiatry*, *3*(4), e250.

Talairach, J., & Tournoux, P. (1988). Co-planar stereotaxic atlas of the human brain. 3-Dimensional proportional system: an approach to cerebral imaging.

Thibault, R. T., MacPherson, A., Lifshitz, M., Roth, R. R., & Raz, A. (2017). Neurofeedback with fMRI: A critical systematic review. *NeuroImage*.

Van Dijk, K. R., Sabuncu, M. R., & Buckner, R. L. (2012). The influence of head motion on intrinsic functional connectivity MRI. *Neuroimage*, ***59***(1), 431-438.

Veit, R., Singh, V., Sitaram, R., Caria, A., Rauss, K., & Birbaumer, N. (2012). Using real-time fMRI to learn voluntary regulation of the anterior insula in the presence of threat-related stimuli. *Social cognitive and affective neuroscience*, *7*(6), 623-634.

Weddell, R. A. (1994). Effects of subcortical lesion site on human emotional behavior. *Brain and cognition*, ***25***(2), 161-193.

Yao, S., Becker, B., Geng, Y., Zhao, Z., Xu, X., Zhao, W., ... & Kendrick, K. M. (2016). Voluntary control of anterior insula and its functional connections is feedback-independent and increases pain empathy. *NeuroImage*, *130*, 230-240.

Young, K. D., Zotev, V., Phillips, R., Misaki, M., Yuan, H., Drevets, W. C., & Bodurka, J. (2014). Real-time FMRI neurofeedback training of amygdala activity in patients with major depressive disorder. *PloS one*, *9*(2), e88785.

Young, K. D., Siegle, G. J., Zotev, V., Phillips, R., Misaki, M., Yuan, H., ... & Bodurka, J. (2017). Randomized clinical trial of real-time fMRI amygdala neurofeedback for major depressive disorder: effects on symptoms and autobiographical memory recall. *American Journal of Psychiatry*, *174*(8), 748-755.

Zilverstand, A., Sorger, B., Sarkheil, P., & Goebel, R. (2015). fMRI neurofeedback facilitates anxiety regulation in females with spider phobia. *Frontiers in behavioral neuroscience*, ***9***.

Zilverstand, A., Sorger, B., Slaats-Willemse, D., Kan, C. C., Goebel, R., & Buitelaar, J. K. (2017). fMRI neurofeedback training for increasing anterior cingulate cortex activation in adult attention deficit hyperactivity disorder. An exploratory randomized, single-blinded study. *PloS one*, 12(1), e0170795.

Zotev, V., Krueger, F., Phillips, R., Alvarez, R. P., Simmons, W. K., Bellgowan, P., ... & Bodurka, J. (2011). Self-regulation of amygdala activation using real-time fMRI neurofeedback. *PloS one*, *6*(9), e24522.