This paper was published by Oxford University Press as:
Klock, L., Voss, M., Weichenberger, M., Kathmann, N., & Kühn, S. (2021). The thought from the machine: Neural basis of thoughts with a coherent and diminished sense of authorship. *Schizophrenia Bulletin*, 47(6), 1631–1641.
https://doi.org/10.1093/schbul/sbab074

The Thought From the Machine: Neural Basis of Thoughts With a Coherent and Diminished Sense of Authorship

Leonie Klock*[,1,2], Martin Voss[3], Markus Weichenberger[4], Norbert Kathmann[5], and Simone Kühn[2,6]

[1]Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, Berlin, Germany; [2]Clinic and Policlinic for Psychiatry and Psychotherapy, University Clinic Hamburg-Eppendorf, Hamburg, Germany; [3]Department of Psychiatry and Psychotherapy, Charité University Medicine and St. Hedwig-Krankenhaus, Berlin, Germany; [4]Center for Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany; [5]Department of Clinical Psychology, Humboldt-Universität zu Berlin, Berlin, Germany; [6]Lise Meitner Group for Environmental Neuroscience, Max Planck Institute for Human Development, Berlin, Germany.

*To whom correspondence should be addressed; Clinic and Policlinic for Psychiatry and Psychotherapy, University Clinic Hamburg-Eppendorf, Martinistraße 52, 20246, Hamburg, Germany; tel: 040-7410-24122, e-mail: l.klock@uke.de

Patients with schizophrenia who experience inserted thoughts report a diminished sense of thought authorship. Based on its elusive neural basis, this functional neuroimaging study used a novel setup to convince healthy participants that a technical device triggers thoughts in their stream of consciousness. Self-reports indicate that participants experienced their thoughts as self-generated when they believed the (fake) device was deactivated, and attributed their thoughts externally when they believed the device was activated—an experience usually only reported by patients diagnosed with schizophrenia. Distinct activations in the medial prefrontal cortex (mPFC) were observed: ventral mPFC activation was linked to a sense of thought authorship and dorsal mPFC activation to a diminished sense of thought authorship. This functional differentiation corresponds to research on self- and other-oriented reflection processes and on patients with schizophrenia who show abnormal mPFC activation. Results thus support the notion that the mPFC might be involved in thought authorship as well as anomalous self-experiences.

Key words: self/externally-generated/self-generated/ fMRI/ventral medial prefrontal cortex/dorsal medial prefrontal cortex/cortical midline structures

Introduction

Individuals who experience the symptom of thought insertion report that they are not the cause of their thoughts, and attribute authorship to an external entity.[1] Accordingly, conceptual approaches assume that thought insertion is characterized by a lack of control and authorship of thoughts.[2] Thought insertion is one of the "first-rank symptoms" of schizophrenia[3]; its persistence for over a month leads to a diagnosis of schizophrenia as per the International Classification of Diseases (ICD10).[4] Moreover, it is one of the extreme examples of a range of transformations in the basic sense of self that have been collectively named self-disorders and regarded as key features of schizophrenia since early descriptions of the illness.[5,6] It is thus surprising that there has not yet been a systematic investigation of its neurocognitive and neurobiological basis.

Healthy individuals build and maintain a continuous sense of self that binds together subjective experiences, body, actions and thoughts distinct from others. Neuroimaging studies consistently report that the cortical midline structures (CMS) are crucially involved in self-reflection and show enhanced activation when individuals process information closely linked to themselves.[7,8] CMS include the ventral medial prefrontal cortex (vmPFC), dorsal medial prefrontal cortex (dmPFC), anterior cingulate cortex, and posterior cingulate cortex, encompassing the precuneus.[7] Accordingly, CMS are assumed to play a key role in an altered sense of self as experienced by patients with schizophrenia.[9,10]

Given the lack of research on thought insertion, the presented study sought to explore the neural basis of thoughts for which participants did or did not report a sense of authorship. This research has been limited by the challenge of provoking the symptom within the context of a functional magnetic resonance imaging (fMRI) study. Consequently, a novel experimental design was developed, in which healthy participants would experience thoughts for which they did not report a sense of authorship. In particular, it convinced participants that a fictional transcranial magnetic stimulation (TMS) device was able to trigger thoughts, thus creating an unusual

context for them to question the causal origin of their own thoughts. Self-report measures indicate that the experimental design successfully provoked healthy participants to attribute the cause of their thoughts to the fictional TMS, indicating an altered sense of authorship—often reported by patients with schizophrenia. This makes the presented study the first to directly compare neural correlates of thoughts with a coherent and diminished sense of authorship. Based on their role in self-referential processes[7,8] and previous speculations that alterations in CMS might be linked to anomalies in the sense of self,[9,10] we expected activation in the CMS to be associated with thought monitoring processes of thought authorship.

Methods

Participants

Thirty right-handed[11] participants were recruited via an e-mail list of students at the Humboldt-Universität zu Berlin. Individuals who did not study psychology, medicine, or physics and never underwent TMS or MRI examinations were eligible (one reported a previous MRI). Additional exclusion criteria were a history of psychiatric disorders, assessed via the Mini-International Neuropsychiatric Interview,[12] neurological diseases, head trauma or metallic implants. One participant was excluded due to excessive head movement (10 mm), which rendered a sample of 29 participants (Mean (M) age $= 22.7\ y$, 17 females). Volunteers were compensated 30€ and consented upfront to participate in a combined TMS and fMRI study. The ethics committee of the German Society for Psychology in Berlin approved the procedure.

Task Design

This study sought to alter the sense of thought authorship of healthy individuals. To initially conceal the study objective, participants were invited to a "combined fMRI and TMS study" exploring neural processes of thinking about own thoughts. To allow for participants to attribute their thoughts externally, they were presented with a normal MRI head coil and falsely informed that it was an MRI-compatible TMS device that would be used to stimulate their brain and thus to externally generate thoughts.

Task Instructions. First, the investigator provided participants with illustrative and accurate information regarding the methods of TMS and fMRI. Participants were then introduced to an alleged TMS expert. He presented a video from a real educational program depicting how TMS can induce involuntary hand movements and explained that a special MRI-compatible TMS device would be employed, which is built into a helmet to cover the participant's entire head. Next, the investigator provided participants with the crucial (but fallacious) information that thinking about an animal activates an area in the prefrontal cortex and that this area would be stimulated during the experiment to trigger a thought of an animal. The investigator explained that a structural MRI scan would first locate this cortical area to subsequently focus the TMS to it. Moreover, the investigator misleadingly compared the process of triggering a thought of a certain animal to the process of triggering an involuntary hand movement as seen in the video. To enhance plausibility and to set expectations on the perceivability of effects of the TMS, the investigator explained that, unlike in the video, only an imperceptible and inaudible stimulation would be employed.

Main Experimental Design. First, a colored cue informed participants about the mode of the fictional TMS device: a green circle signified that the TMS was activated to trigger a thought of an animal (ON), a red circle that it was deactivated (OFF), and an orange circle signified an ambiguous condition, in which no explicit (and false) information was provided whether the TMS was activated or deactivated (ambiguous; figure 1). During a 10s monitoring interval, participants were instructed to think of one animal that was coming to their mind during each trial and to consider whether the occurring thought was self- or externally generated, while a row of dots slowly appeared on the screen to indicate stimulation activity. Next, a "Thought Authorship Question" prompted participants to indicate how sure they were that the thought was self-generated and a "Thought Control Question" whether they were able to control it. Participants answered by moving a cursor with their right hand along a continuous visual analogue scale ranging from "very sure/Me" to "not sure/TMS" for the Authorship and from "full control" to "no control" for the Thought Control question. The scales were continuous in that they had no visible numbers, markers, or increments, and could be clicked on anywhere in between. Scale direction varied randomly to prevent anticipation. Participants were not time restricted and were encouraged to use the entire scale. A "TMS Question" concluding each trial required participants to indicate whether they believed that the TMS was activated and stimulated their brain (by clicking on a green circle), or whether they believed that it was deactivated (by clicking on a red circle). The ambiguous condition, in which participants did not receive information regarding the TMS modus, was especially important because it allowed an assessment of what participants *thought* happened during the monitoring interval (subsequently divided into rated_ON and rated_OFF trials, based on participants' response). A new trial started with the presentation of a fixation cross that varied in its duration between 4 s and 6 s in steps of 500 ms. In total, 80 trials (20 ON, 20 OFF, 40 ambiguous) were presented randomly in two blocks without informing participants about their frequency.

It is important to emphasize that only participants' belief regarding the influence of the TMS differed between experimental conditions. We assumed

Originally published in: Schizophrenia Bulletin, 47, 2021, p. 1632

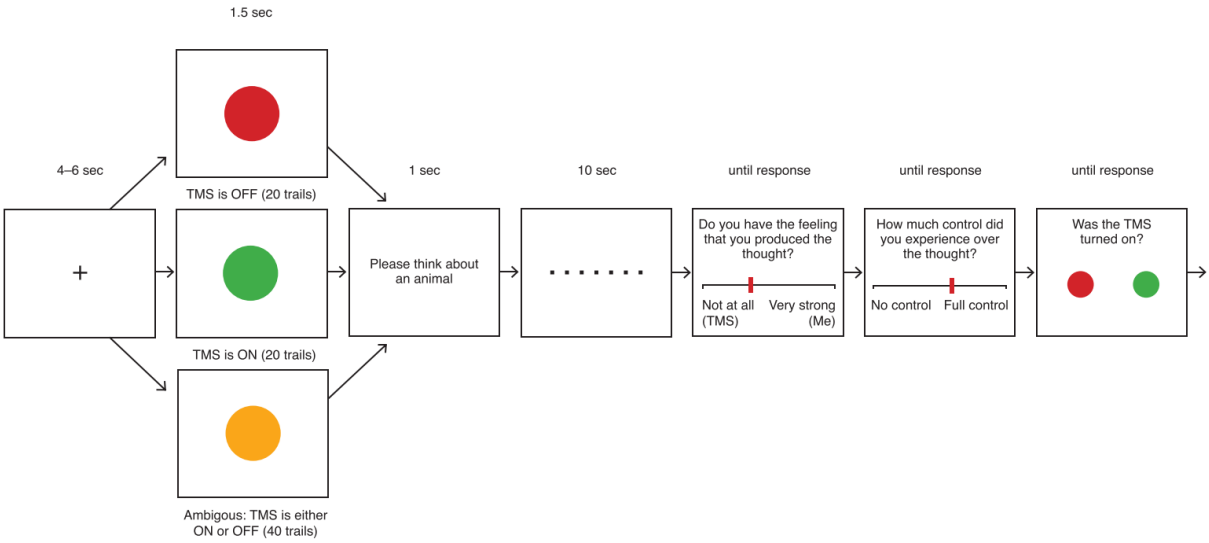

Fig. 1. Experimental task: In three experimental conditions participants received different information regarding the operation mode of the fictional TMS: OFF (TMS is turned off; red cue), ON (TMS is turned on; green cue), ambiguous (TMS is either turned off or on; orange cue). The experimental setup thus allowed the manipulation of participant's belief whether the fictional TMS device influenced their stream of thoughts and triggered a thought of an animal.

participants would adopt two different stances: when informed that the TMS was off (OFF trials), we expected them to experience a coherent sense of thought authorship. In contrast, when informed that the TMS was on (ON trials), we expected them to closely monitor their thoughts and possibly experience them to be externally triggered (diminished sense of authorship). Crucially, we assumed that only *profound* alterations in the sense of thought authorship during ambiguous trials (during which no information on the TMS mode was provided) would cause participants to rate the TMS as activated (rated_ON). More precisely, we expected particularly low authorship ratings in ambiguous trials rated_ON, and thus considered these trials critical for the investigation of a diminished sense of authorship. Furthermore, we considered the number of ambiguous rated_ON trials to indicate the credibility of the experimental design. We directly probed participants' conviction after the final trial by asking how sure they were that the TMS device could influence thoughts (Global Conviction), and influenced their thoughts (Personal Conviction). The continuous response scale ranged from "not at all sure" to "very sure."

Questionnaires. After leaving the scanner and before debriefing, participants completed the Schizotypal Personality Questionnaire (SPQ), a self-report measure of 74 binary questions to assess nine schizotypal traits according to the DSM-III-R.[13]

At the end of the session, participants were debriefed by a psychologist and informed that no TMS device was present and that it is currently not possible to trigger specific thoughts by means of cortical stimulation, ensuring that every participant understood the rationale of us having to use this experimental setup in the context of the study before leaving.

Data acquisition and Analysis

fMRI Image Acquisition. Images were acquired on a 3-Tesla Magnetom Trio MRI scanner (Siemens Medical Systems) using a 12-channel radiofrequency head coil. T1-weighted structural images were obtained (repetition time (TR) = 2500 ms; echo time (TE) = 4.77 ms; inversion time = 1100 ms, acquisition matrix = 256 × 256 × 176, flip angle = 7˚; 1 × 1 × 1 mm^3 voxel size). Functional MRI images were collected using a T2*-weighted echo-planar imaging (EPI) sequence (TR = 2000 ms, TE = 30 ms, image matrix = 64 × 64, 3 × 3 × 3 mm^3 voxel size, 36 slices). Presentation software (www.neurobs.com) and Visuastim video goggles (Resonance Technology, Inc., USA) were used for visual presentation and responses were obtained using an MR compatible trackball (Current Designs Inc., USA).

FMRI Data Preprocessing and Analysis. Preprocessing and statistical analysis were computed using SPM12 (http://www.fil.ion.ucl.ac.uk/spm/) implemented in Matlab (The Mathwork Inc.). First, EPI images were slice-time corrected and realigned to correct for head movements. To spatially realign images, a mean EPI image was constructed and used as a reference. The structural T1 image was coregistered with that mean EPI. First, T1 images were segmented and then normalized to the Montreal Neurological Institute (MNI) template together with the EPI images (unified segmentation). Finally, images were smoothed with a Gaussian kernel of 8 mm FWHM. Preprocessed images were analyzed

with two different statistical models (common general linear model (GLM) and parametric GLM) as implemented in SPM12. The monitoring intervals of the relevant trials were modeled with a duration of 10 s. On the single-subject level, the design matrix included the relevant onsets that were convolved with a canonical hemodynamic response function and its temporal derivatives and six realignment parameters. A high-pass filter of 128 was used.

Common GLM. A GLM with four regressors was calculated to compare the blood-oxygen-level-dependent (BOLD) signals in trials of different conditions: regressors referred to the onsets of ON, OFF, ambiguous rated_ON, and rated_OFF trials, respectively. Statistical parametric effects for each voxel and each subject were calculated for the resulting contrasts. Individual contrast images were then entered into a second-level analysis to calculate between-subject statistics.

Parametric GLM. To identify whether signal intensities during the monitoring interval were modulated by individual authorship ratings at the end of the trial, the onsets of all monitoring intervals were modeled as one regressor and the corresponding authorship ratings as a second. Contrast maps for the authorship regressor were obtained for each individual and then entered in a second-level random-effects analysis. A *t*-test across all subjects tested for statistically significant positive and negative modulations.

For all reported contrasts, resulting statistical values were thresholded with a significance level of $p < .001$ on a voxel level, and with a family-wise error correction (FWE) of $p < .05$ on a cluster level. Based on previous literature implicating the CMS in self-referential processes,[7,8] we hypothesized that thought authorship monitoring processes would be linked to enhanced neural activation in these areas and employed small volume correction in the CMS to correct for multiple comparisons. The region-of-interest (ROI) was based on the Automatic Anatomical Labeling (AAL) atlas[14] employing WFU pickatlas.[15] The ROI encompassed medial superior frontal gyrus, medial orbital gyrus, gyrus rectus, supplemental motor area, precuneus, paracentral lobe, and cingulum (anterior, mid and posterior). For exploratory purposes of this novel fMRI task, we additionally report whole-brain results with a level of significance of $p < .001$, with an FWE correction of $p < .05$ on cluster level (see supplementary material. Supplementary material to this article is available. For more information see http://hdl.handle.net/21.11116/0000-0008-979E-C). Labeling of activation clusters was performed based on the AAL atlas.

Behavioral Data Analysis. Behavioral data were analyzed using Matlab and SPSS21 (IBM). Participants' responses on the continuous visual analogue scales were converted from their precise location on the scales to values ranging from 0 to 100. High values represented a coherent sense of authorship (eg, self-generated)/full control; low values represented a diminished sense of authorship (eg, externally-generated)/no control. On the Conviction Scales, high values indicated strong and low values weak conviction. Self-ratings across different conditions were compared using paired sample t-tests. To explore experimentally induced alterations in thought authorship, we computed mean differences in authorship ratings between ambiguous rated_ON and rated_OFF trials (Altered Authorship Score). To investigate authorship ratings throughout the experiment, we split the total amount of 80 trials into three bins (first part: first 27 trials/second part: middle 26 trials/third part: last 27 trials) and employed a repeated measure ANOVA to compare mean authorship ratings as well as the number of ambiguous rated_ON trials.

Results

Behavioral Results

TMS Question. Participants indicated in 44.1% of ambiguous trials that the fictional TMS was activated. Ambiguous trials were thus divided into two groups depending on the individuals' judgment in the TMS question (rated_ON and rated_OFF). Throughout the experiment the number of trials participants judged the fictional TMS device to be activated increased numerically (table 1).

Thought Authorship Question. Participants reported significantly less authorship, ie, attributed the thought more to the TMS in ON compared to OFF trials and in rated_ON compared to rated_OFF trials (table 1, Figure 2A). Importantly, participants reported even less authorship in ambiguous rated_ON trials when they received no external suggestion about the TMS modus compared to ON trials when they were informed that it was activated. This finding confirms our hypothesis that ambiguous rated_ON trials would be accompanied by low authorship ratings, indicating profound alterations in the experience of thought authorship. The ambiguous condition was therefore a priori the most critical for the investigation of the neural basis of an anomalous sense of thought authorship. Throughout the experiment, participants reported significantly less thought authorship.

Thought Control Question. Participants reported significantly less control over thoughts in ON than in OFF trials and in rated_ON compared to rated_OFF trials. Comparing ON and rated_ON trials, participants reported significantly less control in rated_ON trials than in ON trials.

To explore the relationship between the Thought Authorship and Thought Control Questions, we calculated the correlation coefficient for the mean difference scores of the ratings in rated_ON and rated_OFF trials, which was statistically significant ($r(29) = 0.752$; $p < .001$). Based on our primary research interest in an anomalous sense of thought authorship, we focus here on the results of the Thought Authorship Question and its relation to fMRI data.

Table 1. Results of Self-Reports.

	OFF	ON	Ambiguous				Bins of trials		
			rated_ON	rated_OFF	Global	Personal	1st	2nd	3rd
TMS							5.4 *(0.35)*	5.7 *(0.48)*	6.5 *(0.46*
Question							$F(2\ 56) = 2.309, p = .109$		
Authorship	85.0 *(1.4)*	67.5 *(3.1)*	58.5 *(3.5)*	85.4 *(1.5)*			77.0 *(2.1)*	75.8 *(2.5)*	73.8 *(2.5*
Question	$\text{Mean}_{\text{Difference}} = 17.5$ *(2.1)* $t(28) = 8.2, p < .001$		$\text{Mean}_{\text{Difference}} = -27.0$ *(2.8)* $t(28) = -9.7, p < .001$				$F(2\ 56) = 3.441, p = .039$		
			$\text{Mean}_{\text{Difference}} = 9.0$ *(1.8)* $t(28) = 5.0, p < 0.001$						
Control	80.4 *(1.8)*	58.8 *(3.5)*	48.4 *(3.2)*	82.1 *(1.7)*					
Question	$\text{Mean}_{\text{Difference}} = 21.7$ *(2.8)* $t(28) = 7.7, p < .001$		$\text{Mean}_{\text{Difference}} = -33.7$ *(2.5)* $t(28) = -13.4, p < .001$						
			$\text{Mean}_{\text{Difference}} = 10.4$ *(1.7)* $t(28) = 6.1, p < .001$						
Conviction					33.6 *(4.6)*	39.6 *(4.8)*			
Questions					$r(28) = 0.851, p < .001$				

Mean behavioral ratings and standard error of the mean (SEM) are reported for $N = 29$ participants as well as p values for paired sample t-test and Pearson correlation.

Conviction Question. Participants rated their belief in the functioning of the TMS device on a global and personal conviction scale (one rating is missing due to technical problems.) The two conviction ratings were significantly correlated with each other (table 1) and with the Altered Authorship Score (general: $r(29) = 0.445$, $p = .016$; personal: $r(28) = 0.388$, $p = .042$) indicating that participants who believed more strongly in the experimental setup reported greater differences in their authorship experience between ambiguous rated_ON and rated_OFF trials.

Questionnaires. To explore whether experimentally induced transformations in the sense of authorship were related to personality traits, correlations between the altered authorship score and the nine subscales of the SPQ (Ideas of reference: $M = 0.90$, *standard error of the mean* $(SEM) = 0.2$; Excessive social anxiety: $M = 1.41$, $SEM = 0.26$; Magical thinking: $M = 0.45$, $SEM = 0.18$; Unusual perceptual experiences: $M = 0.28$, $SEM = 0.1$; Odd behavior $M = 0.76$, $SEM = 0.21$; No close friends: $M = 0.83$, $SEM = 0.30$; Odd speech: $M = 1.72$, $SEM = 0.33$; Constricted affect: $M = 1.1$, $SEM = 0.30$; Suspiciousness: $M = 0.69$, $SEM = 0.16$; Sum score: $M = 8.1$, $SEM = 1.11$) were calculated. Only the ideas of reference subscale yielded a significant relationship with the altered authorship score ($r(29) = 0.477$, $p = .009$ (not surviving a Bonferroni correction with an adjusted alpha level of 0.0056 (0.05/9)).

Imaging Results

Based on previous literature that suggests an involvement of CMS in thought monitoring processes of thought authorship, we used a small volume correction within a defined mask of these ROIs. The following results are reported on an FWE corrected cluster-level within that small volume correction (table 2; Figure 2B).

Thoughts with a Coherent Sense of Authorship. We compared OFF > ON, rated_OFF > rated_ON, and tested for positive parametric modulations. The contrast of OFF > ON revealed increased activation in the vmPFC and in the posterior cingulate extending to the precuneus while the rated_OFF > rated_ON contrast and the parametric GLM did not yield any significant activation cluster.

Thoughts with a Diminished Sense of Authorship. We compared ON > OFF, rated_ON > rated_OFF, and tested for negative parametric modulation of authorship ratings. Of those, ambiguous trials were central for the investigation of anomalous authorship experiences as mean authorship ratings were lowest in rated_ON trials, suggesting fundamental transformations in the experience of authorship. We observed enhanced activation in the dmPFC for the comparison of both ON > OFF and rated_ON > rated_OFF. For the latter contrast, increased activation was also observed in the medial parietal cortex including the precuneus. The parametric GLM did not

Table 2. FMRI Results.

Region	BA	Cluster extent	MNI (x y z)	z Value
Thoughts with a Coherent Sense of Authorship				
OFF > ON				
Bilateral vmPFC: medial orbitofrontal gyrus, gyrus rectus	10/11	118	−3 47 −7	5.53
L posterior cingulate: calcarine gyrus, lingual gyrus, precuneus, cerebellum	17/27/30	27	6 −55 14	4.06
Rated_OFF > Rated_ON				
Positive parametric modulation				
Thoughts with a Diminished Sense of Authorship				
ON > OFF				
Bilateral dmPFC: superior frontal gyrus, middle cingulum, supplementary motor area	8/9/32	133	9 26 59	4.01
Rated_ON > Rated_OFF				
Bilateral parietal lobe: precuneus, cuneus	7	52	−6 −70 32	4.56
Bilateral dmPFC: superior medial frontal gyrus	8/9/32	137	6 50 44	4.52
Negative parametric modulation				

BA, Brodmann area; MNI, Montreal Neurological Institute; L left; R right; dmPFC, dorsal medial prefrontal cortex; vmPFC, ventral medial prefrontal cortex.
Results are corrected with a small volume correction and thresholded with a FWE correction of $p < 0.05$ on a cluster-level.

show any activation cluster negatively modulated by individual authorship ratings.

Discussion

Individuals reporting the symptom of thought insertion describe anomalous experiences regarding the authorship of own thoughts. To explore neural structures involved in this unusual experience, we developed an experimental setup that successfully provoked an altered sense of thought authorship among healthy subjects. We introduced participants to a fictional TMS device that was allegedly able to trigger thoughts, which created the context for them to question the source of their thoughts. Self-ratings indicate that participants adopted two different stances: when informed that the fictional TMS was activated, they were more likely to believe that it was generating their thoughts; when informed that it was off, they believed they generated their thoughts. Even more telling is that in 44% of ambiguous trials (those with no indication of TMS modus) participants judged the TMS to be activated and felt more strongly that it generated their thoughts than in trials in which it was explicitly activated. Furthermore, authorship ratings decreased over time, suggesting that the setup remained convincing. We observed a relationship between a schizotypal personality trait, namely ideas of reference, and the mean difference in authorship judgments between ambiguous rated_ON and rated_OFF trials (although that finding does not survive Bonferroni correction). More explicitly, those healthy individuals who tend to perceive external events such as conversations of other people, advertisements, and information in the media in reference to themselves, experienced a stronger influence of the TMS on their stream of thoughts. In particular, they reported less authorship and attributed their thoughts more strongly to the TMS device. This is particularly interesting as it was observed that healthy high schizotypical individuals experience more frequently self-disorder symptoms than individuals low on schizotypic traits.[16] As it is difficult to study unusual psychopathologies such as self-disorders in patients, this finding further highlights the usefulness of this approach of inducing anomalous experiences in healthy participants. Earlier studies also successfully convinced healthy participants that their thoughts were manipulated[17–20]; the present study, however, induced anomalous authorship experiences in more trials.[19] Altogether, this experimental fMRI design led healthy participants to question the source of their thoughts and believe that an external entity could trigger thoughts—a conviction commonly held by patients with schizophrenia who report inserted thoughts. Furthermore, it allowed us to discern the neural bases of coherent and diminished senses of thought authorship.

Consistent with our hypothesis and various imaging studies exploring self-referential mental processes[7,8] enhanced neural activation along the CMS occurred when participants reflected about their thought authorship. Specifically, both authorship experiences were associated with activation in the anterior CMS. In the medial prefrontal cortex (mPFC), enhanced activation was observed in the vmPFC in OFF trials when participants reported coherent thought authorship, and overlapping activation in the dmPFC in ON and rated_ON trials when they

Originally published in: Schizophrenia Bulletin, 47, 2021, p. 1636

reported a diminished sense of thought authorship. This functional differentiation within the mPFC corresponds to a meta-analysis that revealed a spatial gradient within the mPFC that links the vmPFC to thinking about oneself and the dmPFC to thinking about others.[21]

The present results indicate increased vmPFC activation in trials where subjects experienced coherent thought authorship, consistent with previous findings implicating the vmPFC as a key structure in self-referential thoughts.[21–23] These earlier studies typically relied on self-trait judgments[24] (focusing on the "me"[25] or "narrative self"[26])[27]; this task is novel in that participants monitor their sense of authorship (potentially related to the "I"[25] or "minimal self"[26]). We thus speculate that the vmPFC plays a crucial role in both aspects of the self— including thought authorship. Moreover, the vmPFC has been implicated in the value an individual assigns to different types of rewards[22,28]—D'Argembeau[22] proposed its importance in self-referential processes as it attaches unique value to information tied to oneself. In this study, suprathreshold vmPFC activation was only observed in trials participants were informed that the TMS was deactivated, and not in ambiguous rated_ON trials. In OFF trials participants were certain that they generated their own thoughts, thus we speculate increased vmPFC activity in these trials represents high personal significance.

Enhanced dmPFC activation was observed in the contrasts of ambiguous rated_ON > rated_OFF trials (lowest vs highest authorship ratings, respectively), and ON > OFF trials. The results thus suggest that neural activity in the dmPFC is associated with diminished thought authorship, an unusual experience for healthy participants, which, we propose, shares similarities with inserted thought experiences reported by patients with schizophrenia.

Previous research implicates the dmPFC in reflection processes focusing on the self and others,[23] especially others that we perceive as dissimilar to ourselves[29,30] including distant public others,[31] and in ascribing independent mental states to other people, termed mentalizing.[32–34] Other research links the dmPFC to intentional inhibition, the processes of canceling an action.[35–37] Referring to this functional overlap, Lynn and colleagues[38] proposed that disengaging from a self-centered perspective might be a common mechanism underlying both processes. In light of the current findings, we speculate that this disengagement process associated with the dmPFC allowed healthy participants to question their thought authorship and attribute it externally. Preliminary evidence for this putative functional role comes from research on mindfulness meditation, the practice of observing one's momentary experiences in a nonjudgmental, detached way.[39] Different meditation practices activate the mPFC.[40] Long-term mindfulness meditators show altered dmPFC connectivity at rest[41] and enhanced dmPFC activity during mindfulness meditation[42] and self-appraisal, associated with reports of nonreactance towards experiences.[43] Chinese Buddhists, whose beliefs challenge the existence of static self, showed enhanced activation in the dmPFC (and not vmPFC) during self-referential processing.[44] We speculate that whereas experienced meditators *intentionally* detach from own thoughts, the ambiguity of the study setup triggered detachment from thoughts and induced a diminished sense of authorship in healthy participants—an experience comparable to what patients with schizophrenia report. One patient reported, "I had lost myself, a constant feeling that my self no longer belonged to me. (…) I was simply sectioned again, detached from my real self, observing what was being done to me. in a third-person perspective".[45]

The current study set out to investigate elusive neural underpinnings potentially involved in such self-disorders. Our findings indicate involvement of the mPFC in both coherent and diminished senses of thought authorship. The latter is in accordance with reported structural[46,47] and functional[48] alterations in the mPFC among patients diagnosed with schizophrenia. Regarding self- and other-related processes, patients showed lower activation in the vmPFC during self-reflection[49–52] and in dmPFC during mentalizing.[53] Alterations in these areas might be related to impaired differentiation between self and others, and thus to blurred boundaries between the two—commonly reported among persons with schizophrenia.

Furthermore, the mPFC belongs to a default mode network that activates when individuals are resting without instruction (called resting-state), and shows decreased activation when they are engaging in cognitively demanding tasks.[54,55] The mPFC is assumed to be linked to self-focused resting-state mental activity.[56,57] People with schizophrenia exhibit reduced vmPFC activity[58] and altered connectivity[59] during this resting-state. Hence, Northoff[10] proposed a direct link between these rest-related alterations and self-disorders in patients with schizophrenia. A TMS study that disturbed rest-related mPFC activation accordingly reported abnormal self- awareness among healthy subjects.[60] Furthermore, Nelson and colleagues[9] highlight the crucial role the CMS might play in maintaining a coherent sense of self, and that alterations might be linked to a disturbed sense of self.

In the current study enhanced neural activation was also observed in more posterior regions of the CMS, namely the parietal cortex including different subregions of the precuneus. More specifically, the current findings suggest that the precuneus is involved in monitoring thought authorship with regional specificities: ventral precuneus was activated during thoughts with a coherent sense of thought authorship and dorsal precuneus during thoughts with a diminished sense of thought authorship. The posterior cingulate cortex/precuneus has consistently been found to be involved in self-referential processing,[8,23] the default mode network during episodes of rest[53,55] and its core-self system,[61] which is in line with the current finding of increased activation during OFF trials. Furthermore, a

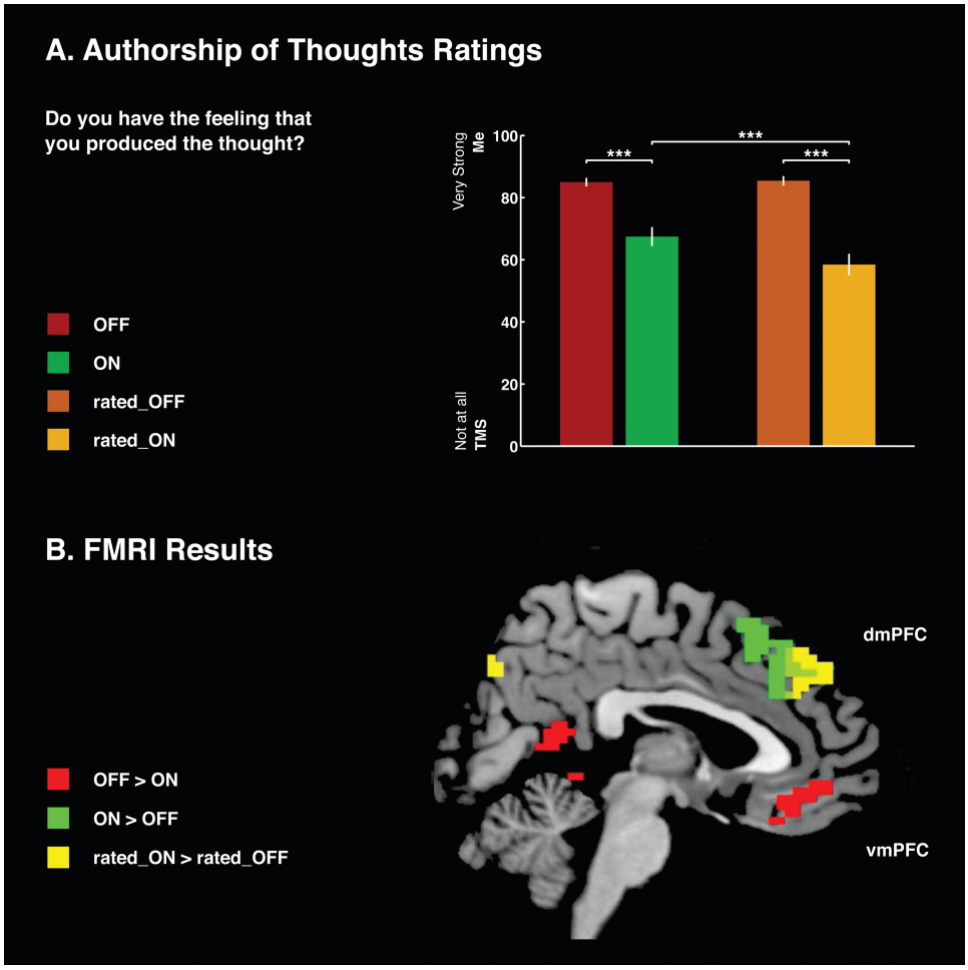

Fig. 2. Behavioral and FMRI results: (A) Significant difference in the authorship of thought ratings between ON and OFF trials and between ambiguous trials divided in rated_OFF and rated_ON. When participants believed that the fictional TMS was turned on (ON, rated_ON), they perceived less authorship and attributed the thought more to the TMS compared to trials in which they believed it was turned off (OFF, rated_OFF). Moreover, when they subjectively indicated in the ambiguous condition that the TMS was turned on (rated_ON), they experienced less authorship than when they were informed that it was turned on (ON). Error bars represent the standard error of the mean (SEM) and asterisk indicates significant differences at $p < .001$. (B) Increased activation in the cortical midline structures was linked to these two subjective experiences regarding the authorship of thoughts: activation in the vmPFC and posterior cingulate/ventral precuneus was observed when participants were informed that the TMS was turned off and monitored their own thoughts (red) and overlapping activation in dmPFC was found when they were informed (green) or believed that the TMS was on (orange; results from a small volume correction are displayed on an FWE corrected level of $p < .05$) and experienced a diminished sense of thought authorship. When participants believed the TMS was turned on, increased activation was also observed in the parietal lobe/ dorsal precuneus.

functional specification within the precuneus for self-referential and episodic memory processes has been reported[62] that corresponds to the current differential findings in that region. It would be an intriguing avenue for future studies to further investigate the role of the precuneus in thoughts with a diminished sense of thought authorship and whether and how these are related to episodic memory processes.

Limitations. We introduced an experimental setup that manipulates the sense of thought authorship among healthy individuals based on the difficulty to repeatedly trigger the symptom of thought insertion in an fMRI session. Clearly, subjective experiences reported here are qualitatively different than those of patients experiencing inserted thoughts and are thus not directly comparable. One important difference relates to the content of external thoughts that feel alien. The instruction to think of an animal may limit results' generalizability. To address this, we conducted a supplemental block in which we instructed participants to let their minds wander freely while maintaining the fictional TMS story (see supplementary material). 33% of ambiguous trials were rated as ON, which we take as evidence that experimentally induced changes in sense of thought authorship are not limited to the instruction to think of an animal. However, we did not observe significant activation in the comparison of ambiguous trials, possibly due to a lack of power of this analysis (for more details, see

supplementary material). Following the supplemental task, we directly tested the credibility of the setup with two conviction questions. Low ratings on these scales might be interpreted as participants being unsure about the credibility of the setup. However, we believe that simply prompting scales that question the TMS device might have created those doubts, but only after the experiment. Concordantly, authorship ratings decreased throughout the experiment, suggesting that the setup remained plausible throughout and that our approach of inducing an anomalous experience in healthy participants might be useful to study unusual psychopathologies. Finally, as this is the first fMRI study to investigate experiences of thought authorship (with $N = 29$), the discussion is preliminary and based on post-hoc interpretations, which emphasizes the importance of future research including a larger sample size to increase statistical power to address the neurobiological basis of this core symptom of schizophrenia.

Conclusion. Healthy participants were introduced to a (fake) TMS device that was allegedly capable of triggering thoughts in their stream of consciousness and thus provoked them to question the authorship of their thoughts. Thoughts that participants subsequently judged as self- and externally-produced were associated with neural activation in different subunits in the mPFC. This finding corresponds to self- and other-referential processes in healthy individuals and observed alterations in patients with schizophrenia who commonly experience anomalous self-experiences. Therefore, the results support the notion that the mPFC might play a central role in the experience of a coherent sense of self, including a sense of thought authorship, and in anomalous self-experiences.

Supplementary Material

Supplementary material to this article is available. For more information see http://hdl.handle.net/21.11116/0000-0008-979E-C

Acknowledgments

We would like to thank Dr Carsten Bogler for his advice throughout the study and Prof Dr Gottfried Vosgerau for comments on the manuscript.

Funding

This study was funded by the Volkswagen Foundation (grant no. VW II/85 067) and LK received a scholarship from Evangelisches Studienwerk Villigst.

Conflict of Interest Statement

None.

References

1. Mullins S, Spence SA. Re-examining thought insertion. Semistructured literature review and conceptual analysis. *Br J Psychiatry.* 2003;182:293–298.
2. Vosgerau G, Voss M. Authorship and control over thoughts. *Mind Lang*. 2014;29:534–565. doi:10.1111/mila.12065
3. Schneider K. *Clinical Psychopathology*. New York, NY: Grune & Stratton; 1959.
4. World Health Organization. *International Classification of Diseases and Related Health Problems*. 10th revised. Geneva: World Health Organization; 1992.
5. Mishara AL, Lysaker PH, Schwartz MA. Self-disturbances in schizophrenia: history, phenomenology, and relevant findings from research on metacognition. *Schizophr Bull.* 2014;40:5–12.
6. Sass LA, Parnas J. Schizophrenia, consciousness, and the self. *Schizophr Bull.* 2003;29:427–444.
7. Northoff G, Bermpohl F. Cortical midline structures and the self. *Trends Cogn Sci*. 2004;8:102–107. doi:https://doi.org/10.1016/j.tics.2004.01.004
8. Northoff G, Heinzel A, de Greck M, Bermpohl F, Dobrowolny H, Panksepp J. Self-referential processing in our brain–a meta-analysis of imaging studies on the self. *Neuroimage.* 2006;31:440–457.
9. Nelson B, Fornito A, Harrison BJ, et al. A disturbed sense of self in the psychosis prodrome: linking phenomenology and neurobiology. *Neurosci Biobehav Rev.* 2009;33:807–817.
10. Northoff G. How is our self-altered in psychiatric disorders? A neurophenomenal approach to psychopathological symptoms. *Psychopathology.* 2014;47:365–376.
11. Oldfield RC. The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia.* 1971;9:97–113.
12. Sheehan DV, Lecrubier Y, Sheehan KH, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry.* 1998;59(Suppl 20):22–33;quiz 34.
13. Raine A. The SPQ: a scale for the assessment of schizotypal personality based on DSM-III-R criteria. *Schizophr Bull.* 1991;17:555–564.
14. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*. 2002;15:273–289. doi:10.1006/ nimg.2001.0978
15. Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage.* 2003;19:1233–1239.
16. Torbet G, Schulze D, Fiedler A, Reuter B. Assessment of self-disorders in a non-clinical population: reliability and association with schizotypy. *Psychiatry Res.* 2015;228:857–865.
17. Swiney L, Sousa P. When our thoughts are not our own: investigating agency misattributions using the mind-to-mind paradigm. *Conscious Cogn.* 2013;22:589–602.
18. Ali SS, Lifshitz M, Raz A. Empirical neuroenchantment: from reading minds to thinking critically. *Front Hum Neurosci.* 2014;8:357.
19. Olson JA, Landry M, Appourchaux K, Raz A. Simulated thought insertion: influencing the sense of agency using deception and magic. *Conscious Cogn.* 2016;43:11–26.

Originally published in: Schizophrenia Bulletin, 47, 2021, p. 1639

20. Walsh E, Oakley DA, Halligan PW, Mehta MA, Deeley Q. The functional anatomy and connectivity of thought insertion and alien control of movement. *Cortex*. 2015;64:380–393.
21. Denny BT, Kober H, Wager TD, Ochsner KN. A meta-analysis of functional neuroimaging studies of self- and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex. *J Cogn Neurosci.* 2012;24:1742–1752.
22. D'Argembeau A. On the role of the ventromedial prefrontal cortex in self-processing: the valuation hypothesis. *Front Hum Neurosci*. 2013;7:372. doi:10.3389/fnhum.2013.00372
23. van der Meer L, Costafreda S, Aleman A, David AS. Self-reflection and the brain: a theoretical review and meta-analysis of neuroimaging studies with implications for schizophrenia. *Neurosci Biobehav Rev.* 2010;34:935–946.
24. Kelley WM, Macrae CN, Wyland CL, Caglar S, Inati S, Heatherton TF. Finding the self? An event-related fMRI study. *J Cogn Neurosci.* 2002;14:785–794.
25. James W. *The Principles of Psychology*. New York, NY: Henry Holt and Company; 1890.
26. Gallagher S. Philosophical conceptions of the self: implications for cognitive science. *Trends Cogn Sci*. 2000;4:14–21. doi:10.1016/S1364-6613(99)01417-5
27. Christoff K, Cosmelli D, Legrand D, Thompson E. Specifying the self for cognitive neuroscience. *Trends Cogn Sci.* 2011;15:104–112.
28. Levy DJ, Glimcher PW. The root of all value: a neural common currency for choice. *Curr Opin Neurobiol.* 2012;22:1027–1038.
29. Mitchell JP, Macrae CN, Banaji MR. Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron.* 2006;50:655–663.
30. Tamir DI, Mitchell JP. Neural correlates of anchoring-and-adjustment during mentalizing. *Proc Natl Acad Sci U S A.* 2010;107:10827–10832.
31. Murray RJ, Schaer M, Debbané M. Degrees of separation: a quantitative neuroimaging meta-analysis investigating self-specificity and shared neural activation between self- and other-reflection. *Neurosci Biobehav Rev.* 2012;36(3):1043–1059.
32. Frith CD, Frith U. The neural basis of mentalizing. *Neuron.* 2006;50:531–534.
33. Gallagher HL, Frith CD. Functional imaging of 'theory of mind'. *Trends Cogn Sci.* 2003;7:77–83.
34. Zaki J, Weber J, Bolger N, Ochsner K. The neural bases of empathic accuracy. *Proc Natl Acad Sci U S A.* 2009;106:11382–11387.
35. Brass M, Haggard P. To do or not to do: the neural signature of self-control. *J Neurosci.* 2007;27:9141–9145.
36. Filevich E, Kühn S, Haggard P. Intentional inhibition in human action: the power of 'no'. *Neurosci Biobehav Rev.* 2012;36:1107–1118.
37. Kühn S, Haggard P, Brass M. Intentional inhibition: how the "veto-area" exerts control. *Hum Brain Mapp*. 2009;30:2834–2843. doi:10.1002/hbm.20711
38. Lynn MT, Muhle-Karbe PS, Brass M. Controlling the self: the role of the dorsal frontomedian cortex in intentional inhibition. *Neuropsychologia.* 2014;65:247–254.
39. Kabat-Zinn J. An outpatient program in behavioral medicine for chronic pain patients based on the practice of mindfulness meditation: theoretical considerations and preliminary results. *Gen Hosp Psychiatry.* 1982;4:33–47.
40. Sperduti M, Martinelli P, Piolino P. A neurocognitive model of meditation based on activation likelihood estimation (ALE) meta-analysis. *Conscious Cogn.* 2012;21:269–276.
41. Taylor VA, Daneault V, Grant J, et al. Impact of meditation training on the default mode network during a restful state. *Soc Cogn Affect Neurosci.* 2013;8:4–14.
42. Hölzel BK, Ott U, Hempel H, et al. Differential engagement of anterior cingulate and adjacent medial frontal cortex in adept meditators and non-meditators. *Neurosci Lett.* 2007;421:16–21.
43. Lutz J, Brühl AB, Doerig N, et al. Altered processing of self-related emotional stimuli in mindfulness meditators. *Neuroimage*. 2016;124:958–967. doi:10.1016/j. neuroimage.2015.09.057
44. Han S, Gu X, Mao L, Ge J, Wang G, Ma Y. Neural substrates of self-referential processing in Chinese Buddhists. *Soc Cogn Affect Neurosci*. 2009;5:332–339. doi:10.1093/scan/nsp027
45. Kean C. Silencing the self: schizophrenia as a self-disturbance. *Schizophr Bull.* 2009;35:1034–1036.
46. Fornito A, Yücel M, Patti J, Wood SJ, Pantelis C. Mapping grey matter reductions in schizophrenia: an anatomical likelihood estimation analysis of voxel-based morphometry studies. *Schizophr Res.* 2009;108:104–113.
47. Glahn DC, Laird AR, Ellison-Wright I, et al. Meta-analysis of gray matter anomalies in schizophrenia: application of anatomic likelihood estimation and network analysis. *Biol Psychiatry.* 2008;64:774–781.
48. Pomarol-Clotet E, Canales-Rodríguez EJ, Salvador R, et al. Medial prefrontal cortex pathology in schizophrenia as revealed by convergent findings from multimodal imaging. *Mol Psychiatry.* 2010;15:823–830.
49. Blackwood NJ, Bentall RP, Ffytche DH, Simmons A, Murray RM, Howard RJ. Persecutory delusions and the determination of self-relevance: an fMRI investigation. *Psychol Med.* 2004;34:591–596.
50. Holt DJ, Cassidy BS, Andrews-Hanna JR, et al. An anterior-to-posterior shift in midline cortical activity in schizophrenia during self-reflection. *Biol Psychiatry.* 2011;69:415–423.
51. Pankow A, Katthagen T, Diner S, et al. Aberrant salience is related to dysfunctional self-referential processing in psychosis. *Schizophr Bull.* 2016;42:67–76.
52. Shad MU, Brent BK, Keshavan MS. Neurobiology of self-awareness deficits in schizophrenia: a hypothetical model. *Asian J Psychiatr.* 2011;4:248–254.
53. Sugranyes G, Kyriakopoulos M, Corrigall R, Taylor E, Frangou S. Autism spectrum disorders and schizophrenia: meta-analysis of the neural correlates of social cognition. *PLoS One.* 2011;6:e25322.
54. Gusnard DA, Raichle ME, Raichle ME. Searching for a baseline: functional imaging and the resting human brain. *Nat Rev Neurosci.* 2001;2:685–694.
55. Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL. A default mode of brain function. *Proc Natl Acad Sci U S A.* 2001;98:676–682.
56. Andrews-Hanna JR, Reidler JS, Sepulcre J, Poulin R, Buckner RL. Functional-anatomic fractionation of the brain's default network. *Neuron.* 2010;65:550–562.
57. Andrews-Hanna JR, Smallwood J, Spreng RN. The default network and self-generated thought: component processes, dynamic control, and clinical relevance. *Ann N Y Acad Sci.* 2014;1316:29–52.

Originally published in: Schizophrenia Bulletin, 47, 2021, p. 1640

58. Kühn S, Gallinat J. Resting-state brain activity in schizophrenia and major depression: a quantitative meta-analysis. *Schizophr Bull*. 2013;39:358–365. doi:10.1093/ schbul/sbr151
59. Whitfield-Gabrieli S, Thermenos HW, Milanovic S, et al. Hyperactivity and hyperconnectivity of the default network in schizophrenia and in first-degree relatives of persons with schizophrenia. *Proc Natl Acad Sci U S A*. 2009;106:1279–1284.
60. Gruberger M, Levkovitz Y, Hendler T, et al. I think there-fore I am: rest-related prefrontal cortex neural activity is involved in generating the sense of self. *Conscious Cogn.* 2015;33:414–421.
61. Davey CG, Pujol J, Harrison BJ. Mapping the self in the brain's default mode network. *Neuroimage.* 2016;132:390–397.
62. Sajonz B, Kahnt T, Margulies DS, et al. Delineating self-referential processing from episodic memory retrieval: common and dissociable networks. *Neuroimage.* 2010;50:1606–1617.

Originally published in: Schizophrenia Bulletin, 47, 2021, p. 1641