Functional Brain Biomarkers of Self-Referential Bias in Remitted Depressed Outpatients

Liliana C. Wu (liliana.wu@mail.utoronto.ca)[1]; Jordan L. Livingston[1,2]; Zindel V. Segal[2]; Norman A.S. Farb[1,2]

[1]Department of Psychology, University of Toronto Mississauga
[2]Graduate Department of Psychological Clinical Science, University of Toronto Scarborough

Abstract

Negative self-referential bias (NSB) is a core feature of depression, yet its neural correlates and relevance for relapse remains unclear. This study examined whether behavioral and neural markers of NSB predict relapse following prophylactic psychotherapy. In a two-year prospective fMRI study embedded within a randomized trial of Mindfulness-Based Cognitive Therapy and Well-being-focused Cognitive Therapy, remitted depressed outpatients completed a self-referential encoding task before and after treatment and were followed longitudinally. NSB predicted greater depressive symptoms but not relapse. In contrast, greater frontal default mode network and salience network reactivity during negative self-referential processing predicted relapse. Whole-brain analysis identified elevated prefrontal activation and somatosensory deactivation as relapse biomarkers, with treatment-related attenuation of somatosensory deactivation predicting lower relapse risk. Somatosensory deactivation emerged as the strongest overall predictor of relapse. These findings suggest that relapse vulnerability reflects not only exaggerated prefrontal negative self-processing but also insufficient recruitment of somatosensory systems supporting embodied self-experience.

Keywords: Depression; fMRI; Self-Reference; Relapse; Sensory Deactivation; Negative Self

Introduction

Major Depressive Disorder (MDD) is among the most prevalent psychiatric conditions (Kessler et al., 2012) and a leading cause of disability worldwide (Herrman et al., 2022). MDD is highly recurrent; over half of individuals diagnosed with MDD experience multiple episodes, developing a more chronic course of illness (Burcusa & Iacono, 2007; Eaton et al., 2008; Monroe & Harkness, 2011).

Given its recurrent nature, elucidating the cognitive and neural mechanisms underlying depression vulnerability is critical. Vulnerability has long been linked to negative self-referential bias (NSB), whereby individuals engage in self-criticism, preferentially endorsing negative over positive traits (Brockmeyer et al., 2015; LeMoult et al., 2017; Pyszczynski et al., 1987; Watkins & Teasdale, 2004) —a bias associated with greater symptom severity (Bradley & Mathews, 1983; Gotlib et al., 2004; Timbremont & Braet, 2004).

At a neural level, the dorsal nexus hypothesis (Sheline et al., 2010) posits that excessive self-criticism and rumination in depression stem from resting-state hyperconnectivity of dorsal medial prefrontal cortex with three large-scale brain networks: the Default Mode Network (DMN), the Central Executive Network (CEN), and the Salience Network (SN). Critically, frontal DMN regions such as dmPFC support evaluative self-referential processing, the appraisal of self-relevant information for personal meaning and affective significance, which is functionally distinct from the autobiographical self-reference associated with posterior DMN regions such as precuneus(Buckner & Carroll, 2007; Lemogne et al., 2012). In parallel, SN regions including the anterior cingulate cortex contribute to assigning personal meaning to negative information and sustaining attention toward affectively relevant content (Burklund et al., 2014; Disner et al., 2011; Jiang, 2024). Frontal DMN and SN regions therefore represent the most theoretically motivated targets for paradigms probing self-critical evaluation.

Empirical evidence confirms prefrontal hyperconnectivity in active depression (Bertocci et al., 2023; Chou et al., 2023; Lemogne et al., 2010; Sabbah & Northoff, 2025; Sheline et al., 2009) and recurrent episodes (Bertocci et al., 2023; Dunlop et al., 2023; Hamilton et al., 2012). In addition to connectivity abnormalities, prior work has implicated altered task-related activation in frontal regions including medial prefrontal cortex, DLPFC, and anterior cingulate cortex during self-referential processing in depression, suggesting these regions also support dysphoric self-focus during task performance (Li et al., 2017; Lou et al., 2019; Nejad et al., 2013). Yet, the extent to which abnormalities across these systems support dysphoric self-referential bias that promotes episode return remains unclear.

Vulnerability may stem from more than aberrant prefrontal processing alone. Emerging research implicates abnormal sensorimotor activity as an additional candidate vulnerability factor (Farb et al., 2011, 2022; Ray et al., 2021; Seth & Friston, 2016). Sensorimotor systems may support a more embodied representation of present-moment experience (Farb et al., 2007). Consistent with this view, prior analysis from the current cohort implicated both interoceptive and sensory deactivation, including the somatosensory cortex, insula, and visual cortex, with future relapse vulnerability (Farb et al., 2022). Related functional connectivity findings have likewise highlighted sensorimotor dysfunction as a potential predictor of MDD recurrence (Ray et al., 2021).

The present study integrated indicators of NSB, prefrontal and sensorimotor processing, and the impact of prophylactic intervention in a single prospective study of MDD vulnerability. We operationalized NSB via patterns of both behavioral self-endorsement and neural (fMRI) activity—

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

searching for indicators of future depression vulnerability in remitted individuals undergoing prophylactic psychotherapy.

We hypothesized that greater NSB would predict future relapse (primary hypothesis), higher depressive symptoms, and lower decentering, a measure of psychological distancing (secondary hypothesis). We also hypothesized that relapse would be predicted by dysphoric self-referential brain activity, indexed by greater activation for negative than positive traits within frontal regions of the DMN and SN, as well as by stronger functional connectivity between these regions (primary hypothesis). Finally, we hypothesized that non-relapsers would exhibit treatment-related reductions in somatosensory inhibition during self-referential processing, a dynamic biomarker reflecting changes in relapse vulnerability (secondary hypothesis).

We conducted a prospective neuroimaging trial of prophylactic psychotherapy aimed at reducing depression relapse vulnerability**.** Participants underwent fMRI scanning before and after the intervention and were followed for two years**.** Self-referential processing was assessed using the Self-Referential Encoding Task (SRET) a validated paradigm (Derry & Kuiper, 1981) for measuring trait endorsement patterns and corresponding neural activations.

Methods

Fully remitted participants were randomized to receive eight group sessions of either Mindfulness Based Cognitive Therapy (MBCT) (Segal et al., 2012) or Cognitive Behavior Therapy with a Well-Being focus (WB-CT) (Fava, 2016) before entering a 2-year follow-up period. Participants completed fMRI and self-report assessments both before and after treatment. All provided informed consent under a protocol approved by the institutional review board at the Centre for Addiction and Mental Health (CAMH), and the study was pre-registered on the Open Science Framework (https://osf.io/s4n8j).

Participants

Eligibility criteria required that remitted depressed participants (1) did not currently meet a diagnosis of Major Depressive Disorder (MDD) according to DSM-IV criteria, (2) scored ≤12 on the Hamilton Depression Rating Scale (HRSD-17), (3) had ≥1 previous episode of MDD, (4) were between 18 and 65 years of age, (5) English speaking, and could provide informed consent. Excluded participants were (1) confirmed of a current diagnosis of Bipolar Disorder, Substance Abuse Disorder, Schizophrenia, or Borderline Personality Disorder, and (2) currently receiving psychotherapy or practicing meditation > once per week or yoga > twice per week.

The neuroimaging subsample did not differ from the parent trial sample on demographic or clinical variables (Farb et al., 2018; Segal et al., 2019). A final sample of 81 participants completed both baseline and post-intervention scans and with adequate behavioral task variance.

Randomization, Masking, and Intervention

Eligible patients were randomized in blocks of four, using computer generated quasi-random numbers, to receive eight weekly group sessions of either MBCT or WB-CT. Randomization was performed by the study coordinator via a randomization table; investigators were blind to group allocation during the intervention and follow-up assessment periods. Details of the prophylactic psychotherapy interventions are reported elsewhere (Farb et al., 2018)

Assessment and Outcomes

The primary outcome was time to relapse, defined as meeting DSM-IV criteria for a major depressive episode. Following treatment, participants entered a 2-year follow-up period during which relapse status was monitored every two months confirmed using SCID and HRSD-17 scores ≥16. Residual symptoms, lingering depressive symptoms after clinical remission, and decentering, reflecting psychological distance and perspective-taking (Segal et al., 2019), were also examined longitudinally across the follow-up interval.

Functional Magnetic Resonance Imaging Task

Participants completed the SRET during fMRI at baseline and post-treatment, consisting of two 80-trial runs of the self-referential bias task, interleaved with a dysphoric mood induction task (Farb, 2010; Farb et al., 2011, 2022). On each trial, participants viewed a cue (self, other, or case) paired with a trait word (positive or negative) and made a yes/no response for endorsement. In the self condition, participants judged whether the adjective described themselves; in the other condition, they judged whether the adjective described another person; and in the case condition, they made a non-self-referential perceptual judgment about the word (i.e., whether the word was presented in upper- or lowercase letters). Thus, the self and other conditions indexed evaluative processing, whereas the case condition served as a non-evaluative comparison. Unlike our prior report in this cohort, which focused on neural responses related to dysphoric mood induction, the present study focuses on self-referential encoding and its relation to relapse vulnerability. The present analyses focused on behavioral and neural responses during the SRET conditions. Each fMRI session included two blocks of 48 task trials and 32 fixation trials**.** fMRI acquisition and preprocessing parameters were performed using a standardized pipeline. Briefly, functional images were spatially smoothed with an 8 mm FWHM Gaussian kernel, and first-level SPM models applied a 180 s high-pass filter and included motion regressors, CSF signal, framewise displacement, and aCompCor components as nuisance covariates.

SRET endorsement patterns were analyzed to derive participant-level positive and negative self-endorsement rates. Endorsement rates for positive and negative words were separately calculated as the proportion of "yes" responses across all valid trials of the corresponding valence. NSB was quantified as the behavioral index ratio of negative to positive endorsement rates at baseline and post-

intervention. To mitigate the influence of outliers, self-referential bias scores were winsorized at the 5th and 95th percentiles.

Statistical Analysis

Positive and Negative Endorsement Neural Activity Whole-brain, voxel wise analyses were conducted on participants (N = 81) who completed both fMRI scanning at pre- and post-intervention with adequate behavioral task variance. At the first level**,** event-related general linear models in SPM12 modeled the self**,** other, and case conditions, each crossed with positive and negative valence. Standard motion parameters and global CSF signals were included as nuisance regressors (Birn, 2012; Kong et al., 2012). Contrast images for each Condition × Valence combination were generated for each participant and session. At the second level, these contrasts entered a mixed factorial model with within-subject factors of time (baseline, post-intervention)**,** condition (self, other, case)**,** and valence (positive, negative). Positive and negative endorsement rates were included as between-subject covariates to account for individual differences in endorsement bias. Exploratory conjunction analyses tested two opposing endorsement-related neural response profiles: (1) self-criticism: regions associated with decreases in positive endorsement and increases in negative endorsement, and (2) self-affirmation: regions associated with increases in positive endorsement and decreases in negative endorsement. Conjunction analyses were performed by thresholding each component contrast at p=√.005, yielding a conjoint significance level of p =.005, with cluster extent k=400. Modelling self-criticism and self-affirmation in separate covariate models (as opposed to a single model with both covariates) did not meaningfully alter the observed response pattern.

Associations of Self-Referential Bias and Depression Vulnerability Survival analyses were conducted to investigate whether NSB predicted relapse risk. Mixed-effects models examined associations between NSB, residual symptoms and decentering, adjusting for age, gender, past depressive episodes, and antidepressant medication (ADM). Additional linear mixed-effects models assessed whether baseline, and post-treatment NSB predicted residual symptoms and decentering over the follow-up period, controlling for the prior timepoint value, time, past episodes, age, and ADM. Variance Inflation Factor values for all covariates were below 5 and above 0.2, indicating no multicollinearity concerns.

Dysphoric Self-Referential Activity and Connectivity within frontal DMN and SN and Relapse Vulnerability Dysphoric self-referential activity was quantified using region of interest (ROI)-based extraction from a priori anatomical masks derived from the Harvard–Oxford cortical atlas. The frontal DMN was defined as a dorsomedial prefrontal cortex (dmPFC) mask created by combining the superior frontal gyrus and paracingulate gyrus. The frontal SN ROI was defined as the anterior cingulate cortex (ACC), using the anterior portion of the cingulate cortex from the Havard-Oxford atlas. For each participant and session, contrast images for Self-Negative and Self-Positive conditions were extracted and median activation values were computed within each ROI. The median was used as a robust summary measure to reduce the influence of extreme voxel values within the mask; this approach was preregistered. Dysphoric self-referential activity was indexed as the Self-Negative minus Self-Positive difference within each ROI, yielding separate frontal DMN and frontal SN metrics.

Functional connectivity between frontal DMN and SN was estimated using ROI-based correlations of denoised voxelwise BOLD time-series. For each participant and session, voxelwise time series were extracted from the anatomically defined dmPFC and ACC masks across the full task runs, mean-centered, band-pass filtered, and denoised by standard motion parameters and global CSF signal (Birn, 2012; Kong et al., 2012). Between-network and within-network connectivity were quantified as the mean Fisher z–transformed correlation between all frontal DMN and SN voxels.

Dysphoric self-referential activity and connectivity measures were entered into mixed-effects models comparing relapsers and non-relapsers, controlling for ADM, past episodes, residual symptoms, and their interaction with time. Cox proportional hazards models further tested whether frontal DMN and SN dysphoric self-referential activity predicted relapse, adjusting for baseline and post-treatment residual symptoms, ADM, and past episodes. Preregistration deviations are detailed on the Open Science Framework (https://osf.io/s4n8j). Briefly, preregistration deviations included replacing the endorsed-only contrast with the Self-Negative > Self-Positive contrast due to insufficient endorsed trials, using a priori anatomical ROIs rather than participant-specific functional-anatomical intersections, adopting the Self > Case contrast for later whole-brain analyses following sensitivity analyses, and not entering interoceptive dysfunction variables separately because they loaded onto the residual symptoms factor.

Static and Dynamic Neural Markers of Relapse Vulnerability during Self-referential processing Whole-brain static and dynamic neural markers were derived from the Self > Case contrast. First-level models estimated self, other, and case conditions collapsed across valence, and resulting contrasts were entered into second-level analyses to distinguish static vulnerability markers from dynamic treatment-related changes.

Static markers were identified from the Relapse > Non-Relapse and Non-Relapse > Relapse contrasts within a 2 (Time: baseline, post-intervention) × 2 (Relapse status: Relapse, Non-relapse) × 2 (Treatment: MBCT, WB-CT) mixed factorial model. Thus, “static” refers to stable between-group differences in self-referential processing between relapsers and non-relapsers across assessment timepoints. Dynamic markers were identified using a

conjunction analysis isolating regions showing consistent increases among non-relapsers across both time points (non-relapse growth) and regions where increases were greater in non-relapsers than relapsers at post-intervention session. Thus, "dynamic" refers to treatment-related changes in neural activity associated with non-relapse. Whole-brain results were thresholded at $p<.005$, cluster extent $k=400$, and FDR corrected at $p<.05$.

Parameter estimates from all significant ROIs were entered into mixed-effects models comparing relapsers and non-relapsers, while controlling for ADM, past episodes, residual symptoms, and their interaction with time. To identify optimal neural thresholds for relapse prediction and treatment response, receiver operating characteristic curve analyses were conducted using the *cutpointr* package(Thiele & Hirschfeld, 2021). Cox proportional hazards models tested whether static and dynamic markers, separately and jointly, predicted relapse, adjusting for baseline and post-treatment residual symptoms, ADM, and past episodes.

Results

Participants who relapsed reported higher residual depressive symptoms (d = .59) and fewer days well (d = -1.10) than non-relapsers, with no other demographic or clinical differences observed.

Self-Referential Patterns of Relapse Vulnerability

We first characterized behavioral and whole-brain neural correlates of self-referential bias at the task level. Figure 1. shows neural activity correlates of self-criticism (Panel A) and self-affirmation (Panel B). Self-criticism was associated with higher activations in a broader set of prefrontal and posterior regions, including orbitofrontal/insula, cerebellum, left dorsolateral prefrontal cortex, left posterior cingulate gyrus, and left lateral occipital cortex, whereas self-affirming responses were linked to posterior somatosensory regions. Higher residual symptoms were positively associated NSB (b = .10, 95% CI [.06, .14]), while both Time (b = -.07, 95% CI [-.12, -.01]) and Decentering (b = -.04, 95% CI [-.08, -.01]) were negatively associated with NSB. However, no significant associations between NSB and relapse status were observed.

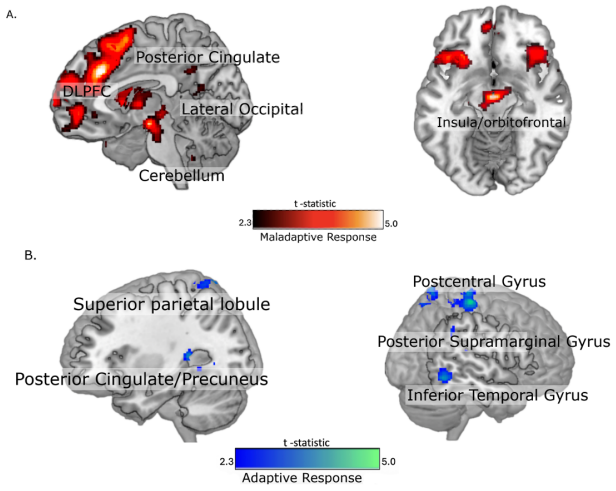

Figure 1: Neural activity correlates of self-affirmation and self-criticism response. Panel A shows self-criticism response clusters (red). Panel B shows self-affirmation response clusters (blue).

Frontal DMN–SN Dysphoric Self-Referential Activity and Connectivity as Markers of Relapse Vulnerability

Greater dysphoric self-referential activity within frontal DMN and SN regions was associated with relapse status (DMN: b = .32, 95% CI [.09, .55]; SN: b = .36, 95% CI [.09, .63]). Relapsers showed higher baseline dysphoric activity in both networks compared with non-relapsers (DMN: b = .32, 95% CI [.08, .56; SN: b = .36, 95% CI [.09, .63]). A Relapse × Time interaction was observed for frontal DMN activity only (b = -.32, 95% CI [-.64, -.01]). However, frontal DMN and salience network dysphoric activity did not independently predict relapse risk in survival analyses. Functional connectivity within or between DMN and SN was not associated with relapse status.

Static Markers of Relapse Vulnerability and Dynamic Markers of Treatment Response

Figure 2 illustrates whole-brain differences in self-referential activity by relapse status. Relapsers exhibited greater activation in the left subgenual cingulate (b = .74, 95% CI [.28, 1.20]) and right lateral occipital cortex (b = .61, 95% CI [.28, .95]), as well as greater deactivation in the right postcentral gyrus (b = -.34, 95% CI [-.57, -.11]), compared with non-relapsers. Survival analyses indicated increased relapse risk associated with elevated subgenual cingulate activity (HR = 1.81, 95% CI[1.20, 2.72]), elevated lateral occipital activity (HR = 2.07, 95% CI[1.12, 3.84]), and greater postcentral deactivation (HR = .26, 95% CI[.12,.58]).

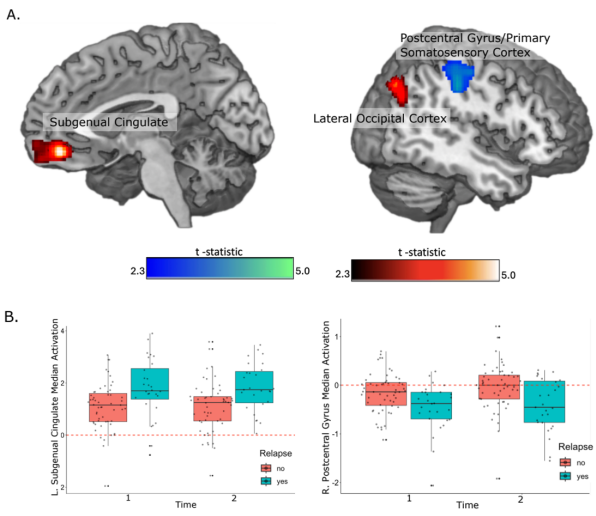

Figure 2: Static neural markers of relapse vulnerability during self-referential processing. Panel A shows group-level activation maps by relapse status with higher in red, and lower activation in blue. Panel B presents mean activation within each region of interest across time.

Dynamic analyses revealed that non-relapsers showed attenuated reductions over time in somatosensory and associative cortices relative to relapsers, indicating preserved engagement of these regions following treatment. Although longitudinal changes were observed, no significant treatment main effects, treatment-by-time effects, or post-intervention treatment effects within the non-relapse group were detected, suggesting that these longitudinal neural changes were observed across the intervention context.

A combined survival model incorporated both static and dynamic ROIs, adjusting for ADM, past episodes, and baseline and post-treatment residual symptoms (Figure 3). Higher mean activation in the right postcentral gyrus (HR = .16, 95% CI[.05, .52]) and greater increases over time in the left supramarginal gyrus (HR = .12, 95% CI[.03, .51]) were associated with reduced relapse risk.

Discussion

We conducted a prospective neuroimaging investigation of prophylactic psychotherapy, aimed at clarifying how self-referential processing contributes to relapse vulnerability. Greater NSB was associated with greater symptom burden and lower decentering both at post-treatment and across the two-year follow-up. Notably, endorsement patterns did not directly predict future depressive episodes.

Self-criticism, which consisted of affirming negative traits and denying positive traits, engaged a broader set of prefrontal and posterior regions in the whole-brain analysis, including medial prefrontal, orbitofrontal/insula, posterior cingulate, and occipital regions, consistent with models linking depression to exaggerated self-focused and affectively salient processing (Lemogne et al., 2009, 2010, 2012). This pattern included left DLPFC activation, a core node of the CEN, providing exploratory support for the third arm of the dorsal nexus hypothesis. DLPFC recruitment during negative self-referential processing may reflect increased cognitive control demands or compensatory regulatory effort in the face of dysphoric self-focus, consistent with accounts linking prefrontal control circuitry to the maintenance of negative attentional bias in depression (De Raedt & Koster, 2010; Johnstone et al., 2007). Within preregistered frontal DMN and SN regions, dysphoric self-referential activity was elevated among relapsers compared with non-relapsers. However, these specific ROI measures did not independently predict time to relapse in survival analyses, suggesting that frontal dysphoric self-referential activity may reflect relapse-related group differences without robustly forecasting episode timing on its own.

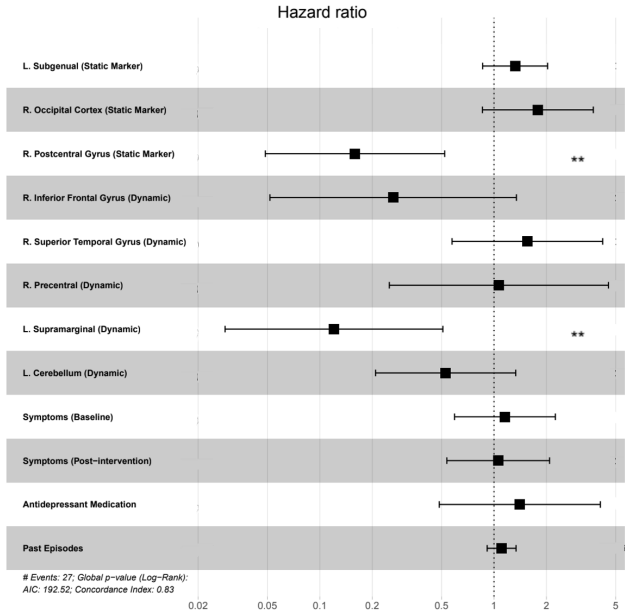

Figure 3: Integrated Cox Model of Static and Dynamic Neural Relapse Markers. Hazard ratios are shown from the combined Cox regression model predicting relapse, incorporating both static neural markers and dynamic markers reflecting change over the intervention period.

The divergence between behavioral NSB and frontal brain activity, where only frontal brain activity distinguished relapse status, may reflect the operation of partially distinct mechanisms underlying negative self-referential processing. Attentional bias toward negative self-relevant information, indexed behaviorally by endorsement rates, may primarily reflect a conscious, reportable tendency that tracks concurrent symptom burden without necessarily driving episode recurrence. By contrast, the neural activity patterns that predicted relapse, particularly sustained prefrontal recruitment during self-referential processing, may more closely reflect internal elaboration, an automatic

amplification and maintenance of negative self-relevant content that contributes more directly to relapse risk (Disner et al., 2011; Watkins & Teasdale, 2004). These distinctions suggest that behavioral and neural indicators of self-referential bias are best understood as complementary rather than redundant markers of depression vulnerability.

Whole brain analyses further characterized relapse vulnerability, specifically implicating the left subgenual cingulate, a region consistently implicated in persistent depressive risk and targeted in therapeutic interventions (Benschop et al., 2022; Drevets et al., 2008; Mayberg et al., 2005). Relapsers also showed greater activation in right lateral occipital cortex, a region more recently implicated in MDD as a site of perceptual disruption (Li et al., 2025; Wu et al., 2023), and whose baseline reactivity additionally predicted worsening residual symptoms across treatment-consistent with evidence that lateral occipital cortex hyperactivation reflects difficulty disengaging from negative emotional content in depression (Colich et al., 2017). Together, these findings align with attentional accounts of depression, where excessive engagement with negative schema-relevant information is maintained by prefrontal and posterior perceptual circuitry (De Raedt & Koster, 2010; Johnstone et al., 2007; Pizzagalli & Roberts, 2022).

By contrast, deficits in self-affirmation were associated with aberrant somatosensory processing. Greater positive and lower negative trait endorsement recruited the posterior somatosensory cortex, and relapsers consistently showed greater deactivations in this region, with treatment-related attenuation of these deactivations in somatosensory-parietal regions characterizing the non-relapse trajectory. Reduced somatosensory activity during self-referential processing may reflect diminished access to bodily and sensory cues that ground self-experience in the present moment (Farb et al., 2010, 2015), shifting self-evaluation toward more enduring, narrative forms of self-awareness that support rigid rather than flexible updating of self-relevant information (Farb et al., 2007, 2015).

This interpretation aligns with broader theoretical accounts proposing that self-representation is grounded in embodied sensory experience and ongoing bodily signals, such that accessible interoceptive representations support moment-to-moment adaptation while reduced access contributes to dysregulated self-processing(Paulus & Stein, 2010; Seth, 2013; Seth & Friston, 2016). Decentering, a construct linked to psychological distance and perspective taking (Segal et al., 2019), was negatively associated with depressive symptoms and self-critical endorsement but positively associated with self-affirmation, suggesting that greater somatosensory engagement over treatment may be expressed psychologically as an increased capacity to relate to self-relevant experience as contextual and changing, rather than as fixed evidence of negative self-worth.

Our combined survival model suggests that somatosensory markers explained unique variance in relapse risk when static and longitudinal treatment-related neural markers were considered together. Higher right postcentral activation and greater increases in left supramarginal activation predicted *lower* relapse likelihood, even after adjusting for clinical covariates. The left supramarginal gyrus may be especially relevant as a dynamic marker because treatment-related attenuated deactivations were associated with lower relapse risk. This region has been implicated in multisensory integration and somatosensory feedback (Isernia et al., 2023; Proske & Gandevia, 2012), and has also been implicated in emotion processing and regulation in psychiatric populations (Gong et al., 2020; Madeira et al., 2020). Accordingly, increased supramarginal engagement over treatment may reflect improved multisensory and embodied processing that supports self-affirmation and reduces vulnerability to relapse. In contrast, prefrontal markers, although predictive in isolation, no longer contributed significantly when somatosensory markers were included, suggesting that somatosensory markers may capture unique protective variance in this sample that is not fully accounted for by prefrontal indices alone.

These findings have important clinical implications. Much of the treatment literature focuses on reducing prefrontal engagement with negative content, challenging self-critical thoughts, reducing rumination, or modulating subgenual activity. Our results suggest that strengthening access to somatosensory representations may be at least as relevant for relapse prevention. Both WB-CT and MBCT emphasize updating negative beliefs and cultivating embodied awareness (Fava, 2016; Segal et al., 2012), and the neural pattern observed here, stronger somatosensory engagement predicting reduced relapse risk, provides a mechanistic account for why embodied components of these interventions may be particularly relevant for long-term protection against recurrence.

Limitations

Several limitations should be noted. The absence of a healthy control group limits conclusions about disorder specificity. The sample was demographically homogeneous, restricting generalizability. Finally, psychiatric and medical comorbidities were not systematically assessed.

Conclusions

This study challenges the traditional prefrontal-centric view of depression vulnerability by demonstrating that relapse risk may depend as much on insufficient recruitment of somatosensory systems as on heightened negative evaluative processing. Interventions that enhance embodied awareness, sensory integration, and bottom-up updating may therefore represent promising targets for future treatment optimization.

Acknowledgements

This study was funded by the Canadian Institute of Health Research (Grant# 243812).

References

Benschop, L., Vanhollebeke, G., Li, J., Leahy, R. M., Vanderhasselt, M.-A., & Baeken, C. (2022). Reduced subgenual cingulate–dorsolateral prefrontal connectivity as an electrophysiological marker for depression. *Scientific Reports*, *12*(1), 16903. https://doi.org/10.1038/s41598-022-20274-9

Bertocci, M. A., Afriyie-Agyemang, Y., Rozovsky, R., Iyengar, S., Stiffler, R., Aslam, H. A., Bebko, G., & Phillips, M. L. (2023). Altered patterns of central executive, default mode and salience network activity and connectivity are associated with current and future depression risk in two independent young adult samples. *Molecular Psychiatry*, *28*(3), 1046–1056. https://doi.org/10.1038/s41380-022-01899-8

Birn, R. M. (2012). The role of physiological noise in resting-state functional connectivity. *NeuroImage*, *62*(2), 864–870. https://doi.org/10.1016/j.neuroimage.2012.01.016

Bradley, B., & Mathews, A. (1983). Negative self-schemata in clinical depression. *The British Journal of Clinical Psychology*, *22 (Pt 3)*, 173–181. https://doi.org/10.1111/j.2044-8260.1983.tb00598.x

Brockmeyer, T., Zimmermann, J., Kulessa, D., Hautzinger, M., Bents, H., Friederich, H.-C., Herzog, W., & Backenstrass, M. (2015). Me, myself, and I: Self-referent word use as an indicator of self-focused attention in relation to depression and anxiety. *Frontiers in Psychology*, *6*, 1564. https://doi.org/10.3389/fpsyg.2015.01564

Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *Trends in Cognitive Sciences*, *11*(2), 49–57. https://doi.org/10.1016/j.tics.2006.11.004

Burcusa, S. L., & Iacono, W. G. (2007). Risk for Recurrence in Depression. *Clinical Psychology Review*, *27*(8), 959–985. https://doi.org/10.1016/j.cpr.2007.02.005

Burklund, L. J., Creswell, J. D., Irwin, M. R., & Lieberman, M. D. (2014). The common and distinct neural bases of affect labeling and reappraisal in healthy adults. *Frontiers in Psychology*, *5*, 221. https://doi.org/10.3389/fpsyg.2014.00221

Chou, T., Deckersbach, T., Dougherty, D. D., & Hooley, J. M. (2023). The default mode network and rumination in individuals at risk for depression. *Social Cognitive and Affective Neuroscience*, *18*(1), nsad032. https://doi.org/10.1093/scan/nsad032

Colich, N. L., Ho, T. C., Foland-Ross, L. C., Eggleston, C., Ordaz, S. J., Singh, M. K., & Gotlib, I. H. (2017). Hyperactivation in Cognitive Control and Visual Attention Brain Regions During Emotional Interference in Adolescent Depression. *Biological Psychiatry. Cognitive Neuroscience and Neuroimaging*, *2*(5), 388–395. https://doi.org/10.1016/j.bpsc.2016.09.001

De Raedt, R., & Koster, E. H. W. (2010). Understanding vulnerability for depression from a cognitive neuroscience perspective: A reappraisal of attentional factors and a new conceptual framework. *Cognitive, Affective & Behavioral Neuroscience*, *10*(1), 50–70. https://doi.org/10.3758/CABN.10.1.50

Derry, P. A., & Kuiper, N. A. (1981). Schematic processing and self-reference in clinical depression. *Journal of Abnormal Psychology*, *90*(4), 286–297. https://doi.org/10.1037//0021-843x.90.4.286

Disner, S. G., Beevers, C. G., Haigh, E. A. P., & Beck, A. T. (2011). Neural mechanisms of the cognitive model of depression. *Nature Reviews. Neuroscience*, *12*(8), 467–477. https://doi.org/10.1038/nrn3027

Drevets, W. C., Savitz, J., & Trimble, M. (2008). The Subgenual Anterior Cingulate Cortex in Mood Disorders. *CNS Spectrums*, *13*(8), 663–681. https://doi.org/10.1017/s1092852900013754

Dunlop, B. W., Cha, J., Choi, K. S., Nemeroff, C. B., Craighead, W. E., & Mayberg, H. S. (2023). Functional connectivity of salience and affective networks among remitted depressed patients predicts episode recurrence. *Neuropsychopharmacology*, *48*(13), 1901–1909. https://doi.org/10.1038/s41386-023-01653-w

Eaton, W. W., Shao, H., Nestadt, G., Lee, B. H., Bienvenu, O. J., & Zandi, P. (2008). Population-Based Study of First Onset and Chronicity in Major Depressive Disorder. *Archives of General Psychiatry*, *65*(5), 513–520. https://doi.org/10.1001/archpsyc.65.5.513

Farb, N. (2010). Minding One's Emotions: Mindfulness Training Alters the Neural Expression of Sadness. *Emotion*. https://doi.org/10.1037/a0017151.supp

Farb, N. A. S., Anderson, A. K., Bloch, R. T., & Segal, Z. V. (2011). Mood-Linked Responses in Medial Prefrontal Cortex Predict Relapse in Patients with Recurrent Unipolar Depression. *Biological Psychiatry*, *70*(4), 366–372. https://doi.org/10.1016/j.biopsych.2011.03.009

Farb, N. A. S., Anderson, A. K., Mayberg, H., Bean, J., McKeon, D., & Segal, Z. V. (2010). Minding one's emotions: Mindfulness training alters the neural expression of sadness. *Emotion*, *10*(1), 25–33. https://doi.org/10.1037/a0017151

Farb, N. A. S., Desormeau, P., Anderson, A. K., & Segal, Z. V. (2022). Static and treatment-responsive brain biomarkers of depression relapse vulnerability following prophylactic psychotherapy: Evidence from a randomized control trial. *NeuroImage : Clinical*, *34*, 102969. https://doi.org/10.1016/j.nicl.2022.102969

Farb, N. A. S., Segal, Z. V., Mayberg, H., Bean, J., McKeon, D., Fatima, Z., & Anderson, A. K. (2007). Attending to the present: Mindfulness meditation reveals distinct neural modes of self-reference. *Social Cognitive and Affective Neuroscience*, *2*(4), 313–322. https://doi.org/10.1093/scan/nsm030

Farb, N., Anderson, A., Ravindran, A., Hawley, L., Irving, J., Mancuso, E., Gulamani, T., Williams, G., Ferguson, A., & Segal, Z. V. (2018). Prevention of relapse/recurrence in major depressive disorder with either mindfulness-based cognitive therapy or cognitive therapy. *Journal of Consulting and Clinical Psychology*, *86*(2), 200–204. https://doi.org/10.1037/ccp0000266

Farb, N., Daubenmier, J., Price, C. J., Gard, T., Kerr, C., Dunn, B. D., Klein, A. C., Paulus, M. P., & Mehling, W. E.

(2015). Interoception, contemplative practice, and health. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00763

Fava, G. A. (2016). Well-Being Therapy: Current Indications and Emerging Perspectives. *Psychotherapy and Psychosomatics*, *85*(3), 136–145. https://doi.org/10.1159/000444114

Ford, C. G., Kiken, L. G., Haliwa, I., & Shook, N. J. (2023). Negatively biased cognition as a mechanism of mindfulness: A review of the literature. *Current Psychology*, *42*(11), 8946–8962. https://doi.org/10.1007/s12144-021-02147-y

Gong, J., Wang, J., Qiu, S., Chen, P., Luo, Z., Wang, J., Huang, L., & Wang, Y. (2020). Common and distinct patterns of intrinsic brain activity alterations in major depression and bipolar disorder: Voxel-based meta-analysis. *Translational Psychiatry*, *10*(1), 353. https://doi.org/10.1038/s41398-020-01036-5

Gotlib, I. H., Kasch, K. L., Traill, S., Joormann, J., Arnow, B. A., & Johnson, S. L. (2004). Coherence and specificity of information-processing biases in depression and social phobia. *Journal of Abnormal Psychology*, *113*(3), 386–398. https://doi.org/10.1037/0021-843X.113.3.386

Hamilton, J. P., Etkin, A., Furman, D. J., Lemus, M. G., Johnson, R. F., & Gotlib, I. H. (2012). Functional Neuroimaging of Major Depressive Disorder: A Meta-Analysis and New Integration of Baseline Activation and Neural Response Data. *American Journal of Psychiatry*, *169*(7), 693–703. https://doi.org/10.1176/appi.ajp.2012.11071105

Herrman, H., Patel, V., Kieling, C., Berk, M., Buchweitz, C., Cuijpers, P., Furukawa, T. A., Kessler, R. C., Kohrt, B. A., Maj, M., McGorry, P., Reynolds, C. F., Weissman, M. M., Chibanda, D., Dowrick, C., Howard, L. M., Hoven, C. W., Knapp, M., Mayberg, H. S., … Wolpert, M. (2022). Time for united action on depression: A Lancet-World Psychiatric Association Commission. *Lancet (London, England)*, *399*(10328), 957–1022. https://doi.org/10.1016/S0140-6736(21)02141-3

Isernia, S., Blasi, V., Baglio, G., Cabinio, M., Cecconi, P., Rossetto, F., Cazzoli, M., Blasi, F., Bruckmann, C., Giunco, F., Sorbi, S., Clerici, M., & Baglio, F. (2023). The key role of depression and supramarginal gyrus in frailty: A cross-sectional study. *Frontiers in Aging Neuroscience*, *15*. https://doi.org/10.3389/fnagi.2023.1292417

Jiang, Y. (2024). A theory of the neural mechanisms underlying negative cognitive bias in major depression. *Frontiers in Psychiatry*, *15*, 1348474. https://doi.org/10.3389/fpsyt.2024.1348474

Johnstone, T., Reekum, C. M. van, Urry, H. L., Kalin, N. H., & Davidson, R. J. (2007). Failure to Regulate: Counterproductive Recruitment of Top-Down Prefrontal-Subcortical Circuitry in Major Depression. *Journal of Neuroscience*, *27*(33), 8877–8884. https://doi.org/10.1523/JNEUROSCI.2063-07.2007

Kerr, C. E., Sacchet, M. D., Lazar, S. W., Moore, C. I., & Jones, S. R. (2013). Mindfulness starts with the body: Somatosensory attention and top-down modulation of cortical alpha rhythms in mindfulness meditation. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00012

Kessler, R. C., Petukhova, M., Sampson, N. A., Zaslavsky, A. M., & Wittchen, H.-U. (2012). Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States. *International Journal of Methods in Psychiatric Research*, *21*(3), 169–184. https://doi.org/10.1002/mpr.1359

Kong, Y., Jenkinson, M., Andersson, J., Tracey, I., & Brooks, J. C. W. (2012). Assessment of physiological noise modelling methods for functional imaging of the spinal cord. *NeuroImage*, *60*(2), 1538–1549. https://doi.org/10.1016/j.neuroimage.2011.11.077

Lemogne, C., Delaveau, P., Freton, M., Guionnet, S., & Fossati, P. (2012). Medial prefrontal cortex and the self in major depression. *Journal of Affective Disorders*, *136*(1–2), e1–e11. https://doi.org/10.1016/j.jad.2010.11.034

Lemogne, C., le Bastard, G., Mayberg, H., Volle, E., Bergouignan, L., Lehéricy, S., Allilaire, J.-F., & Fossati, P. (2009). In search of the depressive self: Extended medial prefrontal network during self-referential processing in major depression. *Social Cognitive and Affective Neuroscience*, *4*(3), 305–312. https://doi.org/10.1093/scan/nsp008

Lemogne, C., Mayberg, H., Bergouignan, L., Volle, E., Delaveau, P., Lehéricy, S., Allilaire, J.-F., & Fossati, P. (2010). Self-referential processing and the prefrontal cortex over the course of depression: A pilot study. *Journal of Affective Disorders*, *124*(1), 196–201. https://doi.org/10.1016/j.jad.2009.11.003

LeMoult, J., Kircanski, K., Prasad, G., & Gotlib, I. H. (2017). Negative Self-Referential Processing Predicts the Recurrence of Major Depressive Episodes. *Clinical Psychological Science*, *5*(1), 174–181. https://doi.org/10.1177/2167702616654898

Li, J., Tan, Y., Zheng, Z., Feng, C., Fang, W., Huang, X., Lin, S., So, K.-F., Huang, L., Ren, C., & Tao, Q. (2025). Reduced neural suppression at occipital cortex in subthreshold depression. *Translational Psychiatry*, *15*(1), 220. https://doi.org/10.1038/s41398-025-03446-9

Li, Y., Li, M., Wei, D., Kong, X., Du, X., Hou, X., Sun, J., & Qiu, J. (2017). Self-referential processing in unipolar depression: Distinct roles of subregions of the medial prefrontal cortex. *Psychiatry Research: Neuroimaging*, *263*, 8–14. https://doi.org/10.1016/j.pscychresns.2017.02.008

Lou, Y., Lei, Y., Mei, Y., Leppänen, P. H. T., & Li, H. (2019). Review of Abnormal Self-Knowledge in Major Depressive Disorder. *Frontiers in Psychiatry*, *10*. https://doi.org/10.3389/fpsyt.2019.00130

Madeira, N., Duarte, J. V., Martins, R., Costa, G. N., Macedo, A., & Castelo-Branco, M. (2020). Morphometry and gyrification in bipolar disorder and schizophrenia: A comparative MRI study. *NeuroImage. Clinical*, *26*, 102220. https://doi.org/10.1016/j.nicl.2020.102220

Mayberg, H. S., Lozano, A. M., Voon, V., McNeely, H. E., Seminowicz, D., Hamani, C., Schwalb, J. M., & Kennedy, S. H. (2005). Deep Brain Stimulation for Treatment-Resistant Depression. *Neuron*, *45*(5), 651–660. https://doi.org/10.1016/j.neuron.2005.02.014

Monroe, S. M., & Harkness, K. L. (2011). Recurrence in major depression: A conceptual analysis. *Psychological Review*, *118*(4), 655–674. https://doi.org/10.1037/a0025190

Nejad, A. B., Fossati, P., & Lemogne, C. (2013). Self-Referential Processing, Rumination, and Cortical Midline Structures in Major Depression. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00666

Paulus, M. P., & Stein, M. B. (2010). Interoception in anxiety and depression. *Brain Structure & Function*, *214*(5), 451–463. https://doi.org/10.1007/s00429-010-0258-9

Pizzagalli, D. A., & Roberts, A. C. (2022). Prefrontal cortex and depression. *Neuropsychopharmacology*, *47*(1), 225–246. https://doi.org/10.1038/s41386-021-01101-7

Proske, U., & Gandevia, S. C. (2012). The proprioceptive senses: Their roles in signaling body shape, body position and movement, and muscle force. *Physiological Reviews*, *92*(4), 1651–1697. https://doi.org/10.1152/physrev.00048.2011

Pyszczynski, T., Holt, K., & Greenberg, J. (1987). Depression, self-focused attention, and expectancies for positive and negative future life events for self and others. *Journal of Personality and Social Psychology*, *52*(5), 994–1001. https://doi.org/10.1037//0022-3514.52.5.994

Ray, D., Bezmaternykh, D., Mel'nikov, M., Friston, K. J., & Das, M. (2021). Altered effective connectivity in sensorimotor cortices is a signature of severity and clinical course in depression. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(40), e2105730118. https://doi.org/10.1073/pnas.2105730118

Sabbah, S. G., & Northoff, G. (2025). The self in depression and anxiety as a transdiagnostic and differential-diagnostic neural marker: A systematic review. *Neuroscience & Biobehavioral Reviews*, *169*, 106034. https://doi.org/10.1016/j.neubiorev.2025.106034

Segal, Z. V., Anderson, A. K., Gulamani, T., Williams, L. A. D., Desormeau, P., Ferguson, A., Walsh, K., & Farb, N. A. S. (2019). Practice of therapy acquired regulatory skills and depressive relapse/ recurrence prophylaxis following cognitive therapy or mindfulness based cognitive therapy. *Journal of Consulting and Clinical Psychology*, *87*(2), 161–170. https://doi.org/10.1037/ccp0000351.supp

Segal, Z., Williams, M., & Teasdale, J. (2012). *Mindfulness-Based Cognitive Therapy for Depression*. Guilford Press.

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, *17*(11), 565–573. https://doi.org/10.1016/j.tics.2013.09.007

Seth, A. K., & Friston, K. J. (2016). Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1708), 20160007. https://doi.org/10.1098/rstb.2016.0007

Sheline, Y. I., Barch, D. M., Price, J. L., Rundle, M. M., Vaishnavi, S. N., Snyder, A. Z., Mintun, M. A., Wang, S., Coalson, R. S., & Raichle, M. E. (2009). The default mode network and self-referential processes in depression. *Proceedings of the National Academy of Sciences*, *106*(6), 1942–1947. https://doi.org/10.1073/pnas.0812686106

Sheline, Y. I., Price, J. L., Yan, Z., & Mintun, M. A. (2010). Resting-state functional MRI in depression unmasks increased connectivity between networks via the dorsal nexus. *Proceedings of the National Academy of Sciences*, *107*(24), 11020–11025. https://doi.org/10.1073/pnas.1000446107

Thiele, C., & Hirschfeld, G. (2021). cutpointr: Improved Estimation and Validation of Optimal Cutpoints in R. *Journal of Statistical Software*, *98*, 1–27. https://doi.org/10.18637/jss.v098.i11

Timbremont, B., & Braet, C. (2004). Cognitive vulnerability in remitted depressed children and adolescents. *Behaviour Research and Therapy*, *42*(4), 423–437. https://doi.org/10.1016/S0005-7967(03)00151-7

Watkins, E., & Teasdale, J. D. (2004). Adaptive and maladaptive self-focus in depression. *Journal of Affective Disorders*, *82*(1), 1–8. https://doi.org/10.1016/j.jad.2003.10.006

Wu, F., Lu, Q., Kong, Y., & Zhang, Z. (2023). A Comprehensive Overview of the Role of Visual Cortex Malfunction in Depressive Disorders: Opportunities and Challenges. *Neuroscience Bulletin*, *39*(9), 1426–1438. https://doi.org/10.1007/s12264-023-01052-7