NeuroImage 279 (2023) 120334

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Training volitional control of the theory of mind network with real-time fMRI neurofeedback

Abhishek Saxena [a], Bridget J. Shovestul [a], Emily M. Dudek [b], Stephanie Reda [a], Arun Venkataraman [c], J. Steven Lamberti [d], David Dodell-Feder [a,e,*]

[a] Department of Psychology, University of Rochester, 500 Wilson Blvd Rochester, NY 14627 USA
[b] Department of Psychology, University of Houston, 3695 Cullen Boulevard Houston, TX 77204 USA
[c] School of Medicine and Dentistry, University of Rochester Medical Center, 601 Elmwood Avenue, Rochester, NY 14642 USA
[d] Department of Psychiatry, University of Rochester Medical Center, 601 Elmwood Avenue, Rochester, NY 14642 USA
[e] Department of Neuroscience, University of Rochester Medical Center, 601 Elmwood Avenue, Rochester, NY 14642 USA

ARTICLE INFO

Keywords:
Real-time fMRI
Neurofeedback
Theory of mind
Social cognition
Social neuroscience
Temporo-parietal junction

ABSTRACT

Is there a way improve our ability to understand the minds of others? Towards addressing this question, here, we conducted a single-arm, proof-of-concept study to evaluate whether real-time fMRI neurofeedback (rtfMRI-NF) from the temporo-parietal junction (TPJ) leads to volitional control of the neural network subserving theory of mind (ToM; the process by which we attribute and reason about the mental states of others). As additional aims, we evaluated the strategies used to self-regulate the network and whether volitional control of the ToM network was moderated by participant characteristics and associated with improved performance on behavioral measures. Sixteen participants underwent fMRI while completing a task designed to individually-localize the TPJ, and then three separate rtfMRI-NF scans during which they completed multiple runs of a training task while receiving intermittent, activation-based feedback from the TPJ, and one run of a transfer task in which no neurofeedback was provided. Region-of-interest analyses demonstrated volitional control in most regions during the training tasks and during the transfer task, although the effects were smaller in magnitude and not observed in one of the neurofeedback targets for the transfer task. Text analysis demonstrated that volitional control was most strongly associated with thinking about prior social experiences when up-regulating the neural signal. Analysis of behavioral performance and brain-behavior associations largely did not reveal behavior changes except for a positive association between volitional control in RTPJ and changes in performance on one ToM task. Exploratory analysis suggested neurofeedback-related learning occurred, although some degree of volitional control appeared to be conferred with the initial self-regulation strategy provided to participants (i.e., without the neurofeedback signal). Critical study limitations include the lack of a control group and pre-rtfMRI transfer scan, which prevents a more direct assessment of neurofeedback-induced volitional control, and a small sample size, which may have led to an overestimate and/or unreliable estimate of study effects. Nonetheless, together, this study demonstrates the feasibility of training volitional control of a social cognitive brain network, which may have important clinical applications. Given the study's limitations, findings from this study should be replicated with more robust experimental designs.

1. Introduction

Reasoning about the unobservable beliefs, thoughts, desires, and intentions of others in everyday social interaction—a process known as "theory of mind" (ToM; Premack & Woodruff, 1978)—can be challenging (Birch & Bloom, 2007; Keysar et al., 2003). And yet, our ability to solve this challenge carries important consequences for our social relationships across contexts and developmental periods (Blatt et al., 2010; Cahill et al., 2020; Caputi et al., 2012; Fink et al., 2015; Galinsky et al., 2008; Goldstein et al., 2014; Imuta et al., 2016; Lecce et al., 2017; Slaughter et al., 2002, 2015; Watson et al., 1999), which in turn, carries important consequences for our health and well-being (Hawkley, 2022; Hawkley & Cacioppo, 2010; Holt-Lunstad et al., 2010, Holt-Lunstad et al., 2015, 2017; Smith & Christakis, 2008; Yang et al., 2016). Given

* Corresponding author at: Department of Psychology, University of Rochester, 500 Wilson Blvd Rochester, NY 14627 USA.
E-mail address: d.dodell-feder@rochester.edu (D. Dodell-Feder).

https://doi.org/10.1016/j.neuroimage.2023.120334
Received 17 March 2023; Received in revised form 12 July 2023; Accepted 14 August 2023
Available online 15 August 2023

these associations, it stands to reason that improving ToM may improve our social lives and carry with it the broader concomitant benefits to other aspects of our non-social lives.

In addition to making the social benefits of improved ToM more easily attainable for the general population, the clinical implications of being able to improve ToM would be profound. Many mental and neurological disorders are characterized by marked disruption to ToM (Cotter et al., 2018). Interventions for ToM deficits in those with mental disorders exist, which have been carefully evaluated in individuals with schizophrenia spectrum disorders. However, several meta-analyses have shown that the ToM improvements from such interventions for individuals with schizophrenia spectrum disorders tend to be moderate and dependent on the scope of the intervention, and often do not generalize beyond the training tasks to daily social behavior (Kurtz & Richardson, 2012; Nijman et al., 2020; Yeo et al., 2022). These findings may be explained, in part, by the notion that these interventions do not necessarily attempt to target the underlying neurobiological processes mediating ToM deficits, leaving a notoriously challenging problem, contributing to daily social difficulties (Fett et al., 2011; Thibaudeau et al., 2021), without a good solution.

A now large body of work has demonstrated that ToM is subserved by a network of brain regions including left and right temporo-parietal junction (LTPJ, RTPJ), superior temporal sulcus (STS), dorsal to ventral aspects of medial prefrontal cortex (MPFC), and the precuneus (PC) (Mar, 2011; Molenberghs et al., 2016; Schurz et al., 2014, 2021; Van Overwalle, 2009), often referred to as the "ToM network." This network responds preferentially to mental state information across a wide variety of stimulus presentation formats and tasks in which mental state attribution is either deliberate and directed or implicit and spontaneous (Mar, 2011; Molenberghs et al., 2016; Schurz et al., 2014, 2021). Importantly, activity in this network has been positively associated with performance on social cognitive measures (Bowman et al., 2019; De Coster et al., 2019; Dodell-Feder et al., 2014; Dodell-Feder et al., 2021; Gweon et al., 2012; Kana et al., 2009; Kanske et al., 2015) and real-world social outcomes in both individuals with and without mental disorders (Dodell-Feder et al., 2014; Dodell-Feder et al., 2016; Dodell-Feder et al., 2014; Hildebrandt et al., 2021; Masten et al., 2011; Morelli et al., 2014; Mukerji et al., 2019; Powers et al., 2016; Rameson et al., 2012; Tusche et al., 2016). These data suggest that the ToM network may be a promising neurobiological target for improving the accuracy and granularity of mental state attribution and/or the tendency to think and inquire about the mental states of others (which has been shown to improve mental state reasoning accuracy; Damen et al., 2021; Eyal et al., 2018). This in turn may improve one's social responsiveness such as providing effective support and validation (Finkel et al., 2017; Reis et al., 2004), and perhaps more generally, make one's social partner feel better understood, which too carries important relationship benefits (Reis et al., 2017). Thus, positively altering function in the ToM network could carry important, real-world social implications.

One method of directly training the response of brain regions mediating complex cognitive processes is real-time functional magnetic resonance imaging neurofeedback (rtfMRI-NF). With rtfMRI-NF, brain function is analyzed and presented back to users in real-time in the form of neurofeedback. Using the neurofeedback signal, users can learn to self-modulate a given brain region(s), and in turn, the cognitive and behavioral processes mediated by that region. The promise of neurofeedback with rtfMRI has been long recognized and well summarized by others (deCharms, 2007, 2008; Paret et al., 2019; Stoeckel et al., 2014; Sulzer et al., 2013; Weiskopf, 2012; Weiskopf et al., 2007). Increasingly, research shows that rtfMRI-NF's promise has at least partly been realized in a variety of applications. Several qualitative reviews have demonstrated that volitional control over a variety of brain region(s) and networks mediating a variety of cognitive, behavioral, and pathophysiological processes can be gained through rtfMRI-NF in samples with and without mental disorders (Martz et al., 2020; Pindi et al., 2022; Scharnowski & Weiskopf, 2015; Sitaram et al., 2017; Taschereau-Dumouchel et al., 2022; Thibault et al., 2018; Tursic et al., 2020; Watanabe et al., 2017). A recent quantitative review of randomized controlled trials of rtfMRI-NF for mental disorders demonstrated a medium-sized effect (g=.59) on the targeted brain region while actively receiving neurofeedback, and a large-sized effect (g=.84) on tests of generalization when no neurofeedback is provided (Dudek & Dodell-Feder, 2021). Several studies have also demonstrated the feasibility of using rtfMRI-NF to train volitional control of brain regions involved in social information processing such as the anterior insula (Kanel et al., 2019; Ruiz et al., 2013; Yao et al., 2016), posterior superior temporal sulcus (Direito et al., 2021), fusiform face area (Pereira et al., 2019), and brain areas associated with affiliative emotions (Moll et al., 2014), which, in certain cases, was associated with improved behavioral performance on social tasks associated with the brain regions targeted for neurofeedback (Ruiz et al., 2013; Yao et al., 2016). However, we are unaware of attempts to train volitional control of the ToM network with rtfMRI-NF, which, if effective, could carry important implications for fostering, maintaining, and enhancing social interactions and relationships in the general population as well as for individuals with mental disorders characterized by social deficits.

Thus, we conducted a preregistered, single-arm, proof-of-concept study to evaluate whether individuals could gain volitional control of the ToM network with rtfMRI-NF. Participants completed a battery of pre-rtfMRI-NF measures to assess potential moderators and behavioral performance on tasks associated with our neurofeedback target (i.e., mental state attribution and attentional reorienting), which were repeated post-rtfMRI-NF. We selected the temporo-parietal junction (TPJ) as the neurofeedback source. Increasing work has shown that the TPJ demonstrates the most selective profile for mental state information (Molenberghs et al., 2016; Saxe & Kanwisher, 2003; Saxe & Powell, 2006; Schurz et al., 2014; Van Overwalle, 2009). Further, patient and neuromodulation studies have suggested a causal role for the TPJ in ToM (Apperly et al., 2004; Bardi et al., 2017; Mai et al., 2016; Samson et al., 2004; Young et al., 2010). Given the co-activation among regions in the network in response to mental state information, we expected that successful self-regulation of the TPJ would generalize to the rest of the ToM network. After individually localizing the TPJ with an fMRI task, participants completed three separate rtfMRI-NF scanning sessions. In each session, participants completed (a) multiple *training* runs in which they were directed to either increase or decrease activity in their TPJ and received intermittent, activation-based feedback, and, subsequently, (b) a single *transfer* run in which they completed the same task, but without receiving neurofeedback in order to test whether learning occurred. After each rtfMRI-NF session, participants reported the strategies they used to up-regulate and down-regulate the TPJ, which we subjected to text analysis to better understand the strategies associated with volitional control. Using all data collected, we addressed the following questions: (1) can participants volitionally control the ToM network during training, and does this generalize to the transfer task when no neurofeedback is provided; (2) is success in self-regulating the TPJ associated with the ability to vividly visualize scenes, trait perspective-taking, and trait empathic concern; (3) what strategies are associated with volitional control of the TPJ; (4) does performance on tasks associated with the TPJ (e.g., ToM, attentional reorienting) change after rtfMRI-NF; and (5) is volitional control of the ToM network associated with changes in behavioral performance on measures related to TPJ function.

2. Material and methods

2.1. Open science practices

This study was pre-registered on the Open Science Framework (https://osf.io/cfut6). Due to additional funding, we were able to recruit more participants than the N described in our preregistration. Data for one preregistered behavioral task (*Own-Body Transformation Task*), was unable to be analyzed because of a technical error in the stimulus presentation code. Data from the primary pre-registered analyses described in the current manuscript and analysis code are available on the Open Science Framework (https://osf.io/jbnpt/?view_only=2582ccd3cda644eb9f267928c8ca4688). Whole-brain random effects maps are available upon reasonable request. The *Consensus on the Reporting and Experimental Design of Clinical and Cognitive-Behavioural Neurofeedback Studies (CRED-nf)* best practices checklist (Ros et al., 2020) is included in the Supplementary Material.

2.2. Participants

Enrollment was open to individuals of any sex, gender, race, and ethnicity who were between the ages of 18–65 years, fluent in English, and had normal or corrected-to-normal vision and hearing. Exclusion criteria were a current mental disorder as assessed with the Structured Clinical Interview for DSM-5 Disorders (First et al., 2015), prior psychiatric hospitalization, first-degree relative with a schizophrenia spectrum disorder, cognitive impairment (IQ≤70) as assessed with the Wechsler Abbreviated Scale of Intelligence (Wechsler, 2011), neurological disorder, and the presence of an MRI contraindicator.

Sixteen participants were recruited from the greater Rochester area through prior involvement with the lab's research, ResearchMatch, or the University of Rochester Clinical & Translational Science Institute Health Research Website. Participants were on average 46 years old (*SD*=14, range=19–65; data were missing for two participants), predominantly female at-birth (56%), identified as women (50%; 38% male, 6% genderqueer, 6% genderfluid), racially White (75%; 19% Black or African American, 6% Multiracial) and non-Hispanic/Latino (100%), married (52%; 25% single/never married, 19% divorced), with a Master's Degree (50%; 19% high school or equivalent, 6% Associate's Degree, 25% Bachelor's Degree).

Participants were monetarily compensated for their time. Written informed consent was obtained from all participants in accordance with the Declaration of Helsinki and local guidelines. The study was approved by the University of Rochester Research Subjects Review Board.

2.3. Sample size and power

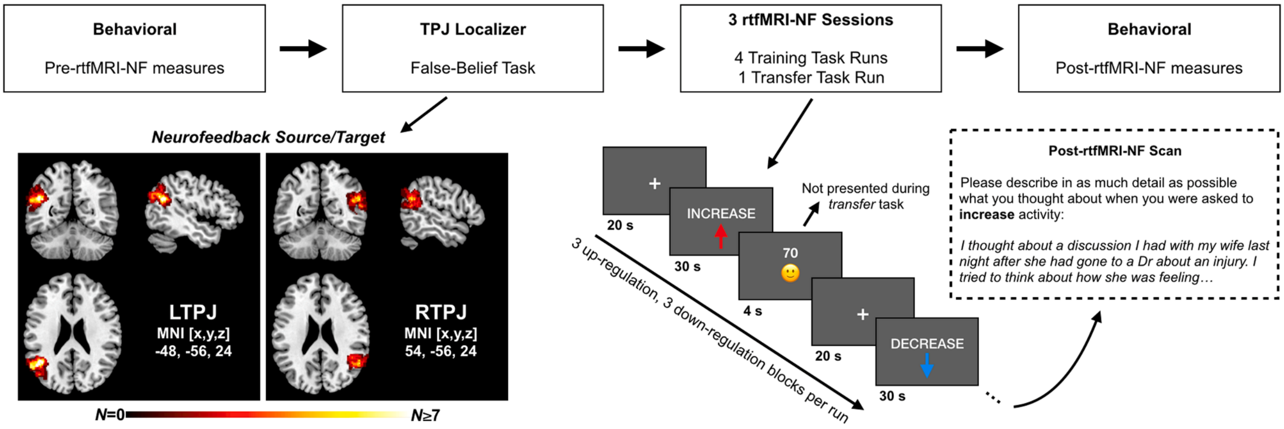
Sample size was determined by financial constraints. Given the analytic strategy described below, assuming a conservative correlation of r=0.50 between repeated-measures, N=16 afforded us 80% power to detect within-subject condition differences (i.e., up- versus down-regulation, pre- versus post-rtfMRI-NF behavioral performance; alpha=0.05) of f=0.37 (d=0.74). We could detect correlations of r=0.43 (alpha=0.05, one-tailed). We note that increasing research suggests that large samples are needed to obtain reliable estimates of brain-behavior associations (Grady et al., 2021; Marek et al., 2022), and any effects described herein are likely to overestimate the true magnitude of association.

2.4. Design and procedure

This was a single-arm study in which all participants received active real-time fMRI; there was no non-active control condition (see Sorger et al., 2019 for a discussion on this limitation). The study was conducted in 5–6 separate study visits (n=6, 38% completed the post-rtfMRI-NF behavioral session on the same day as their third and last rtfMRI-NF session; Fig. 1). In the first session, participants completed eligibility assessments and behavioral measures assessing aspects of cognition putatively associated with neural activity in the ToM network. In the second session, participants underwent fMRI while completing the TPJ localizer task. In the third through fifth session, participants underwent fMRI while completing the rtfMRI procedures. After each of these scan sessions, participants described what they thought of while up- and down-regulating their TPJ and rated their enjoyment and difficulty of self-regulating the neural target. In the final session, participants completed the same behavioral assessments completed in the initial behavioral session. The median interval between rtfMRI-NF sessions was 4.5 days and the median interval between the last rtfMRI-NF session and post- rtfMRI-NF behavioral assessment was 2.5 days.

2.5. MRI data acquisition and analysis

MRI data were collected on a 3T Siemens Prisma scanner at the University of Rochester Center for Advanced Brain Imaging & Neurophysiology with a 64-channel head coil. We acquired an anatomical image with a T1-weighted MPRAGE sequence (192 sagittal slices, voxel size=1 × 1 × 1 mm^3). We collected functional data using an echo-planar imaging (EPI) sequence (TR=2000 ms, TE=30 ms, flip angle=90°, FoV=220 mm, 58 axial slices, voxel size=2 × 2 × 2 mm^3). fMRI data were preprocessed in SPM12 in the following steps: realigned to the first image, co-registered to the anatomical, normalized to the MNI template, and smoothed using a 6 mm FWHM Gaussian kernel. We used the Artifact Detection Tools (www.nitrc.org/projects/artifact_detect/) to identify outlier scans in global signal (±3 *SD*) and motion (>1 mm of composite motion relative to prior scan). fMRI data were also analyzed in SPM12 in the whole brain using GLMs that included terms for task conditions (see below) convolved with the standard hemodynamic response function, and nuisance regressors for the movement



Please describe in as much detail as possible what you thought about when you were asked to **increase** activity:

I thought about a discussion I had with my wife last night after she had gone to a Dr about an injury. I tried to think about how she was feeling…

Fig. 1. Study design.
Note. Neurofeedback source/target image depicts the overlap of participant LTPJ/RTPJ ROIs identified from the False-Belief Task and used as the source/target of the neurofeedback. MNI coordinates depict peak overlap of each participant's TPJ. Post-rtfMRI-NF, in addition to reporting strategies used while self-regulating, participants reported the number of different strategies used in a given block on average, and their enjoyment of and difficulty in self-regulating the TPJ.

parameters and outlier scans identified with the Artifact Detection Tools. Data were high-pass filtered at 128 s.

2.6. Localization of the TPJ for neurofeedback

The TPJ was localized using the *False-Belief Task* (Dodell-Feder et al., 2011; Saxe & Kanwisher, 2003), which is one of the most widely-used tasks for assessing the neuroanatomical basis of ToM in the fMRI literature (Schurz et al., 2014, 2021). The false-belief task depicts individuals who have a belief that is inconsistent with reality (e.g., a person who believes an object is in location A, but the object was surreptitiously moved by another person to location B). These scenarios present a compelling test of belief understanding since the person's actions would differ if they acted in accordance with their beliefs (searching in location A) versus the true state of affairs (searching in location B) (Saxe & Kanwisher, 2003; Wellman et al., 2001). These stories are contrasted with false-photograph and map stories that similarly depict outdated representations of the world (e.g., a photograph that no longer accurately depicts a landscape due to erosion). As argued by others (Schaafsma et al., 2015), ToM encompasses a range of psychological processes involving the representation of mental states, and the false-belief task can be said to assess one specific component of ToM. That said, fMRI meta-analyses of ToM consistently implicate the TPJ as part of a core network, recruited across many different types of ToM tasks and levels of mental state reasoning (Molenberghs et al., 2016; Schurz et al., 2014, 2021). Thus, while it is possible that a different ToM localizer would have led to the selection of a different cluster of voxels in the TPJ, we do not have reason to believe that these differences would be substantive in terms of anatomy or functional profile.

In the version of the task used, participants read 10 stories describing outdated (i.e., false) beliefs and 10 stories describing outdated physical representations in the world (i.e., photos, maps), divided into two functional runs. After each story, participants answered a *true/false* question about the story. Story order was pseudo-randomized and presented for 12 s, followed by the question for 6 s, and then 12 s of fixation on a central cross. The task was presented with MATLAB and Psychophysics Toolbox (Brainard, 1997; Kleiner et al., 2007).

Individual subject contrasts were generated for belief>physical representation following the analysis procedures described above. We used data from 462 neurotypical participants who completed the *False-Belief Task* (Dufour et al., 2013) as a guide for ROI selection. To select neurofeedback targets, images were viewed using a voxel-wise family wise error rate (FWER)-corrected threshold of $p<.05$, $k>20$. In order to ensure we had at least one TPJ to use as the neurofeedback target, if we could not localize either TPJ at that threshold, we iteratively lowered the threshold to $p<.0001$ uncorrected and then $p<.001$ uncorrected. We were able to localize left and right TPJ in 12 participants; in 4 participants we were able to localize just LTPJ ($n=2$) or RTPJ ($n=2$). The entire cluster that surpassed the threshold was used as the neurofeedback target (Fig. 1). We note that ROI size ranged considerably as a result, number of voxels $M \pm SD=230 \pm 186$, min=23, max=698.

2.7. rtfMRI-NF

At the start of the first rtfMRI-NF session, participants were oriented to the rtfMRI procedures. We described the function of the TPJ to participants in plain language (i.e., being responsive to information about mental states). Participants were then asked to generate strategies for increasing and decreasing neural activity in this region based on the information provided to ensure that as a starting point, strategies involved some social content (e.g., thinking about a recent interaction with a friend). We emphasized that though participants could use these strategies to begin with, they should alter their strategy based on the neurofeedback. These procedures align with other rtfMRI studies (e.g., Sukhodolsky et al., 2020) and current recommendations (Fede et al., 2020).

RtfMRI was implemented by exporting the imaging data using scripts from *Multivariate and Univariate Real-Time Functional Imaging* software (Hinds et al., 2011) to a data analysis workstation running *OpenNFT* software (Koush et al., 2017), which was used to process the data in real-time (see Koush et al., 2017 for details) and present the feedback signal to the participant. In each of the three rtfMRI-NF sessions, participants performed two tasks in the scanner: four runs of a *training task* in which they received intermittent, activation-based feedback from the TPJ, followed by one run of a *transfer task* in which no feedback was provided as a test of learning. For each run of the training task, participants completed six blocks where they fixated on a central cross for 20 s (i.e., baseline), were directed to up-regulate or down-regulate their TPJ for 30 s (three blocks of up-regulation and three blocks of down-regulation per run), and then received feedback for 4 s (Fig. 1). Feedback was calculated as the median percent signal change in the left and/or right TPJ in the regulation period relative to the prior baseline period and adaptively scaled to stay within each ROI's mean of the highest 5% and lowest 5% of data points acquired (Koush et al., 2012). These values were converted to a number between 0–100 and visually displayed to the participant along with a smiley face scaled to the success of that regulation period (e.g., for an up-regulation block, a higher number would be accompanied by a larger smile). The transfer task was the same as the training task except no feedback was presented to the participant after the regulation period.

After each of the three rtfMRI sessions, participants completed a survey in which they reported the strategies they used when regulating the neural signal (i.e., what they thought about during the regulation blocks), how many different strategies they used on average during the regulation blocks, and how enjoyable and difficult they found the neurofeedback procedure to be.

2.8. Assessment of possible moderators

Participants completed measures that we predicted may impact one's ability to gain volitional control over the network. This included one's ability to generate vivid imagery for past social scenarios that may serve as the basis for effective self-regulation strategies and trait perspective-taking. We also included a measure of trait empathic concern. Although there exists a preponderance of evidence that empathy and emotion-based processes are distinct from ToM, there is also evidence to suggest that these processes are fundamentally intertwined (Preckel et al., 2018; Zaki & Ochsner, 2012). Thus, one might reasonably expect that empathic individuals, as with individuals who frequently engage in perspective-taking, may be more attuned to others' internal states in a way that might facilitate strategy selection and self-regulation.

2.8.1. Visual imagery

Visual imagery was assessed with the Vividness of Visual Imagery Questionnaire (Marks, 1973), which is a 16-item self-report questionnaire in which participants are asked to generate different visual images and then rate how vivid each image was using a 1 (*No image at all, you only "know" that you are thinking of the object*) to 5 (*Perfectly clear and as vivid as normal vision*) scale. Scores are calculated as the sum of all items.

2.8.2. Trait perspective-taking and empathic concern

Trait perspective-taking and empathic concern were assessed with the Interpersonal Reactivity Index (Davis, 1983), which is a multidimensional scale of empathy. The Perspective-Taking and Empathic Concern subscales assess the tendency to consider others' perspectives and experience compassion and concern for others, respectively. Each subscale consists of 7 items rated with a 0 (*does not describe me well*) to 4 (*describes me very well*) scale. Scores are calculated by summing the items of each subscale.

2.9. Behavioral outcome measures

Participants completed a battery of measures pre- and post-rtfMRI-NF assessing aspects of cognition associated with the TPJ including measures of mental state attribution and attentional reorienting.

2.9.1. Hinting task

In the Hinting Task (Corcoran et al., 1995; Klein et al., 2020), participants are read 10 vignettes and are asked to infer a character's intent from a hint provided by that character. Correct responses are given a score of 2. If the participant provides an inaccurate assessment of the character's intent, they are provided with an additional clue, and can earn 1 point for an accurate assessment. If the participant again provides an inaccurate assessment, no additional clues are given and the response is scored a 0. Scores are summed and range from 0–20. Prior work using a revised, more stringent scoring criteria described in Klein et al. (2020) and used in the current study shows that samples without mental disorders demonstrate test-retest reliability estimates of r=0.55, small magnitude practice effects with repeated testing (d_z=0.22), and minimal ceiling effects.

2.9.2. Social attribution task-multiple choice

In the Social Attribution Task-Multiple Choice (Bell et al., 2010), participants view a silent 64 s animation of geometric objects acting with ostensible agency and social intention. The animation is stopped periodically during which participants are asked a total of 19 questions (e.g., "What are the two triangles doing?"), each of which is presented with four response options: one describes the correct social inference (e.g., fighting), two describe incorrect social inferences, and one describes a nonsocial inference (e.g., rotating). Scores can range from 0–19. The most comprehensive evaluation of the task's utility as a repeated measure for samples without mental disorders (Pinkham et al., 2017) combined alternate versions of the form, which were not used here. However, those data, combing both forms, demonstrate test-retest reliability of r=0.55, small magnitude practice effects (d_z=0.31), and minimal ceiling effects.

2.9.3. Multiracial emotion identification task

In the Multiracial Emotion Identification Task (Dodell-Feder et al., 2020), participants viewed the faces of 48 individuals of varying ages and racial/ethnic groups and judged whether the face has a happy, sad, angry, or fearful expression. Participants have 10 s to respond while the face is on the screen, and then another 10 s to respond after the face is removed from the screen. The reported score is proportion correct. Although data that speak to the measure's utility as a repeated measure are not available, data from a nearly identical emotion identification task, the Penn Emotion Recognition Test 40 (Kohler et al., 2003), demonstrate test-retest reliability of r=0.68, small magnitude practice effects (d_z=0.12), and no ceiling effects (Pinkham et al., 2017). The task was presented with MATLAB and Psychophysics Toolbox. As mentioned above, though emotion identification tasks are thought to be associated with a neural network different from that subserving ToM (Schurz et al., 2021), other work has demonstrated their neural bases to be jointly activated and/or directly influenced by one another (Kanske et al., 2016; Lamm et al., 2011; Schurz et al., 2021; Zaki et al., 2009), suggesting that changes to the ToM network may carry consequences for emotion processing.

2.9.4. Spontaneous theory of mind protocol

In the Spontaneous Theory of Mind Protocol (Rice & Redcay, 2015), participants view a silent 2 min clip of the films *Rear Window* and *John Tucker Must Die* without instruction. After viewing each clip, participants are asked to describe what they saw in 7–10 typed sentences. We coded responses for the amount of spontaneous mental state content by submitting written responses to Linguistic Inquiry and Word Count (LIWC) software (Pennebaker et al., 2015). Using LIWC, we calculated the proportion of words that fell into the following categories: affect (e.g., "emotional", "love", "jealous"), positive emotion (e.g., "happy", "like", "trust"), negative emotion (e.g., "afraid", "sad", "uncomfortable"), and insight (e.g., "believe", "knows", "understands"). To our knowledge, no data exist that speak to the measure's utility as a repeated measure.

2.9.5. Mental state fluency task

The Mental State Fluency Task is a novel measure created for the purposes of the current study that was designed to assess the fluency with which one can infer the mental states of real-world social partners. Based in part on verbal fluency (Lezak, 2012) and future fluency measures (MacLeod et al., 1993), in this task, participants were asked to identify a positive, and separately, a negative meaningful social interaction they experienced in the last year. After providing a brief description of the event, participants were given 60 s to report what their interaction partner might have been reasonably thinking and feeling during the interaction, which the experimenter recorded in a Qualtrics survey. Afterwards, participants rated their confidence in the extent to which their interaction partner actually experienced each mental state described by the participant using a percentage (i.e., 0–100% confidence). As the outcome variable, we calculated the total number of mental states generated weighted by the perceived likelihood of their occurrence (i.e., the product of the number of mental states and the confidence ratings). As this is a novel measure, no data exist that speak to its utility as a repeated measure.

2.9.6. Attentional cueing task

As the TPJ has also been associated with attentional reorienting (Corbetta et al., 2000; Corbetta & Shulman, 2002; Mitchell, 2008), we included an Attentional Cueing Task adapted from Krall et al. (2016) and Vossel et al. (2009) that has been shown to recruit the TPJ. In each of the 200 task trials, participants fixate on a central cross flanked on the left and right side by two empty boxes for 2000 ms. The central cross is replaced by a cue in the form of an arrow pointing either to the left or right box for 200 ms, after which the arrow is replaced by a central cross for 400 or 700 ms. Finally, the target—a white asterisk—appears in either the left or right box for 100 ms. In 160 trials (80%) the target is validly cued; that is, the arrow correctly points to the box in which the asterisk subsequently appears. In 40 trials (20%), the target is invalidly cued. Participants are instructed to indicate which box the target appears as quickly and accurately as possible. The main experimental task is preceded by 10 practice trials with feedback. Following others (Krall et al., 2016), we calculated inverse efficiency scores as *M* reaction time divided by proportion correct, separately for valid and invalid trials. The task was presented with MATLAB and Psychophysics Toolbox.

2.10. Data analysis

Unless otherwise stated, data were analyzed and visualized in R Statistical Software (R Core Team, 2022) and R Studio (RStudio Team, 2020) using the following packages: psych (Revelle, 2022), rstatix (Kassambara, 2021), bootES (Kirby & Gerlanc, 2013), pls (Liland et al., 2021), confintr (Mayer, 2022), ggplot2 (Wickham, 2016), and ggpubr (Kassambara, 2020).

All data were visually inspected. Outcomes in which we observed outliers (values 1.5x the interquartile range above the third quartile or below the first quartile) were subjected to 90% Winsorization.

2.10.1. fMRI data

Our primary question was whether participants demonstrated volitional control of the ToM network during the training task, and more critically, during the transfer task, which, given that no active neurofeedback is provided, serves as a test of learning. To address this question, we conducted ROI analysis in regions of the ToM network. Similar to our localization of the TPJ neurofeedback target, we used the belief>physical representation contrast from the False-Belief Task, thresholded at $p<.001$, $k>10$, uncorrected, to individually-localize the following regions of the ToM network using the Dufour et al. (2013) boundaries and neurotypical group maps as a guide (n refers to number of participants in which the ROI could be localized): LTPJ (n=15), RTPJ (n=14), right superior temporal sulcus (RSTS; n=12), precuneus (PC; n=15), dorsal medial prefrontal cortex (DMPFC; z coordinate $\geq$20; n=12), middle medial prefrontal cortex (MMPFC; $20\geq z\geq 0$; n=11), ventral medial prefrontal cortex (VMPFC; $z\leq 0$; n=11). Since not every ROI could be identified in every participant, in order to preserve power, we defined ROIs from the neurotypical groups maps (belief>physical representation) described in Dufour et al. as a 6 mm sphere surrounding the peak coordinate in that region, and used these group ROIs for any participant in which a given ROI could not be individually-localized. Using these ROIs, we extracted beta values from first-level maps for up-regulation>baseline and down-regulation>baseline, for each of the three rtfMRI-NF sessions, separately for the training and transfer tasks. These beta values were submitted to 2 condition (up-regulation, down-regulation) by 3 session repeated-measures ANOVA, conducted separately with training and transfer task data as the outcome, and are accompanied by η^2_G as the measure of effect size. We also provide the effect size for up-regulation versus down-regulation collapsing across session—our index of volitional control—as Cohen's d_z, which is accompanied by 95% bias-corrected and accelerated (BCa) CI derived from 10,000 bootstrap samples. Given the large age range of the sample, we repeated these analyses including age as a covariate; findings were unchanged.

We followed-up the hypothesis-driven ROI analysis with exploratory whole-brain analysis in SPM12 by conducting a one-sample t-test on up-regulation>down-regulation maps (collapsing across session) from the first-level GLMs. We report findings at a voxel-wise FWER-corrected $p<.05$, $k>20$ and uncorrected $p<.001$, $k>20$ thresholds.

2.10.2. Analysis of self-regulation strategy

Descriptions of the self-regulation strategies used by participants were analyzed with LIWC, which calculated the proportion of words in 46 psychological categories (we excluded syntactic [e.g., prepositions], metadata [e.g., number of words], and semantic categories indexing informal language [e.g., netspeak, such as "btw"]; see Supplementary Materials for category abbreviations; complete description of the categories and example words is available at the following link: https://www.liwc.app/help/psychometrics-manuals). To determine which categories were most strongly associated with volitional control, we used partial least squares regression (PLSR) with the LIWC categories as predictors and the difference between up-regulation and down-regulation (collapsed across session given that we did not find a condition by session interaction) for the transfer task as the outcome. PLSR is a multivariate data analytic technique that identifies components that best predict an outcome. PLSR is well suited to the question of how strategies relate to self-regulation as we wanted to maximize the variance explained in volitional control while reducing the dimensionality of the LIWC data. Additionally, PLSR has been used in similar fMRI-LIWC applications (Finn et al., 2018). We used data from the transfer task as we viewed it to be the most important outcome, indexing whether learning during the training task generalized to a context without active neurofeedback (as in the participants' daily lives). For those interested, full results using the training task data are provided in the Supplementary Materials. We selected the number of components that resulted in the smallest root mean square error of prediction (RMSEP) based on leave-one-out, bias-corrected cross-validated predictions (Mevik & Wehrens, 2007). Thus, for each ROI, we generated an estimate of the number of components that jointly best explained the LIWC data (separately for up- and down-regulation) and volitional control on the transfer task, the amount of variance in LIWC and volitional control explained by the components, and the feature loadings on the selected components.

2.10.3. Behavioral data

For behavioral outcome measures with a single condition (i.e., Hinting Task, Social Attribution Test, Multiracial Emotion Identification Task, Spontaneous Theory of Mind Protocol), we analyzed changes in pre-to-post-rtfMRI-NF behavioral performance with paired-samples Welch's t-tests and report d_z with 95% BCa CI from 10,000 bootstrap samples as the effect size. Tests were one-tailed given our directional hypotheses (Lakens, 2017). For behavioral outcome measures with more than one condition (Mental State Fluency task: positive social experience condition, negative social experience condition; Attentional Cueing Task: validly-cued trials, invalidly-cued trials), we analyzed data with 2 condition by 2 time (pre-rtfMRI-NF, post-rtfMRI-NF) repeated measures ANOVA and report η^2_G as the effect size.

2.10.4. Brain-behavior associations

We evaluated whether the putative moderators were associated with volitional control by conducting Spearman rank correlations (one-tailed given our directional hypotheses) between the moderator and volitional control (up-regulation versus down-regulation) on the transfer task, separately for each ROI. We evaluated whether changes in behavioral performance were associated with volitional control by conducting similar correlations between pre-to-post changes in behavioral performance and volitional control on the transfer task, separately for each ROI. Correlations are accompanied by 95% BCa CI derived from 10,000 bootstrap samples.

3. Results

3.1. NF experience data

On a 0–100 point scale, participants rated the NF task as moderately enjoyable, M=45, 95% CI [38, 52], and moderately difficult, M=50, 95% CI [44, 56]. Enjoyment was negatively associated with difficulty, although the association was not unexpected under the null hypothesis, r_s=-0.08, 95% CI [-0.66, 0.50], p=.757. Across the three sessions, there were no differences in enjoyment, session comparison $bs<3.7$, $ps>0.418$, or difficulty, session comparison $bs<5.7$, $ps>0.137$. When up- and down-regulating the neural signal, participants most often thought about a single thing (e.g., a single social experience when up-regulating or a single object when down-regulating; up=48%, down=52%), followed by two-three things (up=29%, down=35%), four-six things (up=21%, down=8%), and finally, more than six things (up=2%, down=4%).

3.2. NF effects on the brain

3.2.1. ROI analysis

Our main question was whether NF led to control of the ToM network—defined here as greater neural activity for up- versus down-regulation—during the training and transfer task. To test this question, we evaluated the effect of regulation direction (up- versus down-regulation), session, and their interaction on seven individually-localized ROIs. On the training task, across all ROIs except for RSTS, we found greater neural activity for up- versus down-regulation with the difference being consistently large in magnitude (range d_z=1.05 [LTPJ] – 1.86 [PC]; Table 1, Fig. 2). Stating the effect sizes differently (i.e., the probability of superiority; Ruscio, 2008), there is a 77% chance that a randomly selected up-regulation value was greater than a randomly selected down-regulation value in the region demonstrating the smallest, yet significant effect (LTPJ), and a 91% chance that a randomly selected up-regulation value was greater than a randomly selected down-regulation value in the region demonstrating the largest effect (PC). We found no effect of session, nor a regulation by time interaction.

On the transfer task in which no NF was provided, across all ROIs except for MMPFC and RTPJ, neural activity was greater for up- versus down-regulation (Table 1, Fig. 2). The effects were generally smaller

Table 1
ROI analysis results.

ROI	Task	Term	F [a]	η^2_G	d_z [95% CI] [b]
DMPFC					
	Training				
		Regulation	**18.61*****	**.338**	**1.08 [.57, 1.73]**
		Session	.78	.009	
		Interaction	3.17	.012	
	Transfer				
		Regulation	**13.45****	**.207**	**.92 [.24, 1.69]**
		Session	.12	.002	
		Interaction	1.93	.021	
LTPJ					
	Training				
		Regulation	**17.77*****	**.286**	**1.05 [.60, 1.64]**
		Session	.19	.002	
		Interaction	.51	.002	
	Transfer				
		Regulation	**6.68***	**.107**	**.65 [.19, 1.16]**
		Session	.40	.009	
		Interaction	.29	.004	
MMPFC					
	Training				
		Regulation	**19.97*****	**.279**	**1.12 [.62, 1.76]**
		Session	.03	.001	
		Interaction	1.89	.015	
	Transfer				
		Regulation	4.27	.093	.52 [-.06, 1.21]
		Session	.93	.013	
		Interaction	.56	.007	
PC					
	Training				
		Regulation	**55.30*****	**.496**	**1.86 [1.01, 3.00]**
		Session	2.13	.022	
		Interaction	.83	.006	
	Transfer				
		Regulation	**34.71*****	**.281**	**1.47 [.85, 2.22]**
		Session	1.19	.015	
		Interaction	1.26	.020	
RSTS					
	Training				
		Regulation	4.33	.065	.52 [-.003, 1.08]
		Session	1.50	.013	
		Interaction	3.12	.022	
	Transfer				
		Regulation	**6.29***	**.084**	**.63 [.22, 1.04]**
		Session	1.02	.015	
		Interaction	1.46	.015	
RTPJ					
	Training				
		Regulation	**22.92*****	**.177**	**1.20 [.62, 1.85]**
		Session	1.66	.023	
		Interaction	.84	.007	
	Transfer				
		Regulation	2.12	.020	.36 [-.17, .79]
		Session	1.78	.031	
		Interaction	1.65	.034	
VMPFC					
	Training				
		Regulation	**25.56*****	**.284**	**1.26 [.67, 2.08]**
		Session	.24	.004	
		Interaction	3.23	.022	
	Transfer				
		Regulation	**40.42*****	**.247**	**1.59 [.94, 2.36]**
		Session	.08	.002	
		Interaction	1.31	.020	

Note. Results from repeated-measures ANOVAs testing the effect of regulation (up, down), session (1, 2, 3), and the interaction between these terms on neural activity in the ROIs (beta values) from the training and transfer tasks. Findings that are unexpected under the null hypothesis ($p<.05$, unadjusted for multiple tests) are in bold text. DMPFC=dorsal medial prefrontal cortex, LTPJ=left temporo-parietal junction, MMPFC=middle medial prefrontal cortex, PC=precuneus, RSTS=right superior temporal sulcus, RTPJ=right temporo-parietal junction, VMPFC=ventral medial prefrontal cortex.

* $p<.05$
** $p<.01$
*** $p<.001$

[a] Regulation term df=1,15; Session df=2,30; Regulation*Session df=2,30.

[b] Effect size for up- vs down-regulation averaged across all three sessions. Values are accompanied by BCa 95% CIs generated from 10,000 bootstrap samples.

than those observed for the training task, ranging from medium-to-large in those ROIs that showed an effect (range d_z=.63 [RSTS] – 1.59 [VMPFC]). Stated otherwise, there was a 67% and 87% chance of a randomly selected up-regulation value being higher than a randomly selected down-regulation value in the ROIs showing the smallest yet significant effect and the largest effect, respectively (i.e., RSTS and VMPFC). Similar to the training task, there was no effect of session nor a regulation by session interaction meaning that the positive effect of NF was stable and obtained after just a single session. Together, the ROI analyses showed that volitional control was achieved in most of the ToM-related brain regions.

3.2.2. *Exploratory whole-brain analysis*

On the training task, an exploratory random-effects whole-brain analysis revealed greater neural activity for up- versus down-regulation in PC, LTPJ, and dorsal to ventral MPFC, as well as other regions implicated in social cognition, such as the cerebellum and left temporal cortex, and regions not typically implicated in social cognition, such as the hippocampus, parahippocampal cortex, the caudate, and visual cortex (Table 2, Fig. 3). However, no region survived voxel-wise FWER-correction. Findings were less robust for the transfer task (Table 2, Fig. 3). Specifically, whole-brain random-effects analysis revealed greater neural activity for up- versus down-regulation in PC, LTPJ, and VMPFC. Similarly, no effects were observed with voxel-wise FWER-correction.

3.2.3. *Moderators of volitional control*

We tested whether the ability to vividly imagine scenarios, perspective-taking, and empathic concern facilitated volitional control separately for each ROI, which we calculated as the difference between up- and down-regulation on the transfer task averaged across all sessions. On the vividness of visualizing imagery, the associations were in the predicted direction for most ROIs, such that greater ability to vividly visualize imagery was associated with a greater difference between up- and down-regulation (Table 3), and ranged in size from small to large. Surprisingly, we observed negative associations with the two NF targets. However, none of these associations were unexpected under the null hypothesis ($ps>.05$). On perspective-taking, all associations were in the predicted direction, such that greater perspective-taking was associated with greater volitional control on the transfer task, and ranged from small to large, but none of the associations were unexpected under the null hypothesis. On trait empathic concern, contrary to our predictions, most associations were negative, such that greater empathic concern was related to less volitional control on the transfer task. These associations effects were small in magnitude, and none were unexpected under the null hypothesis.

3.3. *Regulation strategies associated with volitional control of the ToM network*

Towards understanding what mental processes facilitated volitional control of the ToM network (i.e., difference between up- and down-regulation on the transfer task), we characterized written descriptions of the strategies used by participants for up- and down-regulating the neural signal, and submitted those data to a text analysis that calculated the proportion of words that fell into 46 psychological categories. We used partial least squares regression to then characterize how principal components generated from these categories were associated with learned control of the ROIs. On up-regulation strategies, the psychological features were able to explain variance in all ROIs except for PC and VMPFC, where an intercept-model was the best fit to the data. In

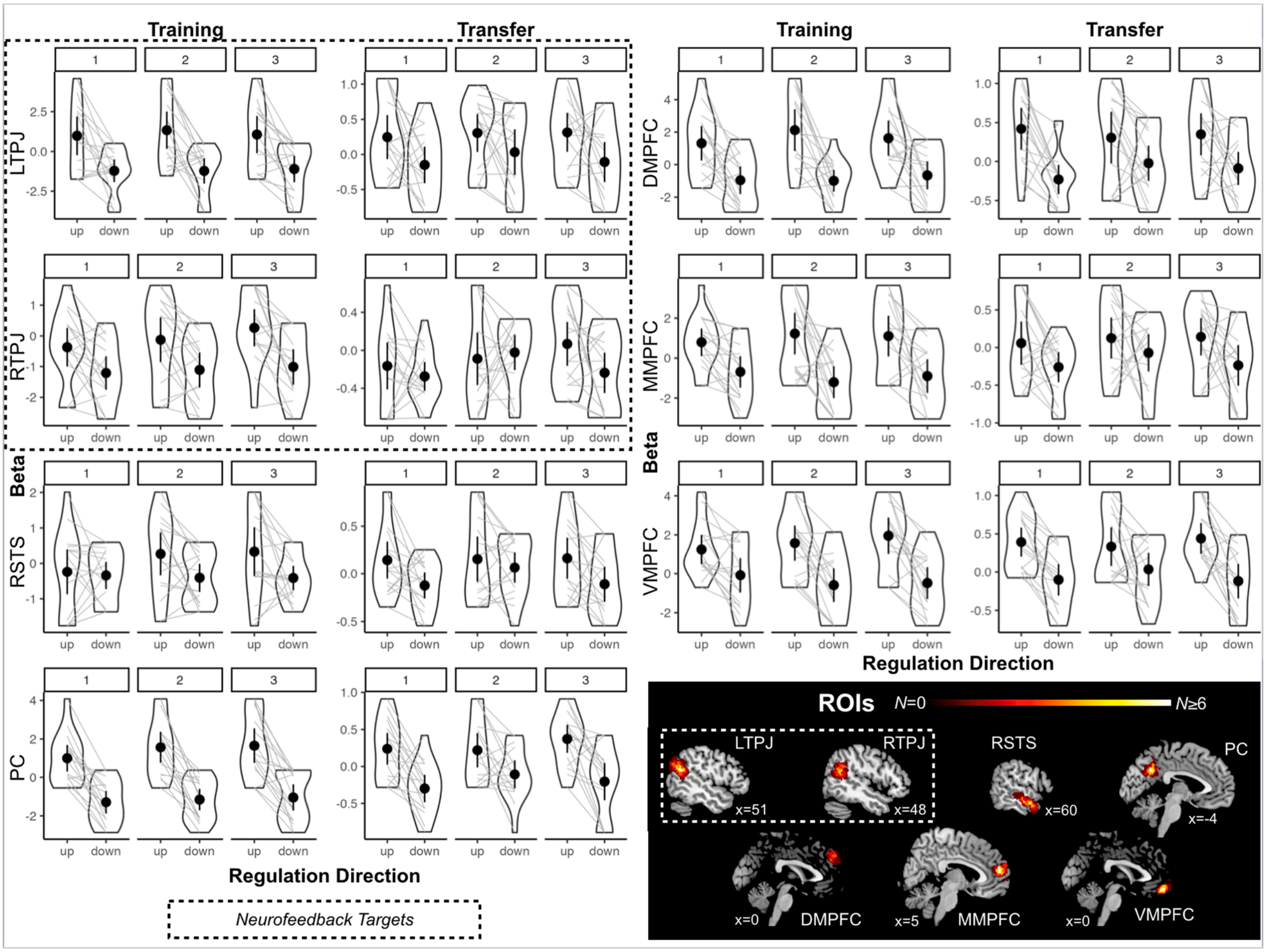


Fig. 2. ROI Analysis Results.

Note. Plots depict beta values extracted for up-regulation>baseline ("up") and down-regulation>baseline ("down") as a function of rtfMRI-NF session (1, 2, 3) in each of the ROIs. Training task data is on the left and transfer task data on the right of each facet. Light gray lines connect paired participant data. Solid dots and error bars depict mean beta values +/- 95% CI. The panel at the bottom right depict the overlap of individually-localized ROIs identified using the belief>physical representation contrast from the False-Belief Task.

DMPFC, MMPFC, RSTS, and RTPJ, the data were best predicted (i.e., resulted in the smallest RMSEP) by a single component, which explained roughly half the variance in the psychological dimensions (53% in DMPFC and RTPJ, 52% in MMPFC, 50% in RSTS), but less than a quarter of the variance in learned control of the ROIs (21% in DMPFC, 24% in MMPFC, 21% in RSTS, 18% in RTPJ). The features with the highest positive loadings across all ROIs were social (e.g., "awkward," "relationship," "party"), affiliation (e.g., "friend," "compassion," "flirting"), and drives (e.g., "success", "bully", "benefit"). In LTPJ, the data were best predicted by 8 components, which explained 95% of the variance in the psychological dimensions, and 99% of the variance in volitional control (Fig. 4). The first three components—which explained roughly three-quarters of the variance in both psychological features (76%) and volitional control of LTPJ (74%)—were similarly characterized by (1) social, affiliation, and drives, (2) time (e.g., "end," "until," "season") and a focus on past experiences (e.g., "did," "talked," "believed"), and (3) work (e.g., "jobs," "office," "work") and male references (e.g., "boy," "his," "dad"), respectively.

On down-regulation strategies, the psychological features were able to explain variance in RSTS and VMPFC (an intercept-model best fit the data for DMPFC, MMPFC, LTPJ, PC, and RTPJ). In both of these ROIs, the data were best predicted by a single component, which explained over one-third of variance in the psychological dimensions (37% in RSTS, 39% in VMPFC) and between approximately one-sixth to one-quarter of the variance in learned control (28% in RSTS, 19% in VMPFC). In all of the ROIs, the highest loading positive features were perceptual processes (e.g., "look," "heard," "feeling") and seeing (e.g., "view," "saw," "seen"). More specifically, many participants reported "staring" at or "seeing" objects on the "screen," and/or visualizing inanimate objects like tables and chairs, although these latter objects were not captured by the LIWC categories.

Results using the training data are reported in the Supplementary Materials. Generally, we were able to explain variance in more ROIs, and more of that variance, with the training data. That said, the features with the highest loading—i.e., the strategies that best explained ROI responses—were the same as those identified with the transfer ROI data.

3.4. Volitional control and behavior

3.4.1. Behavioral performance

To evaluate the possible impact of volitional control of the ToM network on aspects of cognition associated with the TPJ (e.g., mental state attribution, attentional reorienting) and ToM network more broadly (i.e., social cognition), we first evaluated whether there were any changes in performance on the behavioral tasks pre- to post-rtfMRI-NF training. On the social cognitive tasks, performance on the Hinting Task and Social Attribution Task improved from pre- to post-rtfMRI-NF, with the differences being medium and small, respectively; however, the

Table 2
Exploratory random-effects whole-brain analysis results for up-regulation>-down-regulation.

Task	Region	MNI Coordinates (*x, y, x*)	Cluster Extent (voxels)	Peak *t*-value (*p*<.001)
Training				
	Precuneus	-4, -62, 36	3095	9.10
	L Caudate	-18, 32, 0	452	8.35
	R Cerebellum	14, -54, -20	103	7.64
	R Caudate	20, -18, 28	525	7.63
	L Cerebellum	-16, -56, -20	154	6.63
	R Hippocampus	40, -40, 4	97	6.39
	R Calcarine Sulcus	22, -82, 4	86	6.26
	L Hippocampus	-40, -32, -6	99	5.83
	Ventral Medial Prefrontal Cortex	-2, 52, -8	289	5.23
	L Middle Temporal Gyrus	-58, -6, -18	71	5.00
	Dorsal Medial Prefrontal Cortex	-2, 58, 32	46	4.13
Transfer				
	R Hippocampus	30, -44, 12	161	6.65
	L Hippocampus	-32, -40, 4	32	5.46
	Precuneus	-4, -44, 32	450	5.32
	Ventral Medial Prefrontal Cortex	2, 54, -10	115	5.23
	L Angular Gyrus	-40, -72, 40	21	4.26

Note. p<.001, k>20, uncorrected. No regions survive FWER-correction (p<.05, k>20).

differences were not unexpected under the null hypothesis (Table 4). Performance on the Multiracial Emotion Identification Task, Spontaneous Theory of Mind Protocol, and Mental State Fluency Task worsened from pre- to post-rt-fMRI-NF, with the effects being small in magnitude and not unexpected under the null hypothesis. On the attentional cueing paradigm, inverse efficiency scores were lower for valid versus invalid trials as expected, and higher for post- versus pre-rtfMRI-NF, indicating a decrement in performance. These differences were unexpected under the null hypothesis. However, there was no trial type by time interaction.

3.4.2. Associations between volitional control and social cognitive performance

Though performance on the behavioral tasks largely did not change from pre- to post-rtfMRI-NF, small differences in performance may be related to neurofeedback-mediated neural changes. Furthermore, any numerical improvements in performance from pre- to post-rtfMRI-NF, as we observed on the Hinting and Social Attribution Tasks, may simply reflect practice effects. Thus, we evaluated whether neural activity in the ROIs during the transfer task was associated with pre- to post-rtfMRI-NF changes in performance on the behavioral tasks. On the social cognitive measures, the only association that was unexpected under the null hypothesis (p<.05) was a positive correlation between RTPJ and the Hinting Task, such that the greater the NF-related transfer effect, the more positive the change in Hinting Task performance, with the effect being large in magnitude (Table 5). Given the number of tests performed and our sample size, this association, including the magnitude, should be interpreted with caution. Associations between ROI values and task performance were in the predicted direction—greater NF-related ROI activity on the transfer task related to improvements in performance—for the STOMP and Mental State Fluency negative valence task, but not unexpected under the null hypothesis. Associations between ROIs and performance on the other tasks were inconsistent and/or in the opposite direction predicted. On the attentional cueing task, no associations were unexpected under the null hypothesis.

3.5. Exploratory analyses evaluating the utility of neurofeedback for volitional control

To summarize, analyses thus far have demonstrated that participants are able to volitionally control much of the ToM network, even when neurofeedback is not actively being provided. The lack of condition by session interactions though could indicate that neurofeedback is not necessary for volitional control. Instead, the strategy that we initially provided to participants may be sufficient for gaining control of the network. The design of the study (i.e., lack of control group that was provided with the same initial strategy, lack of an initial pre-neurofeedback transfer scan) does not allow us to conclusively rule out this account. Nonetheless, we undertook a set of non-preregistered, exploratory analyses to evaluate potential learning that occurred as the result of neurofeedback.

One possibility is that neurofeedback-related learning across the three sessions is occurring, but may be difficult to detect with the brain data due to small effects (i.e., the regulation by session interactions) that we are underpowered to detect. Such changes may be more readily observable from other data. For example, if participants are using the neurofeedback signal, then we might expect to see changes in strategy across sessions. We evaluated this possibility by looking at changes in strategy—indexed by the LIWC data—across session. Using principal components analysis (with oblimin rotation) as a dimension reduction strategy for the LIWC data, we evaluated changes in components that summarized participants' self-reported strategies, separately for up- and down-regulation, as a function of session (see Supplementary Materials for details). On up-regulations strategies, none of the six identified components differed by session, $Fs \leq 1.88$, $ps \geq .169$, $\eta^2_g s \leq .06$. In contrast, on down-regulation strategies, we found that one of the five identified components, characterized by differentiation (e.g., “hasn't,” “but,”

Table 3
Associations between hypothesized moderators and volitional control.

ROI	Moderator Vividness of Visualizing Imagery	Perspective-Taking	Empathic Concern
DMPFC	.23 [-.35, .63]	.12 [-.42, .59]	.14 [-.38, .64]
LTPJ	-.06 [-.55, .43]	.31 [-.23, .72]	-.03 [-.53, .50]
MMPFC	.19 [-.33, .62]	.02 [-.54, .57]	-.19 [-.63, .44]
PC	.26 [-.34, .70]	.03 [-.57, .55]	-.07 [-.59, .47]
RSTS	.32 [-.28, .76]	.17 [-.47, .69]	-.07 [-.56, .46]
RTPJ	-.11 [-.59, .43]	.29 [-.34, .65]	.16 [-.41, .69]
VMPFC	-.23 [-.68, .29]	.06 [-.55, .62]	-.25 [-.63, .43]

Note. Values represent Spearman's rank correlation coefficient and 95% BCa CI generated from 10,000 bootstrap samples.

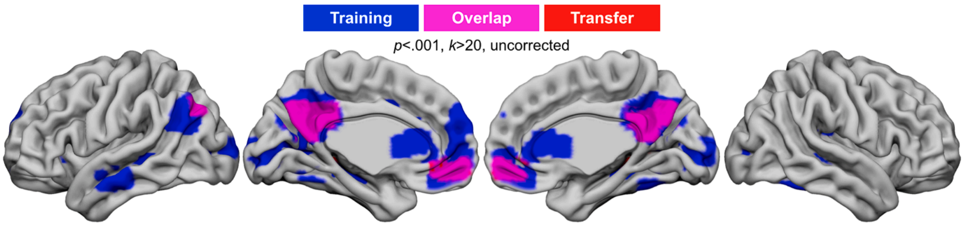


Fig. 3. Exploratory whole-brain analysis of the rtfMRI-NF training and transfer tasks.

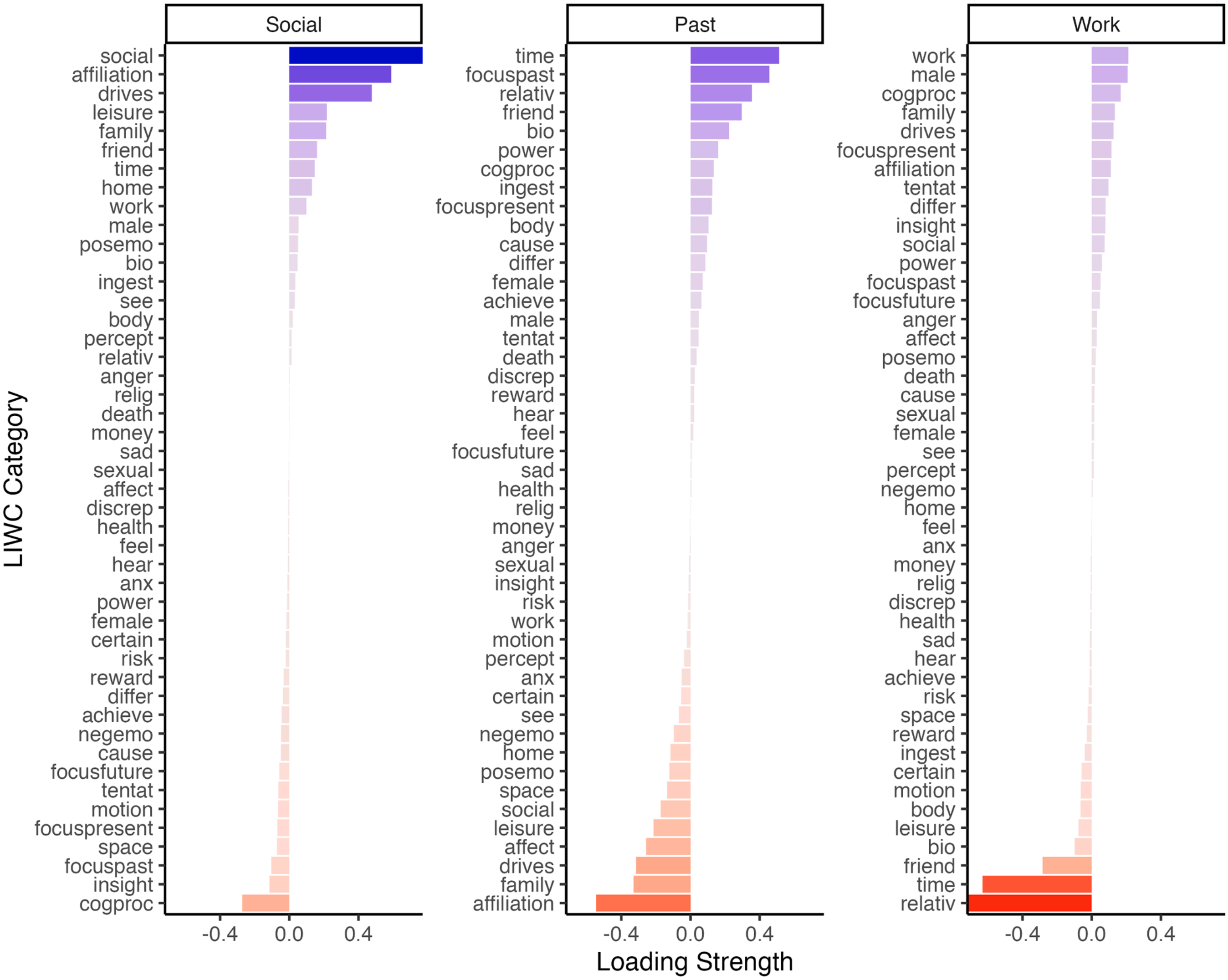


Fig. 4. Association Between Up-Regulation Strategy and Volitional Control of the LTPJ.
Note. Data from the PLSR model using LIWC features for up-regulation strategy descriptions and volitional control of LTPJ, for which we were best able to explain variance in the LIWC features and volitional control in a ROI. The first three components, which together, explain roughly 75% of the variance in LIWC features and LTPJ volitional control, are depicted along with LIWC feature loadings on each component. Blue, positive-going loadings indicate a positive association between the feature and the component; red, negative-going loadings indicate a negative association between the feature and the component. See Supplementary Materials for category abbreviations.

"else") and cognitive processes (e.g., "cause," "ought," "know") changed across session, $F(2,\ 30)=6.18$, $p=.006$, $\eta^2_g=.21$ (other component Fs$\leq$1.07, ps$\geq$.357, $\eta^2_g \leq .04$). Paired-samples t-tests revealed a large magnitude reduction in this component from session 1 to 3, $t(15)=3.53$, $p=.003$, $d_z=.88$, 95% CI [.41, 1.37], a trend decrease from session 1 to 2, $t(15)=1.95$, $p=.071$, $d_z=.49$, 95% CI [-.02, .94], and no change from session 2 to 3, $t(15)=1.18$, $p=.258$, $d_z=.29$, 95% CI [-.25, .82]. These findings provide at least tentative evidence that some change in strategy is occurring over sessions, which may be related to the neurofeedback signal since after the first session, participants only had the neurofeedback signal on which to base changes in strategy.

Another possibility is that neurofeedback-related learning is occurring, but rapidly, within the very first neurofeedback session. To evaluate this possibility, we re-analyzed the training data from just the first session by submitting beta values representing the difference between up- and down-regulation for each of the four runs of session one to a repeated-measures ANOVA. We focused on response in LTPJ and RTPJ, reasoning that since the feedback signal directly relates to activity in these regions, learning-related changes should in theory be most clearly observed in these regions. We found no difference in volitional control across the four runs in LTPJ, $F(3, 45)=1.84$, $p=.153$, $\eta^2_G=.05$, or RTPJ, $F(3, 45)=2.13$, $p=.110$, $\eta^2_G=.07$.

One final possibility is that neurofeedback-related learning is occurring, but even more rapidly, within the first training run of the first session. We conducted another analysis to evaluate this idea by submitting beta values representing the difference between up- and down-regulation for each of the three blocks of the first run of the first session to repeated-measures ANOVAs. Given that participants received no neurofeedback prior to the first block, this serves as a good test of whether volitional control is possible without yet receiving any neurofeedback. In LTPJ, volitional control numerically increased from the first to third block, $d_z=.35$, 95% CI [-.20, .76], and second to third block, $d_z=.47$, 95% CI [-.07, .95]; however, these differences were not unexpected, $F(2, 30)=1.96$, $p=.159$, $\eta^2_G=.06$. In contrast, in RTPJ, differences in volitional control across blocks was unexpected under the null hypothesis, $F(1,\ 21)=5.41$, $p=.021$, $\eta^2_g=.18$. Paired-samples t-tests revealed that volitional control was higher in the first versus second block, $t(15)=2.44$, $p=.028$, $d_z=-.61$, 95% CI [-1.11, -.11], and in the third versus second block, $t(15)=2.74$, $p=.015$, $d_z=.69$, 95% CI [.15, 1.27]; there was no difference between the first and third block, t

Table 4
Behavioral performance.

Measure	Pre, *M* (*SD*)	Post, *M* (*SD*)	Statistical Comparison	Effect Size [a]
Hinting Task	17.3 (2.1)	18.0 (1.6)	$t(15)=1.69$, $p=.055$	$d_z=.42$ [-.10, .93]
Social Attribution Task	15.9 (2.9)	16.7 (1.7)	$t(15)=1.12$, $p=.141$	$d_z=.28$ [-.28, .68]
Multiracial Emotion Identification	.88 (.04)	.88 (.04)	$t(15)=.33$, $p=.627$	$d_z=-.08$ [-.68, .45]
Spontaneous Theory of Mind Protocol	8.5 (2.6)	7.6 (3.2)	$t(15)=1.04$, $p=.843$	$d_z=-.26$ [-.76, .29]
Mental State Fluency			*Time:* $F(1,15)=.95$, $p=.346$; *Valence:* $F(1,15)=8.88$, $p=.009$; *Interaction:* $F(1,15)=.13$, $p=.720$	*Time:* $\eta^2_G=.01$; *Valence:* $\eta^2_G=.09$; *Interaction:* $\eta^2_G=.001$
Positive	611.6 (217.3)	545.7 (225.1)		
Negative	466.5 (240.6)	432.8 (190.9)		
Attentional Cueing Task			*Time:* $F(1,15)=19.55$, $p<.001$; *Trial Type:* $F(1,15)=8.03$, $p=.013$; *Interaction:* $F(1,15)=1.44$, $p=.249$	*Time:* $\eta^2_G=.05$; *Trial Type:* $\eta^2_G=.01$; *Interaction:* $\eta^2_G=.002$
Invalid	.62 (.18)	.67 (.17)		
Valid	.57 (.11)	.65 (.16)		

[a] Negative d_z indicates that performance pre-rtfMRI-NF is greater than post-rtfMRI-NF.

(15)=.88, $p=.396$, $d_z=-.22$, 95% CI [-.81, .31]. Together, these data suggest that some degree of volitional control is present without yet receiving any neurofeedback, although changes in volitional control across block suggest that some neurofeedback-related learning may be occurring.

4. Discussion

Given the importance of ToM for effective interpersonal interaction, it stands to reason that gaining the ability to self-regulate the neural network mediating ToM may carry positive consequences for real-world social behavior. Towards addressing this possibility, in the current study, our primary aim was to test whether rtfMRI-NF conferred volitional control of the ToM network. As additional aims, we evaluated the strategies used to self-regulate the network and whether volitional control of the ToM network was moderated by participant characteristics and associated with improved performance on behavioral measures. In doing so, we attempted to evaluate whether brain activity can be volitionally modulated with neurofeedback in a way that ultimately enhances social processes, as in some other work (Direito et al., 2021; Kanel et al., 2019; Moll et al., 2014; Pereira et al., 2019; Ruiz et al., 2013; Yao et al., 2016). In contrast to other work, we targeted key nodes of the ToM network, using intermittent, activation-based neurofeedback (as opposed to other neurofeedback approaches such as multivariate pattern analysis, as in Moll et al., 2014), and formally analyzed the association between self-regulation strategy and volitional control.

We found that during the training task, when participants were actively provided with activation-based intermittent neurofeedback, participants demonstrated volitional control—operationalized as the reliable difference in neural activity between up- and down-regulation—of all regions of the network except for the RSTS. These effects were consistently large in magnitude across ROIs, and did not differ across session. Our more critical test of volitional control was neural activity during the transfer task when no active neurofeedback was being provided. On this task, volitional control was achieved in all ROIs except in RTPJ and MMPFC. Effects were generally smaller in magnitude compared to the training task, although still medium-to-large. Similar to the training task, volitional control did not change across sessions, suggesting that volitional control could be achieved by the end of the first rtfMRI-NF session. Whole-brain analysis was largely consistent with these ROI analysis findings.

It is perhaps easier to explain the lack of volitional control in non-targeted regions such as the RSTS (training task) and MMPFC (transfer task) since neural activity in these regions did not contribute to the neurofeedback signal. Further, despite being implicated as part of a core ToM network (Molenberghs et al., 2016; Schurz et al., 2014; Van Overwalle, 2009) or one subserving predominantly cognitive, as opposed to affective, mental state inference processes (Schurz et al., 2021), these regions also show distinct response profiles to social and mental state information (Schurz et al., 2014), suggesting different regions are implementing different subprocesses of ToM. For example, the STS has been implicated in various aspects of social perception, including analyzing biological motion (Allison et al., 2000) and inferring an agent's intent from their actions (Pelphrey et al., 2004; Saxe et al., 2004). MPFC has been implicated in the process of making judgments regarding stable social or psychological characteristics of others (Van Overwalle, 2009) in a way that is sensitive to the self-relevance of the target (Tamir & Mitchell, 2010), and exhibits particular sensitivity to the valence of another's mental state (Skerry & Saxe, 2014, 2015). It may be that the strategies employed by participants for self-regulation did not overwhelmingly tap these processes. It

Table 5
Associations between volitional control and change in behavioral performance.

ROI	Hinting Task	Social Attribution Task	Multiracial Emotion Identification	Spontaneous ToM Protocol	Mental State Fluency-Positive	Mental State Fluency-Negative	Attentional Cueing
DMPFC	-.09 [-.60, .42]	.18 [-.50, .71]	-.30 [-.70, .23]	.37 [-.11, .75]	.14 [-.37, .61]	-.14 [-.69, .51]	-.07 [-.52, .50]
LTPJ	-.15 [-.62, .36]	-.10 [-.63, .54]	-.21 [-.62, .29]	.24 [-.37, .72]	-.10 [-.67, .51]	.14 [-.40, .67]	-.09 [-.63, .56]
MMPFC	-.11 [-.61, .39]	.16 [-.50, .67]	-.22 [-.68, .34]	.10 [-.50, .63]	.34 [-.28, .74]	.42 [-.17, .78]	.06 [-.51, .63]
PC	.07 [-.38, .50]	.09 [-.49, .61]	.04 [-.45, .49]	.40 [-.16, .80]	-.02 [-.59, .60]	.08 [-.50, .61]	-0.31 [-.74, .36]
RSTS	.30 [-.26, .69]	-.00 [-.44, .44]	-.26 [-.68, .34]	.09 [-.51, .60]	.27 [-.29, .68]	.43 [-.14, .74]	-.20 [-.69, .36]
RTPJ	**.52 [-.02, .80]**	-.37 [-.75, .25]	-.03 [-.55, .49]	.37 [-.16, .70]	.20 [-.32, .66]	.11 [-.49, .62]	-.01 [-.45, .51]
VMPFC	-.22 [-.69, .34]	-.05 [-.59, .58]	-.42 [-.79, .21]	.07 [-.47, .54]	.41 [-.07, .77]	.30 [-.21, .71]	.12 [-.42, .67]

Note. Values represent Spearman's rho and 95% BCa CI generated from 10,000 bootstrap samples. Bolded values indicate $p<.05$.

is harder to explain the lack of a transfer effect in RTPJ since it was one of the neurofeedback targets. That said, along with MMPFC, the effect of up- vs down-regulation was in the predicted direction and small-to-medium in magnitude, suggesting that we simply may have been underpowered to detect the transfer effect in these ROIs. It may also be that strategies used by the participants may be more effective at self-regulating certain brain regions versus others, as suggested by our text-based analysis of strategy in which we were able to explain more variance in volitional control of LTPJ vs RTPJ. Here too, although research implicates both TPJ regions as part of a core network (Molenberghs et al., 2016; Schurz et al., 2014, 2021), research also suggests subtle differences in the functional profile of LTPJ and RTPJ, with RTPJ showing the most selective profile for mental state information (Aichhorn et al., 2009; Perner et al., 2006; Saxe & Wexler, 2005) and LTPJ showing sensitivity to discrepant viewpoints (as in false-sign vignettes; Aichhorn et al., 2009; Perner et al., 2006) in additional to mental state information, suggesting a more general role for LTPJ in metarepresentational reasoning (Apperly et al., 2007). Unfortunately, the coarseness of the LIWC data do not allow for a more granular mapping between strategy and neural activity, which could provide useful information regarding the processes these regions implement in the context of ToM.

On the impact of session, although it seems reasonable to assume that more training would lead to more volitional control, participants may be able to rapidly learn self-regulation strategies in a way that additional neurofeedback may not add a measurable benefit. Indeed, other rtfMRI-NF studies have shown a positive neural effect after just a single session in samples with and without mental disorders (Bauer et al., 2020; De Filippi et al., 2022; MacDuffie et al., 2018; Okano et al., 2020). It is also possible that there exists a relatively easily achievable ceiling to our neural measure of volitional control. In other words, thinking about social content during up-regulation and non-social content during down-regulation leads to reliable volitional control, and subtle changes in the social and non-social content participants considered during self-regulation produces little in the way of measurable neural change, at least as assessed with LIWC. Thus, part of what participants may be learning with the neurofeedback signal is which subtle changes in strategy lead to this ceiling effect in neural activity, which would make it difficult for us to detect changes in volitional control and map strategies to volitional control.

Although we were not able to explain volitional control as a function of several participant characteristics (i.e., vividness of imagery, trait empathy), we were able to identify in-session strategies that were associated with the extent of volitional control. On up-regulation strategies, volitional control in DMPFC, MMPFC, RSTS, and RTPJ was best explained by a single component defined by social, affiliation, and drive-related features. Said otherwise, volitional control was highest when participants thought about social experiences, behaviors, relationships, and motivations. That said, we were able to account for less than a quarter of the variance in volitional control in these regions. In contrast, we were able to explain close to 100% of the variance in volitional control in LTPJ with eight components, the first of which was similar to the social, affiliation, and drive-related features that emerged from the analysis of other regions. The other two components, which, along with the first, explained nearly 75% of the variance in volitional control, were defined by a focus on past experiences, and work, male references, and family, respectively. Together, these data indicate that, similar to the other regions, thinking about prior social experiences, behaviors, events, and relationships drive volitional control on LTPJ. These findings largely align with what is known about the functional properties of these regions. On down-regulation strategies, we were able to explain less than 50% of the variance in volitional control in just RSTS and VMPFC. The features loading most highly onto the single predictor component involved perceptual processes. More qualitatively, participants described attending to things they saw in the scanner or visualizing inanimate objects, neither of which were well characterized by LIWC. Analysis of the training data demonstrated that we were able to explain more strategy-related variance in volitional control, although the strategies were qualitatively similar, involving social experiences for up-regulation and perceptual processes for down-regulation. Further analysis of these data with more sensitive approaches that are not limited to pre-existing dictionaries, for example, with topic modeling (Berger & Packard, 2022), may yield additional important insights into effective self-regulation strategies.

A key question we aimed to address is whether volitional control impacted behavior in measurable ways. Towards evaluating this question, participants completed several measures that are associated with the TPJ and the ToM network more broadly, including tests of explicit and implicit mental state attribution and attentional reorienting. Performance was unchanged after rtfMRI-NF, although we did find that volitional control in RTPJ was associated with positive changes in Hinting Task performance, suggesting a tangible benefit of volitional control on processing intention from indirect speech. However, towards reducing Type II error rates, we conducted a large number of tests on a small sample, meaning that this association and its magnitude should be interpreted with caution. It is possible that neural self-regulation does translate to behavioral self-regulation, but we were unable to detect it here either because we were underpowered, and behavioral effects may be small, or because our measures were psychometrically not well-suited to detecting within-person change.

One of the most critical questions concerns the necessity or utility of the neurofeedback signal for volitional control. Due to the design of the study, which did not include a control group nor a pre-rtfMRI-NF transfer scan, we cannot rule out the possibility that the initial strategy we provided to participants was sufficient for achieving volitional control. We attempted to address this issue in a series of exploratory analyses. First, we found that participant-reported self-regulation strategy for down-regulation changed across sessions with the change characterized, in part, by an appropriate reduction in the extent to which participants used words indexing cognitive processes including belief states. Given that participants were not provided with any additional instruction on strategy and have only the neurofeedback signal to base strategy changes on, this could be taken as tentative evidence of neurofeedback-related learning. Second, we evaluated the possibility that neurofeedback-related changes in volitional control are occurring, but much more rapidly than would be detected in our main analysis that evaluated changes across session. In line with this idea, we observed changes in volitional control of RTPJ across blocks within the first run of the first session. This change was characterized by higher volitional control in blocks one and three versus block two, with no difference between the first and third block. This could suggest that participants start with a good strategy, adjust that strategy for the second block, see that it does not work as well as indicated by the neurofeedback signal, and either return to the first strategy or use a new strategy for the third block. Since we do not have data on by-block strategies, we cannot adjudicate between these possibilities. Nevertheless, these data offer two important insights. First, some degree of volitional control is possible within the very first block of the task without yet receiving neurofeedback, suggesting that the initial strategy offered is sufficient for volitional control. Second, because volitional control did change across blocks, this too suggests that the neurofeedback signal is being used, although to varying degrees of success. Indeed, there was substantial variability between participants in terms of volitional control over the course of sessions, runs, and blocks, suggesting that there might be participant characteristics that impact the success with which one is able to use the neurofeedback signal to alter self-regulation strategies. This would be an important area for future investigation.

This study is limited by several critical limitations. As discussed above, the primary limitation was that there was no control group, leaving open the possibility that volitional control could be achieved without the neurofeedback signal or that volitional control was somehow mediated by non-specific effects like physiological differences during up- versus down-regulation, among other possibilities. In other words, we cannot conclusively rule-in or rule-out the necessity of the neurofeedback signal for volitional control beyond the initial self-regulation strategy provided to participants. Control conditions for rtfMRI studies and the limitations posed by the lack of appropriate controls (included those aforementioned) have been incisively discussed by others, which we refer the reader to (see Sorger et al., 2019). Second, our sample size was extremely modest, leaving open the possibility that volitional control—in certain ROIs like RTPJ or aspects of behavioral performance—were too small for us to detect and that estimates of brain-behavior/strategy associations overestimate the true effect. Additionally, given the novelty of rtfMRI's application to training volitional control of the ToM network here, in an effort to avoid Type II error, we conducted a large number of tests without correcting for multiple comparisons. Findings should be interpreted accordingly, particularly the one brain-behavior association between RTPJ and change in Hinting Task performance, which provides only weak evidence of an association. Similarly, our statistical models evaluating the association between strategy and volitional control are prone to overfitting, making the percent variance explained values likely overestimates.

Despite these limitations, given the preliminary evidence of volitional control here, it would be worthwhile evaluating the clinical utility of rtfMRI-NF from the ToM network on the large number of clinical groups characterized by social cognitive impairment (Cotter et al., 2018). For individuals with schizophrenia spectrum disorders—a mental disorder associated with marked, pervasive, and persistent social cognitive impairments (Green et al., 2015)—gold standard pharmacological treatment often does not substantially improve aspects of social cognition, including ToM (Kucharska-Pietura & Mortimer, 2013; Penn et al., 2009; Sergi et al., 2007). Although social cognitive interventions produce moderate improvements in social cognition for these individuals, the benefits are less often observed with more naturalistic social cognitive measures (e.g., on The Awareness of Social Inference Test in Fiszdon et al., 2017 or on the Empathic Accuracy Task in Horan et al., 2018) or untrained aspects of real-world aspects of social behavior (Yeo et al., 2022). It seems reasonable that the degree and efficiency with which ToM (and the social behaviors it supports) can be trained, would be greater using methods that more directly target the neural network mediating those processes. Assuming the efficacy of rtfMRI-NF can be confirmed in future controlled work, and continues to be well tolerated, rtfMRI-NF might prove to provide at least an alternative to other interventions. As with any intervention, future work in this area should also investigate for whom rtfMRI-NF works, the durability of neural and behavior change, and ways of promoting generalization to daily social behavior.

In summary, here, in this proof-of-concept study, we find tentative support for the idea that volitional control of the ToM network can be achieved, which is largely driven by using strategies involving thinking about social experiences. Exploratory analysis suggests that neurofeedback-related learning occurred, although some degree of volitional control was achieved with the simple mental strategy initially provided to participants. Although we largely did not find clear support for the idea that neural self-regulation translates to changes in behavior, higher-powered studies with measures well suited to repeated testing will be better equipped to address this question. Ultimately, many critical questions remain, including the superiority of active rtfMRI-NF to sham rtfMRI-NF, the impact of participants characteristics on volitional control (e.g., expectation and motivation; Thibault et al., 2018), and the impact of study parameters on volitional control (e.g., feedback from one versus multiple, other ToM ROIs), among others.

Funding

Funding for this study was generously provided in part by a NARSAD Young Investigator Grant from the Brain & Behavior Research Foundation [grant #27282] and startup funds provided by the University of Rochester.

CRediT authorship contribution statement

Abhishek Saxena: Formal analysis, Investigation, Data curation, Writing – original draft. **Bridget J. Shovestul:** Investigation, Data curation, Writing – review & editing. **Emily M. Dudek:** Conceptualization, Project administration, Investigation, Data curation, Writing – review & editing. **Stephanie Reda:** Investigation, Data curation, Writing – review & editing. **Arun Venkataraman:** Methodology, Software, Writing – review & editing. **J. Steven Lamberti:** Resources, Supervision, Writing – review & editing. **David Dodell-Feder:** Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

none

Data availability

Data and analysis code described in the current manuscript are available on the Open Science Framework (https://osf.io/jbnpt/?view_only=2582ccd3cda644eb9f267928c8ca4688).

Acknowledgments

The authors would like to thank Paul Wighton and André van der Kouwe for their generous assistance in implementing *multivariate and univariate real-time functional imaging* (MURFI) software, the staff at the University of Rochester Center for Advanced Brain Imaging & Neurophysiology for assistance in setting up the real-time fMRI pipeline and data collection, Xiaoyu (Jojo) Dong for assistance with data collection, the University of Rochester Center Clinical & Translational Science Institute for assistance with recruitment, and ResearchMatch, a national health volunteer registry that was created by several academic institutions and supported by the U.S. National Institutes of Health as part of the Clinical Translational Science Award (CTSA) program. Finally, we would like to thank the participants for their participation.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2023.120334.

References

Aichhorn, M., Perner, J., Weiss, B., Kronbichler, M., Staffen, W., Ladurner, G., 2009. Temporo-parietal Junction activity in theory-of-mind tasks: falseness, beliefs, or attention. J. Cogn. Neurosci. 21 (6), 1179–1192. https://doi.org/10.1162/jocn.2009.21082.

Allison, T., Puce, A., McCarthy, G., 2000. Social perception from visual cues: role of the STS region. Trends Cogn. Sci. 4 (7), 267–278. https://doi.org/10.1016/S1364-6613(00)01501-1.

Apperly, I.A., Samson, D., Chiavarino, C., Humphreys, G.W., 2004. Frontal and temporo-parietal lobe contributions to theory of mind: neuropsychological evidence from a false-belief task with reduced language and executive demands. J. Cogn. Neurosci. 16 (10), 1773–1784. https://doi.org/10.1162/0898929042947928.

Apperly, I.A., Samson, D., Chiavarino, C., Bickerton, W.L., Humphreys, G.W., 2007. Testing the domain-specificity of a theory of mind deficit in brain-injured patients: evidence for consistent performance on non-verbal, "reality-unknown" false belief and false photograph tasks. Cognition 103 (2), 300–321. https://doi.org/10.1016/j.cognition.2006.04.012.

Bardi, L., Six, P., Brass, M., 2017. Repetitive TMS of the temporo-parietal junction disrupts participant's expectations in a spontaneous theory of mind task. Soc. Cognit. Affect. Neurosci. 12 (11), 1775–1782. https://doi.org/10.1093/scan/nsx109.

Bauer, C.C.C., Okano, K., Ghosh, S.S., Lee, Y.J., Melero, H., Angeles, de los, C., Nestor, P. G., del Re, E.C., Northoff, G., Niznikiewicz, M.A., Whitfield-Gabrieli, S., 2020. Real-time fMRI neurofeedback reduces auditory hallucinations and modulates resting state connectivity of involved brain regions: Part 2: default mode network -preliminary evidence. Psychiatry Res. 284, 112770 https://doi.org/10.1016/j.psychres.2020.112770.

Bell, M.D., Fiszdon, J.M., Greig, T.C., Wexler, B.E., 2010. Social attribution test — multiple choice (SAT-MC) in schizophrenia: comparison with community sample and relationship to neurocognitive, social cognitive and symptom measures. Schizophr. Res. 122 (1–3), 164–171. https://doi.org/10.1016/j.schres.2010.03.024.

Berger, J., Packard, G., 2022. Using natural language processing to understand people and culture. Am. Psychol. 77 (4), 525–537. https://doi.org/10.1037/amp0000882.

Birch, S.A.J., Bloom, P., 2007. The curse of knowledge in reasoning about false beliefs. Psychol. Sci. 18 (5), 382–386. https://doi.org/10.1111/j.1467-9280.2007.01909.x.

Blatt, B., LeLacheur, S.F., Galinsky, A.D., Simmens, S.J., Greenberg, L., 2010. Does perspective-taking increase patient satisfaction in medical encounters? Acad. Med. 85 (9), 1445–1452. https://doi.org/10.1097/ACM.0b013e3181eae5ec.

Bowman, L.C., Dodell-Feder, D., Saxe, R., Sabbagh, M.A., 2019. Continuity in the neural system supporting children's theory of mind development: longitudinal links between task-independent EEG and task-dependent fMRI. Dev. Cogn. Neurosci. 40, 100705 https://doi.org/10.1016/j.dcn.2019.100705.

Brainard, D.H., 1997. The psychophysics toolbox. Spat. Vis. 10 (4), 433–436.

Cahill, V.A., Malouff, J.M., Little, C.W., Schutte, N.S., 2020. Trait perspective taking and romantic relationship satisfaction: a meta-analysis. J. Fam. Psychol. 34 (8), 1025–1035. https://doi.org/10.1037/fam0000661.

Caputi, M., Lecce, S., Pagnin, A., Banerjee, R., 2012. Longitudinal effects of theory of mind on later peer relations: The role of prosocial behavior. Dev. Psychol. 48 (1), 257–270. https://doi.org/10.1037/a0025402.

Corbetta, M., Shulman, G.L., 2002. Control of goal-directed and stimulus-driven attention in the brain. Nat. Rev. Neurosci. 3 (3), 201–215. https://doi.org/10.1038/nrn755.

Corbetta, M., Kincade, J.M., Ollinger, J.M., McAvoy, M.P., Shulman, G.L., 2000. Voluntary orienting is dissociated from target detection in human posterior parietal cortex. Nat. Neurosci. 3 (3), 292–297. https://doi.org/10.1038/73009.

Corcoran, R., Mercer, G., Frith, C.D., 1995. Schizophrenia, symptomatology and social inference: investigating "theory of mind" in people with schizophrenia. Schizophr. Res. 17 (1), 5–13. https://doi.org/10.1016/0920-9964(95)00024-G.

Cotter, J., Granger, K., Backx, R., Hobbs, M., Looi, C.Y., Barnett, J.H., 2018. Social cognitive dysfunction as a clinical marker: a systematic review of meta-analyses across 30 clinical conditions. Neurosci. Biobehav. Rev. 84, 92–99. https://doi.org/10.1016/j.neubiorev.2017.11.014.

Damen, D., Pollmann, M.M.H., Grassow, T.L., 2021. The benefits and obstacles to perspective getting. Front. Commun. 6, 611187 https://doi.org/10.3389/fcomm.2021.611187.

Davis, M.H., 1983. Measuring individual differences in empathy: Evidence for a multidimensional approach. J. Pers. Soc. Psychol. 44 (1), 113–126. https://doi.org/10.1037/0022-3514.44.1.113.

De Coster, L., Lin, L., Mathalon, D.H., Woolley, J.D., 2019. Neural and behavioral effects of oxytocin administration during theory of mind in schizophrenia and controls: a randomized control trial. Neuropsychopharmacology 44 (11), 1925–1931. https://doi.org/10.1038/s41386-019-0417-5.

De Filippi, E., Marins, T., Escrichs, A., Gilson, M., Moll, J., Tovar-Moll, F., Deco, G., 2022. One session of fMRI-Neurofeedback training on motor imagery modulates whole-brain effective connectivity and dynamical complexity. Cereb. Cortex Commun. 3 (3), tgac027. https://doi.org/10.1093/texcom/tgac027.

deCharms, R.C., 2007. Reading and controlling human brain activation using real-time functional magnetic resonance imaging. Trends Cogn. Sci. 11 (11), 473–481. https://doi.org/10.1016/j.tics.2007.08.014.

deCharms, R.C., 2008. Applications of real-time fMRI. Nat. Rev. Neurosci. 9 (9), 720–729. https://doi.org/10.1038/nrn2414.

Direito, B., Mouga, S., Sayal, A., Simões, M., Quental, H., Bernardino, I., Playle, R., McNamara, R., Linden, D.E., Oliveira, G., Castelo Branco, M., 2021. Training the social brain: clinical and neural effects of an 8-week real-time functional magnetic resonance imaging neurofeedback phase IIa clinical trial in autism. Autism 25 (6), 1746–1760. https://doi.org/10.1177/13623613211002052.

Dodell-Feder, D., Koster-Hale, J., Bedny, M., Saxe, R., 2011. FMRI item analysis in a theory of mind task. Neuroimage 55 (2), 705–712. https://doi.org/10.1016/j.neuroimage.2010.12.040.

Dodell-Feder, D., DeLisi, L.E., Hooker, C.I., 2014a. Neural disruption to theory of mind predicts daily social functioning in individuals at familial high-risk for schizophrenia. Soc. Cognit. Affect. Neurosci. 9 (12), 1914–1925. https://doi.org/10.1093/scan/nst186.

Dodell-Feder, D., Tully, L.M., Lincoln, S.H., Hooker, C.I., 2014b. The neural basis of theory of mind and its relationship to social functioning and social anhedonia in individuals with schizophrenia. NeuroImage Clin. 4, 154–163. https://doi.org/10.1016/j.nicl.2013.11.006.

Dodell-Feder, D., Felix, S., Yung, M.G., Hooker, C.I., 2016. Theory-of-mind-related neural activity for one's romantic partner predicts partner well-being. Soc. Cognit. Affect. Neurosci. 11 (4), 593–603. https://doi.org/10.1093/scan/nsv144.

Dodell-Feder, D., Ressler, K.J., Germine, L.T., 2020. Social cognition or social class and culture? On the interpretation of differences in social cognitive performance. Psychol. Med. 50 (1), 133–145. https://doi.org/10.1017/S003329171800404X.

Dodell-Feder, D., Tully, L.M., Dudek, E., Hooker, C.I., 2021. The representation of mental state information in schizophrenia and first-degree relatives: a multivariate pattern analysis of fMRI data. Soc. Cognit. Affect. Neurosci. 16 (6), 608–620. https://doi.org/10.1093/scan/nsab028.

Dudek, E., Dodell-Feder, D., 2021. The efficacy of real-time functional magnetic resonance imaging neurofeedback for psychiatric illness: a meta-analysis of brain and behavioral outcomes. Neurosci. Biobehav. Rev. 121, 291–306. https://doi.org/10.1016/j.neubiorev.2020.12.020.

Dufour, N., Redcay, E., Young, L., Mavros, P.L., Moran, J.M., Triantafyllou, C., Gabrieli, J.D.E., Saxe, R., 2013. Similar brain activation during false belief tasks in a large sample of adults with and without autism. PLoS One 8 (9), e75468. https://doi.org/10.1371/journal.pone.0075468.

Eyal, T., Steffel, M., Epley, N., 2018. Perspective mistaking: accurately understanding the mind of another requires getting perspective, not taking perspective. J. Pers. Soc. Psychol. 114 (4), 547–571. https://doi.org/10.1037/pspa0000115.

Fede, S.J., Dean, S.F., Manuweera, T., Momenan, R., 2020. A guide to literature informed decisions in the design of real time fMRI neurofeedback studies: a systematic review. Front. Hum. Neurosci. 14, 60. https://doi.org/10.3389/fnhum.2020.00060.

Fett, A.K.J., Viechtbauer, W., Dominguez, M.G., Penn, D.L., van Os, J., Krabbendam, L., 2011. The relationship between neurocognition and social cognition with functional outcomes in schizophrenia: a meta-analysis. Neurosci. Biobehav. Rev. 35 (3), 573–588. https://doi.org/10.1016/j.neubiorev.2010.07.001.

Fink, E., Begeer, S., Peterson, C.C., Slaughter, V., de Rosnay, M., 2015. Friendlessness and theory of mind: a prospective longitudinal study. Br. J. Dev. Psychol. 33 (1), 1–17. https://doi.org/10.1111/bjdp.12060.

Finkel, E.J., Simpson, J.A., Eastwick, P.W., 2017. The psychology of close relationships: fourteen core principles. Annu. Rev. Psychol. 68 (1), 383–411. https://doi.org/10.1146/annurev-psych-010416-044038.

Finn, E.S., Corlett, P.R., Chen, G., Bandettini, P.A., Constable, R.T., 2018. Trait paranoia shapes inter-subject synchrony in brain activity during an ambiguous social narrative. Nat. Commun. 9 (1), 2043. https://doi.org/10.1038/s41467-018-04387-2.

First, M.B., Williams, J.B.W., Karg, R.S., Spitzer, R.L., 2015. Structured Clinical Interview For DSM-5—Research Version (SCID-5 For DSM-5, Research Version; SCID-5-RV). American Psychiatric Association.

Fiszdon, J.M., Roberts, D.L., Penn, D.L., Choi, K.H., Tek, C., Choi, J., Bell, M.D., 2017. Understanding social situations (USS): a proof-of-concept social–cognitive intervention targeting theory of mind and attributional bias in individuals with psychosis. Psychiatr. Rehabil. J. 40 (1), 12–20. https://doi.org/10.1037/prj0000190.

Galinsky, A.D., Maddux, W.W., Gilin, D., White, J.B., 2008. Why It pays to get inside the head of your opponent: the differential effects of perspective taking and empathy in negotiations. Psychol. Sci. 19 (4), 378–384. https://doi.org/10.1111/j.1467-9280.2008.02096.x.

Goldstein, N.J., Vezich, I.S., Shapiro, J.R., 2014. Perceived perspective taking: when others walk in our shoes. J. Pers. Soc. Psychol. 106 (6), 941–960. https://doi.org/10.1037/a0036395.

Grady, C.L., Rieck, J.R., Nichol, D., Rodrigue, K.M., Kennedy, K.M., 2021. Influence of sample size and analytic approach on stability and interpretation of brain-behavior correlations in task-related fMRI data. Hum. Brain Mapp. 42 (1), 204–219. https://doi.org/10.1002/hbm.25217.

Green, M.F., Horan, W.P., Lee, J., 2015. Social cognition in schizophrenia. Nat. Rev. Neurosci. 16 (10), 620–631. https://doi.org/10.1038/nrn4005.

Gweon, H., Dodell-Feder, D., Bedny, M., Saxe, R., 2012. Theory of mind performance in children correlates with functional specialization of a brain region for thinking about thoughts: behavioral and neural development in theory of mind. Child Dev. 83 (6), 1853–1868. https://doi.org/10.1111/j.1467-8624.2012.01829.x.

Hawkley, L.C., Cacioppo, J.T., 2010. Loneliness matters: a theoretical and empirical review of consequences and mechanisms. Ann. Behav. Med. 40 (2), 218–227. https://doi.org/10.1007/s12160-010-9210-8.

Hawkley, L.C., 2022. Loneliness and health. Nat. Rev. Dis. Primers 8 (1), 22. https://doi.org/10.1038/s41572-022-00355-9.

Hildebrandt, M.K., Jauk, E., Lehmann, K., Maliske, L., Kanske, P., 2021. Brain activation during social cognition predicts everyday perspective-taking: a combined fMRI and ecological momentary assessment study of the social brain. Neuroimage 227, 117624. https://doi.org/10.1016/j.neuroimage.2020.117624.

Hinds, O., Ghosh, S., Thompson, T.W., Yoo, J.J., Whitfield-Gabrieli, S., Triantafyllou, C., Gabrieli, J.D.E., 2011. Computing moment-to-moment BOLD activation for real-time neurofeedback. Neuroimage 54 (1), 361–368. https://doi.org/10.1016/j.neuroimage.2010.07.060.

Holt-Lunstad, J., Smith, T.B., Layton, J.B., 2010. Social relationships and mortality risk: a meta-analytic review. PLoS Med. 7 (7), e1000316 https://doi.org/10.1371/journal.pmed.1000316.

Holt-Lunstad, J., Smith, T.B., Baker, M., Harris, T., Stephenson, D., 2015. Loneliness and social isolation as risk factors for mortality: a meta-analytic review. Perspect. Psychol. Sci. 10 (2), 227–237. https://doi.org/10.1177/1745691614568352.

Holt-Lunstad, J., Robles, T.F., Sbarra, D.A., 2017. Advancing social connection as a public health priority in the United States. Am. Psychol. 72 (6), 517–530. https://doi.org/10.1037/amp0000103.

Horan, W.P., Dolinsky, M., Lee, J., Kern, R.S., Hellemann, G., Sugar, C.A., Glynn, S.M., Green, M.F., 2018. Social cognitive skills training for psychosis with community-based training exercises: a randomized controlled trial. Schizophr. Bull. 44 (6), 1254–1266. https://doi.org/10.1093/schbul/sbx167.

Imuta, K., Henry, J.D., Slaughter, V., Selcuk, B., Ruffman, T., 2016. Theory of mind and prosocial behavior in childhood: a meta-analytic review. Dev. Psychol. 52 (8), 1192–1205. https://doi.org/10.1037/dev0000140.

Kana, R.K., Keller, T.A., Cherkassky, V.L., Minshew, N.J., Just, M.A., 2009. Atypical frontal-posterior synchronization of theory of mind regions in autism during mental state attribution. Soc. Neurosci. 4 (2), 135–152. https://doi.org/10.1080/17470910802198510.

Kanel, D., Al-Wasity, S., Stefanov, K., Pollick, F.E., 2019. Empathy to emotional voices and the use of real-time fMRI to enhance activation of the anterior insula. Neuroimage 198, 53–62. https://doi.org/10.1016/j.neuroimage.2019.05.021.

Kanske, P., Böckler, A., Trautwein, F.M., Singer, T., 2015. Dissecting the social brain: introducing the EmpaToM to reveal distinct neural networks and brain–behavior relations for empathy and theory of mind. Neuroimage 122, 6–19. https://doi.org/10.1016/j.neuroimage.2015.07.082.

Kanske, P., Böckler, A., Trautwein, F.M., Parianen Lesemann, F.H., Singer, T., 2016. Are strong empathizers better mentalizers? evidence for independence and interaction between the routes of social cognition. Soc. Cognit. Affect. Neurosci. 11 (9), 1383–1392. https://doi.org/10.1093/scan/nsw052.

Kassambara A. (2020). ggpubr: "ggplot2" based publication ready plots (0.4.0). https://CRAN.R-project.org/package=ggpubr.

Kassambara A. (2021). rstatix: pipe-friendly framework for basic statistical tests (0.7.0). https://CRAN.R-project.org/package=rstatix.

Keysar, B., Lin, S., Barr, D.J., 2003. Limits on theory of mind use in adults. Cognition 89 (1), 25–41. https://doi.org/10.1016/S0010-0277(03)00064-7.

Kirby, K.N., Gerlanc, D., 2013. *BootES*: An R package for bootstrap confidence intervals on effect sizes. Behav. Res. 45, 905–927. https://doi.org/10.3758/s13428-013-0330-5.

Klein, H.S., Springfield, C.R., Bass, E., Ludwig, K., Penn, D.L., Harvey, P.D., Pinkham, A. E., 2020. Measuring mentalizing: a comparison of scoring methods for the hinting task. Int. J. Methods Psychiatr. Res. 29 (2) https://doi.org/10.1002/mpr.1827.

Kleiner, M., Brainard, D., Pelli, D., 2007. What's new in psychtoolbox-3? Perception 36 (14), 1–16.

Kohler, C.G., Turner, T.H., Bilker, W.B., Brensinger, C.M., Siegel, S.J., Kanes, S.J., Gur, R. E., Gur, R.C., 2003. Facial emotion recognition in schizophrenia: intensity effects and error pattern. Am. J. Psychiatry 160 (10), 1768–1774. https://doi.org/10.1176/appi.ajp.160.10.1768.

Koush, Y., Zvyagintsev, M., Dyck, M., Mathiak, K.A., Mathiak, K., 2012. Signal quality and Bayesian signal processing in neurofeedback based on real-time fMRI. Neuroimage 59 (1), 478–489. https://doi.org/10.1016/j.neuroimage.2011.07.076.

Koush, Y., Ashburner, J., Prilepin, E., Sladky, R., Zeidman, P., Bibikov, S., Scharnowski, F., Nikonorov, A., De Ville, D.V., 2017. OpenNFT: an open-source Python/Matlab framework for real-time fMRI neurofeedback training based on activity, connectivity and multivariate pattern analysis. Neuroimage 156, 489–503. https://doi.org/10.1016/j.neuroimage.2017.06.039.

Krall, S.C., Volz, L.J., Oberwelland, E., Grefkes, C., Fink, G.R., Konrad, K., 2016. The right temporoparietal junction in attention and social interaction: a transcranial magnetic stimulation study: RTPJ-TMS in attention and social interaction. Hum. Brain Mapp. 37 (2), 796–807. https://doi.org/10.1002/hbm.23068.

Kucharska-Pietura, K., Mortimer, A., 2013. Can antipsychotics improve social cognition in patients with schizophrenia? CNS Drugs 27 (5), 335–343. https://doi.org/10.1007/s40263-013-0047-0.

Kurtz, M.M., Richardson, C.L., 2012. Social cognitive training for schizophrenia: a meta-analytic investigation of controlled research. Schizophr. Bull. 38 (5), 1092–1104. https://doi.org/10.1093/schbul/sbr036.

Lakens, D., 2017. Will knowledge about more efficient study designs increase the willingness to pre-register? MetaArXiv. https://doi.org/10.31222/osf.io/svzyc.

Lamm, C., Decety, J., Singer, T., 2011. Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. Neuroimage 54 (3), 2492–2502. https://doi.org/10.1016/j.neuroimage.2010.10.014.

Lecce, S., Ceccato, I., Bianco, F., Rosi, A., Bottiroli, S., Cavallini, E., 2017. Theory of mind and social relationships in older adults: the role of social motivation. Aging Ment. Health 21 (3), 253–258. https://doi.org/10.1080/13607863.2015.1114586.

Lezak, M.D., 2012. Neuropsychological Assessment, 5th ed. Oxford University Press.

Liland K.H., Mevik B.H., & Wehrens R. (2021). pls: partial least squares and principal component regression (2.8-0). https://CRAN.R-project.org/package=pls.

MacDuffie, K.E., MacInnes, J., Dickerson, K.C., Eddington, K.M., Strauman, T.J., Adcock, R.A., 2018. Single session real-time fMRI neurofeedback has a lasting impact on cognitive behavioral therapy strategies. NeuroImage Clin. 19, 868–875. https://doi.org/10.1016/j.nicl.2018.06.009.

MacLeod, A.K., Rose, G.S., Williams, J.M.G., 1993. Components of hopelessness about the future in parasuicide. Cogn. Ther. Res. 17 (5), 441–455. https://doi.org/10.1007/BF01173056.

Mai, X., Zhang, W., Hu, X., Zhen, Z., Xu, Z., Zhang, J., Liu, C., 2016. Using tDCS to explore the role of the right temporo-parietal junction in theory of mind and cognitive empathy. Front. Psychol. 7 https://doi.org/10.3389/fpsyg.2016.00380.

Mar, R.A., 2011. The neural bases of social cognition and story comprehension. Annu. Rev. Psychol. 62 (1), 103–134. https://doi.org/10.1146/annurev-psych-120709-145406.

Marek, S., Tervo-Clemmens, B., Calabro, F.J., Montez, D.F., Kay, B.P., Hatoum, A.S., Donohue, M.R., Foran, W., Miller, R.L., Hendrickson, T.J., Malone, S.M., Kandala, S., Feczko, E., Miranda-Dominguez, O., Graham, A.M., Earl, E.A., Perrone, A.J., Cordova, M., Doyle, O., Dosenbach, N.U.F, 2022. Reproducible brain-wide association studies require thousands of individuals. Nature 603 (7902), 654–660. https://doi.org/10.1038/s41586-022-04492-9.

Marks, D.F., 1973. Visual imagery differences in the recall of pictures. Br. J. Psychol. 64 (1), 17–24. https://doi.org/10.1111/j.2044-8295.1973.tb01322.x.

Martz, M.E., Hart, T., Heitzeg, M.M., Peltier, S.J., 2020. Neuromodulation of brain activation associated with addiction: A review of real-time fMRI neurofeedback studies. NeuroImage Clin. 27, 102350 https://doi.org/10.1016/j.nicl.2020.102350.

Masten, C.L., Morelli, S.A., Eisenberger, N.I., 2011. An fMRI investigation of empathy for 'social pain' and subsequent prosocial behavior. Neuroimage 55 (1), 381–388. https://doi.org/10.1016/j.neuroimage.2010.11.060.

Mayer M. (2022). confintr: confidence intervals (R package version 0.1.2). https://CRAN.R-project.org/package=confintr.

Mevik, B.H., Wehrens, R., 2007. The **pls** package: principal component and partial least squares regression in R. J. Stat. Softw. 18 (2) https://doi.org/10.18637/jss.v018.i02.

Mitchell, J.P., 2008. Activity in right temporo-parietal junction is not selective for theory-of-mind. Cereb. Cortex 18 (2), 262–271. https://doi.org/10.1093/cercor/bhm051.

Molenberghs, P., Johnson, H., Henry, J.D., Mattingley, J.B., 2016. Understanding the minds of others: a neuroimaging meta-analysis. Neurosci. Biobehav. Rev. 65, 276–291. https://doi.org/10.1016/j.neubiorev.2016.03.020.

Moll, J., Weingartner, J.H., Bado, P., Basilio, R., Sato, J.R., Melo, B.R., Bramati, I.E., De Oliveira-Souza, R., Zahn, R., 2014. Voluntary enhancement of neural signatures of affiliative emotion using fMRI neurofeedback. PLoS One 9 (5), e97343. https://doi.org/10.1371/journal.pone.0097343.

Morelli, S.A., Rameson, L.T., Lieberman, M.D., 2014. The neural components of empathy: predicting daily prosocial behavior. Soc. Cognit. Affect. Neurosci. 9 (1), 39–47. https://doi.org/10.1093/scan/nss088.

Mukerji, C.E., Lincoln, S.H., Dodell-Feder, D., Nelson, C.A., Hooker, C.I., 2019. Neural correlates of theory-of-mind are associated with variation in children's everyday social cognition. Soc. Cognit. Affect. Neurosci. 14 (6), 579–589. https://doi.org/10.1093/scan/nsz040.

Nijman, S.A., Veling, W., van der Stouwe, E.C.D., Pijnenborg, G.H.M., 2020. Social cognition training for people with a psychotic disorder: a network meta-analysis. Schizophr. Bull. 46 (5), 1086–1103. https://doi.org/10.1093/schbul/sbaa023.

Okano, K., Bauer, C.C.C., Ghosh, S.S., Lee, Y.J., Melero, H., de los Angeles, C., Nestor, P. G., del Re, E.C., Northoff, G., Whitfield-Gabrieli, S., Niznikiewicz, M.A., 2020. Real-time fMRI feedback impacts brain activation, results in auditory hallucinations reduction: part 1: superior temporal gyrus -preliminary evidence. Psychiatry Res. 286, 112862 https://doi.org/10.1016/j.psychres.2020.112862.

Paret, C., Goldway, N., Zich, C., Keynan, J.N., Hendler, T., Linden, D., Cohen Kadosh, K., 2019. Current progress in real-time functional magnetic resonance-based neurofeedback: Methodological challenges and achievements. Neuroimage 202, 116107. https://doi.org/10.1016/j.neuroimage.2019.116107.

Pelphrey, K.A., Morris, J.P., McCarthy, G., 2004. Grasping the intentions of others: the perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. J. Cogn. Neurosci. 16 (10), 1706–1716. https://doi.org/10.1162/0898929042947900.

Penn, D.L., Keefe, R.S.E., Davis, S.M., Meyer, P.S., Perkins, D.O., Losardo, D., Lieberman, J.A., 2009. The effects of antipsychotic medications on emotion perception in patients with chronic schizophrenia in the CATIE trial. Schizophr. Res. 115 (1), 17–23. https://doi.org/10.1016/j.schres.2009.08.016.

Pennebaker, J.W., Boyd, R.L., Joran, K., Blackburn, K., 2015. The Development and Psychometric Properties of LIWC2015. University of Texas at Austin.

Pereira, J.A., Sepulveda, P., Rana, M., Montalba, C., Tejos, C., Torres, R., Sitaram, R., Ruiz, S., 2019. Self-regulation of the fusiform face area in autism spectrum: a feasibility study with real-time fMRI neurofeedback. Front. Hum. Neurosci. 13, 446. https://doi.org/10.3389/fnhum.2019.00446.

Perner, J., Aichhorn, M., Kronbichler, M., Staffen, W., Ladurner, G., 2006. Thinking of mental and other representations: the roles of left and right temporo-parietal junction. Soc. Neurosci. 1 (3–4), 245–258. https://doi.org/10.1080/17470910600989896.

Pindi, P., Houenou, J., Piguet, C., Favre, P., 2022. Real-time fMRI neurofeedback as a new treatment for psychiatric disorders: A meta-analysis. Prog. Neuropsychopharmacol. Biol. Psychiatry. 119, 110605. https://doi.org/10.1016/j.pnpbp.2022.110605.

Pinkham, A.E., Harvey, P.D., Penn, D.L., 2017. Social cognition psychometric evaluation: results of the final validation study. Schizophr. Bull. https://doi.org/10.1093/schbul/sbx117 sbx117–sbx117.

Powers, K.E., Chavez, R.S., Heatherton, T.F., 2016. Individual differences in response of dorsomedial prefrontal cortex predict daily social behavior. Soc. Cognit. Affect. Neurosci. 11 (1), 121–126. https://doi.org/10.1093/scan/nsv096.

Preckel, K., Kanske, P., Singer, T., 2018. On the interaction of social affect and cognition: empathy, compassion and theory of mind. Curr. Opin. Behav. Sci. 19, 1–6. https://doi.org/10.1016/j.cobeha.2017.07.010.

Premack, D., Woodruff, G., 1978. Does the chimpanzee have a theory of mind? Behav. Brain Sci. 1 (4), 515–526. https://doi.org/10.1017/S0140525X00076512.

R Core Team. (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (4.2.0). https://www.R-project.org/.

Rameson, L.T., Morelli, S.A., Lieberman, M.D., 2012. The neural correlates of empathy: experience, automaticity, and prosocial behavior. J. Cogn. Neurosci. 24 (1), 235–245. https://doi.org/10.1162/jocn_a_00130.

Reis, H.T., Clark, M.S., Holmes, J.G., 2004. Perceived partner responsiveness as an organizing construct in the study of intimacy and closeness. Handbook of Closeness and Intimacy. Lawrence Erlbaum, pp. 201–225.

Reis, H.T., Lemay, E.P., Finkenauer, C., 2017. Toward understanding understanding: The importance of feeling understood in relationships. Soc. Pers. Psychol. Compass 11 (3), e12308. https://doi.org/10.1111/spc3.12308.

Revelle, W., 2022. psych: Procedures for Psychological, Psychometric, and Personality Research (2.2.5). Northwester University. https://CRAN.R-project.org/package=psych.

Rice, K., Redcay, E., 2015. Spontaneous mentalizing captures variability in the cortical thickness of social brain regions. Soc. Cognit. Affect. Neurosci. 10 (3), 327–334. https://doi.org/10.1093/scan/nsu081.

Ros, T., Enriquez-Geppert, S., Zotev, V., Young, K.D., Wood, G., Whitfield-Gabrieli, S., Wan, F., Vuilleumier, P., Vialatte, F., Van De Ville, D., Todder, D., Surmeli, T., Sulzer, J.S., Strehl, U., Sterman, M.B., Steiner, N.J., Sorger, B., Soekadar, S.R., Sitaram, R., Thibault, R.T., 2020. Consensus on the reporting and experimental design of clinical and cognitive-behavioural neurofeedback studies (CRED-nf checklist). Brain 143 (6), 1674–1685. https://doi.org/10.1093/brain/awaa009.

RStudio Team. (2020). RStudio: Integrated development for R (1.3.1093). http://www.rstudio.com/.

Ruiz, S., Lee, S., Soekadar, S.R., Caria, A., Veit, R., Kircher, T., Birbaumer, N., Sitaram, R., 2013. Acquired self-control of insula cortex modulates emotion recognition and brain network connectivity in schizophrenia. Hum. Brain Mapp. 34 (1), 200–212. https://doi.org/10.1002/hbm.21427.

Ruscio, J., 2008. A probability-based measure of effect size: Robustness to base rates and other factors. Psychological Methods 13 (1), 19–30. https://doi.org/10.1037/1082-989X.13.1.19.

Samson, D., Apperly, I.A., Chiavarino, C., Humphreys, G.W., 2004. Left temporoparietal junction is necessary for representing someone else's belief. Nat. Neurosci. 7 (5), 499–500. https://doi.org/10.1038/nn1223.

Saxe, R., Kanwisher, N., 2003. People thinking about thinking peopleThe role of the temporo-parietal junction in "theory of mind. Neuroimage 19 (4), 1835–1842. https://doi.org/10.1016/S1053-8119(03)00230-1.

Saxe, R., Powell, L.J., 2006. It's the thought that counts: specific brain regions for one component of theory of mind. Psychol. Sci. 17 (8), 692–699. https://doi.org/10.1111/j.1467-9280.2006.01768.x.

Saxe, R., Wexler, A., 2005. Making sense of another mind: the role of the right temporo-parietal junction. Neuropsychologia 43 (10), 1391–1399. https://doi.org/10.1016/j.neuropsychologia.2005.02.013.

Saxe, R., Xiao, D.K., Kovacs, G., Perrett, D.I., Kanwisher, N., 2004. A region of right posterior superior temporal sulcus responds to observed intentional actions. Neuropsychologia 42 (11), 1435–1446. https://doi.org/10.1016/j.neuropsychologia.2004.04.015.

Schaafsma, S.M., Pfaff, D.W., Spunt, R.P., Adolphs, R., 2015. Deconstructing and reconstructing theory of mind. Trends Cogn. Sci. 19 (2), 65–72. https://doi.org/10.1016/j.tics.2014.11.007.

Scharnowski, F., Weiskopf, N., 2015. Cognitive enhancement through real-time fMRI neurofeedback. Curr. Opin. Behav. Sci. 4, 122–127. https://doi.org/10.1016/j.cobeha.2015.05.001.

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., Perner, J., 2014. Fractionating theory of mind: a meta-analysis of functional brain imaging studies. Neurosci. Biobehav. Rev. 42, 9–34. https://doi.org/10.1016/j.neubiorev.2014.01.009.

Schurz, M., Radua, J., Tholen, M.G., Maliske, L., Margulies, D.S., Mars, R.B., Sallet, J., Kanske, P., 2021. Toward a hierarchical model of social cognition: a neuroimaging meta-analysis and integrative review of empathy and theory of mind. Psychol. Bull. 147 (3), 293–327. https://doi.org/10.1037/bul0000303.

Sergi, M.J., Green, M.F., Widmark, C., Reist, C., Erhart, S., Braff, D.L., Kee, K.S., Marder, S.R., Mintz, J., 2007. Social cognition and neurocognition: effects of risperidone, olanzapine, and haloperidol. Am. J. Psychiatry 164 (10), 1585–1592. https://doi.org/10.1176/appi.ajp.2007.06091515.

Sitaram, R., Ros, T., Stoeckel, L., Haller, S., Scharnowski, F., Lewis-Peacock, J., Weiskopf, N., Blefari, M.L., Rana, M., Oblak, E., Birbaumer, N., Sulzer, J., 2017. Closed-loop brain training: the science of neurofeedback. Nat. Rev. Neurosci. 18 (2), 86–100. https://doi.org/10.1038/nrn.2016.164.

Skerry, A.E., Saxe, R., 2014. A common neural code for perceived and inferred emotion. J. Neurosci. 34 (48), 15997–16008. https://doi.org/10.1523/JNEUROSCI.1676-14.2014.

Skerry, A.E., Saxe, R., 2015. Neural representations of emotion are organized around abstract event features. Curr. Biol. 25 (15), 1945–1954. https://doi.org/10.1016/j.cub.2015.06.009.

Slaughter, V., Dennis, M.J., Pritchard, M., 2002. Theory of mind and peer acceptance in preschool children. Br. J. Dev. Psychol. 20 (4), 545–564. https://doi.org/10.1348/026151002760390945.

Slaughter, V., Imuta, K., Peterson, C.C., Henry, J.D., 2015. Meta-analysis of theory of mind and peer popularity in the preschool and early school years. Child Dev. 86 (4), 1159–1174. https://doi.org/10.1111/cdev.12372.

Smith, K.P., Christakis, N.A., 2008. Social networks and health. Annu. Rev. Sociol. 34 (1), 405–429. https://doi.org/10.1146/annurev.soc.34.040507.134601.

Sorger, B., Scharnowski, F., Linden, D.E.J., Hampson, M., Young, K.D., 2019. Control freaks: towards optimal selection of control conditions for fMRI neurofeedback studies. Neuroimage 186, 256–265. https://doi.org/10.1016/j.neuroimage.2018.11.004.

Stoeckel, L.E., Garrison, K.A., Ghosh, S.S., Wighton, P., Hanlon, C.A., Gilman, J.M., Greer, S., Turk-Browne, N.B., deBettencourt, M.T., Scheinost, D., Craddock, C., Thompson, T., Calderon, V., Bauer, C.C., George, M., Breiter, H.C., Whitfield-Gabrieli, S., Gabrieli, J.D., LaConte, S.M., Evins, A.E, 2014. Optimizing real time fMRI neurofeedback for therapeutic discovery and development. NeuroImage Clin. 5, 245–255. https://doi.org/10.1016/j.nicl.2014.07.002.

Sukhodolsky, D.G., Walsh, C., Koller, W.N., Eilbott, J., Rance, M., Fulbright, R.K., Zhao, Z., Bloch, M.H., King, R., Leckman, J.F., Scheinost, D., Pittman, B., Hampson, M., 2020. Randomized, sham-controlled trial of real-time functional magnetic resonance imaging neurofeedback for tics in adolescents with tourette syndrome. Biol. Psychiatry 87 (12), 1063–1070. https://doi.org/10.1016/j.biopsych.2019.07.035.

Sulzer, J., Haller, S., Scharnowski, F., Weiskopf, N., Birbaumer, N., Blefari, M.L., Bruehl, A.B., Cohen, L.G., deCharms, R.C., Gassert, R., Goebel, R., Herwig, U., LaConte, S., Linden, D., Luft, A., Seifritz, E., Sitaram, R., 2013. Real-time fMRI neurofeedback: progress and challenges. Neuroimage 76, 386–399. https://doi.org/10.1016/j.neuroimage.2013.03.033.

Tamir, D.I., Mitchell, J.P., 2010. Neural correlates of anchoring-and-adjustment during mentalizing. Proc. Natl Acad. Sci. 107 (24), 10827–10832. https://doi.org/10.1073/pnas.1003242107.

Taschereau-Dumouchel, V., Cushing, C.A., Lau, H., 2022. Real-time functional MRI in the treatment of mental health disorders. Annu. Rev. Clin. Psychol. 18 (1), 125–154. https://doi.org/10.1146/annurev-clinpsy-072220-014550.

Thibaudeau, É., Cellard, C., Turcotte, M., Achim, A.M., 2021. Functional impairments and theory of mind deficits in schizophrenia: a meta-analysis of the associations. Schizophr. Bull. 47 (3), 695–711. https://doi.org/10.1093/schbul/sbaa182.

Thibault, R.T., MacPherson, A., Lifshitz, M., Roth, R.R., Raz, A., 2018. Neurofeedback with fMRI: a critical systematic review. Neuroimage 172, 786–807. https://doi.org/10.1016/j.neuroimage.2017.12.071.

Tursic, A., Eck, J., Lührs, M., Linden, D.E.J., Goebel, R., 2020. A systematic review of fMRI neurofeedback reporting and effects in clinical populations. NeuroImage Clin. 28, 102496 https://doi.org/10.1016/j.nicl.2020.102496.

Tusche, A., Böckler, A., Kanske, P., Trautwein, F.M., Singer, T., 2016. Decoding the charitable brain: empathy, perspective taking, and attention shifts differentially predict altruistic giving. J. Neurosci. 36 (17), 4719–4732. https://doi.org/10.1523/JNEUROSCI.3392-15.2016.

Van Overwalle, F., 2009. Social cognition and the brain: a meta-analysis. Hum. Brain Mapp. 30 (3), 829–858. https://doi.org/10.1002/hbm.20547.

Vossel, S., Weidner, R., Thiel, C.M., Fink, G.R., 2009. What is "Odd" in Posner's location-cueing paradigm? neural responses to unexpected location and feature changes compared. J. Cogn. Neurosci. 21 (1), 30–41. https://doi.org/10.1162/jocn.2009.21003.

Watanabe, T., Sasaki, Y., Shibata, K., Kawato, M., 2017. Advances in fMRI real-time neurofeedback. Trends Cogn. Sci. 21 (12), 997–1010. https://doi.org/10.1016/j.tics.2017.09.010.

Watson, A.C., Nixon, C.L., Wilson, A., Capage, L., 1999. Social interaction skills and theory of mind in young children. Dev. Psychol. 35 (2), 386–391. https://doi.org/10.1037/0012-1649.35.2.386.

Wechsler, D., 2011. WASI-II: Wechsler abbreviated Scale of Intelligence. Psychological Corporation.

Weiskopf, N., Sitaram, R., Josephs, O., Veit, R., Scharnowski, F., Goebel, R., Birbaumer, N., Deichmann, R., Mathiak, K., 2007. Real-time functional magnetic resonance imaging: methods and applications. Magn. Reson. Imaging 25 (6), 989–1003. https://doi.org/10.1016/j.mri.2007.02.007.

Weiskopf, N., 2012. Real-time fMRI and its application to neurofeedback. Neuroimage 62 (2), 682–692. https://doi.org/10.1016/j.neuroimage.2011.10.009.

Wellman, H.M., Cross, D., Watson, J., 2001. Meta-analysis of theory-of-mind development: the truth about false belief. Child Dev. 72 (3), 655–684. https://doi.org/10.1111/1467-8624.00304.

Wickham, H., 2016. ggplot2: Elegant Graphics For Data Analysis (2nd ed. 2016). Springer International Publishing :, Imprint: Springer https://doi.org/10.1007/978-3-319-24277-4.

Yang, Y.C., Boen, C., Gerken, K., Li, T., Schorpp, K., Harris, K.M., 2016. Social relationships and physiological determinants of longevity across the human life span. Proc. Nat. Acad. Sci. 113 (3), 578–583. https://doi.org/10.1073/pnas.1511085112.

Yao, S., Becker, B., Geng, Y., Zhao, Z., Xu, X., Zhao, W., Ren, P., Kendrick, K.M., 2016. Voluntary control of anterior insula and its functional connections is feedback-independent and increases pain empathy. Neuroimage 130, 230–240. https://doi.org/10.1016/j.neuroimage.2016.02.035.

Yeo, H., Yoon, S., Lee, J., Kurtz, M.M., Choi, K., 2022. A meta-analysis of the effects of social-cognitive training in schizophrenia: The role of treatment characteristics and study quality. Br. J. Clin. Psychol. 61 (1), 37–57. https://doi.org/10.1111/bjc.12320.

Young, L., Camprodon, J.A., Hauser, M., Pascual-Leone, A., Saxe, R., 2010. Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. Proc. Nat. Acad. Sci. 107 (15), 6753–6758. https://doi.org/10.1073/pnas.0914826107.

Zaki, J., Ochsner, K.N., 2012. The neuroscience of empathy: progress, pitfalls and promise. Nat. Neurosci. 15 (5), 675–680. https://doi.org/10.1038/nn.3085.

Zaki, J., Weber, J., Bolger, N., Ochsner, K., 2009. The neural bases of empathic accuracy. Proc. Nat .Acad. Sci. 106 (27), 11382–11387. https://doi.org/10.1073/pnas.0902666106.