A mind full of self: Self-referential processing as a mechanism underlying the therapeutic effects of mindfulness training on internalizing disorders

Yanli Lin*, Courtney P. Callahan, Jason S. Moser
Department of Psychology, Michigan State University, United States

ARTICLE INFO
Keywords:
Mindfulness
Self-Referential processing
Internalizing disorders

ABSTRACT
The aim of the current review is to advance the hypothesis that change in self-referential processing is a key but under-examined mechanism through which mindfulness training confers its therapeutic benefits for individuals with internalizing disorders. Consequently, we integrated neuroscientific studies on aberrant self-referential processing in internalizing disorders with contemplative science scholarship examining the effects of mindfulness training on the self-referential system. Reviewing these literatures yielded four major conclusions: (1) internalizing disorders can be characterized by excessive self-referential processing and emotion dysregulation; (2) mindfulness training has moderate effects on reducing internalizing symptoms; (3) mindfulness training promotes the shifting from narrative self-focus to present-centered experiential awareness; (4) such mindfulness-induced changes in self-reference is accompanied by reduced activation in overactive self-referential brain regions that have been implicated in internalizing disorders. Clinical and research implications related to delineating the role of self-referential processing in producing the therapeutic effects of mindfulness training are discussed.

1. Introduction

Understanding the self has been one of the most pertinent endeavors in human history. Various religious, philosophical and scientific disciplines have sought to address the “problem of the self”. Indeed, there is substantial richness and diversity for explaining the self, ranging from the Buddhist denial of self to the Cartesian embrace of the self as a distinct entity of subjectivity. In addition to debating the existence of self, history is rife with varying perspectives as to what constitutes the self. For example, self has been construed to reflect a basic sense of agency comprised of what is accessible to immediate self-consciousness—what has been termed the “minimal self” (Gallagher, 2000) or “core self” (Damasio, 1999). Yet, other concepts such as the “narrative self” (Gallagher, 2000; Gallagher and Frith, 2003) point to a more complex experience involving memories, intentions, and social awareness, highlighting the expansive and seemingly intractable nature of the construct.

Recent efforts in neuroscience have sought to develop a systematic understanding of self by identifying overlapping processes that unify the different conceptions of self—grouping these processes under the umbrella term “self-referential” processing, also referred to in the literature as “self-related” or “self-relevant” processing (Craik and Hay, 1999; Kelley et al., 2002; Schore, 2003). The neuroscientific perspective assumes that self-referential processing, defined as the experience of strongly relating to one’s own person, is a core feature common to the various concepts of self (Northoff et al., 2006). Here, “experience” refers to the phenomenal aspect (e.g., the subjective feeling of me) that is prerelative (Legrand, 2007) and not involving reflection, logic, or objective reasoning. Consequently, self-referential processing is presumed to be an essential and fundamental function necessary for discriminating between self and other (Murray et al., 2012), and other socially adaptive forms of higher order cognition (Murray et al., 2015).

Advances in cognitive neuroscience have provided unique opportunities to investigate the neural correlates of self-referential processing. Many neuroimaging studies have found that self-referential tasks involving verbal, memory, emotional, and social stimuli reliably elicit neural activity in the cortical midline structures (CMS; Northoff et al., 2006). Moreover, a growing body of literature has also illuminated an overlap between self and resting state brain activity (Qin and Northoff, 2011; Murray et al., 2015), also known as the default-mode network (DMN)—showing that self-referential processes are not necessarily task dependent but can occur during periods of inactivity. Taken together, these findings provide support for the construct of self-referential processing—organizing the complexity and abstraction associated with the different conceptions of self into a tractable set of observable and predictable neural processes. Indeed, the work to identify the neural
correlates of self-referential processing is notable in that it seeks to bridge what is inherently a subjective experience with observable measurement. Unsurprisingly, the emergence of this approach has had significant impacts on related disciplines.

Clinical science, in particular, has leveraged the neuroscience of self-referential processing to examine the relationship between self and psychopathology in novel ways. Specifically, various theoretical orientations in clinical psychology have implicated the self in understanding internalizing disorders—a descriptive label uniting unipolar mood (i.e., depression) and anxiety disorders based on the shared propensity to express distress inwards (Kotov et al., 2017; Wright et al., 2013). One prevailing perspective is that excessive self-focus, characterized by attentional engagement in self-related processes, exacerbates internalizing symptoms (Pyszczynski and Greenberg, 1987; Ingram, 1990; Mor and Winquist, 2002). Because different forms of self-focus (e.g., rumination) inherently involve self-referential processing, research has examined the neural correlates of self-referential processing in patients with depression and anxiety disorders. The results of these investigations have yielded aberrant activity in self-referential brain regions, supporting the general postulation that excessive self-focus plays an integral role in the development and maintenance of internalizing symptoms (Lemogne et al., 2009; Nejad et al., 2013). Clinically, measuring neural activity may yield valuable insights regarding the degree to which intervention outcomes are driven by changes in self-referential processing, possibly leading to the discovery of novel target mechanisms.

One such intervention is mindfulness training, a practice that is central to the emergence of mindfulness-based therapies (MBTs; Baer et al., 2006). Originating from a 2500-year-old Buddhist contemplative tradition, mindfulness training involves the development of present-moment, nonjudgmental awareness through repeated sustenance of selective attention on internal experiences (see Vago and Silbersweig, 2012 for a review). Given the breadth of the current review, we define mindfulness training as any activity that promotes the cultivation of experiential awareness of present-moment sensory phenomena (i.e., mindful awareness). Consequently, our operationalization of mindfulness training spans multiple “training” modalities, ranging from self-directed contemplative practices (e.g., sitting and walking mindfulness meditation) to guided therapeutic exercises (e.g., body scan) offered within the broader scope of MBTs.

Driven by its therapeutic properties, both the training and associated principles of mindfulness have been integrated in a variety of efficacious psychotherapeutic interventions for internalizing disorders (Baer et al., 2006; Hoffman et al., 2008). Although much work has uncovered the therapeutic potential of mindfulness training, little is known about its underlying neural mechanisms. In Buddhism, mindfulness training served as a means to liberate the practitioner from the suffering associated with the fundamental misperception of the self as a fixed, permanent entity (Anālayo, 2003).

Interestingly, recent neuroscience investigations have yielded findings that are in line with Buddhist teachings, linking mindfulness training with significant changes in self-referential processing (Farb et al., 2007; Brewer et al., 2011; Berkovich-Ohana et al., 2012). Although several prominent conceptual models of mindfulness have referenced the effects of mindfulness training on self-referential processing (Grabovac et al., 2011; Hözel et al., 2011; Lutz et al., 2015; Vago and Silbersweig, 2012; Tang et al., 2015), the clinical implications of mindfulness-induced changes in self-referential processing have received surprisingly little attention. Despite the natural extension, the potentially rich overlap between change in self-referential processing as a function of mindfulness training and internalizing psychopathology remains a highly promising but unexplored area of inquiry. Consequently, we sought to provide a comprehensive critical integration of these separate literatures to develop a theoretical framework involving specific predictions and anticipated methodological limitations that may challenge their testing. Ultimately, the purpose of the review is to stimulate and guide future research at the fruitful intersection between contemplative and clinical neuroscience.

In sum, separate literatures have shown that: (a) aberrant self-referential processing underlies internalizing psychopathology; (b) mindfulness training reduces symptoms of internalizing pathology; (c) mindfulness training produces changes in self-referential processing. Taken together, an intriguing hypothesis emerges: that mindfulness training ameliorates symptoms of internalizing psychopathology through modulating self-referential processing. The primary purpose of this review is to explore this hypothesis through an integrated neuroscientific analysis of self-referential processing in internalizing disorders and mindfulness training. First, we provide an abridged overview of the neuroscience of self-referential processing, drawing primarily from studies involving the CMS and DMN. Due to the breadth of this literature, this section will be circumscribed to viewing the CMS and DMN as broad brain networks underlying self-referential processing and will not contain extensive treatment of functional significance at the regional level. Second, we show how depression and anxiety are characterized by aberrant self-referential processing. Third, we review the therapeutic benefits of mindfulness training in treating symptoms of depression and anxiety. Fourth, we construct the hypothesis by showing that mindfulness training produces change in overlapping self-referential regions implicated in internalizing disorders. Lastly, we conclude by discussing the limitations of the reviewed literature and suggest some directions for future studies to test this hypothesis more directly.

2. The neuroscience of self

2.1. Cortical midline structures and the default mode network

A unique line of neuroimaging research has sought to identify the neural substrates of self-referential processing by contrasting neural activation between self-related versus non-self-related task conditions. In response to early studies yielding considerable heterogeneity in regional activation (Damasio, 2003; Keenan et al., 2003a,b; Kircher and David, 2003; Vogele and Fink, 2003), Northoff and Bermpohl (2004) identified medial cortical regions that were commonly activated across a number of different self-related tasks, grouping these regions under the singular term CMS (see Fig. 1 for a visual depiction). Specifically, these regions include the orbitomedial prefrontal cortex (OMPFC), anterior cingulate cortex (ACC), medial prefrontal cortex (MPFC), and posterior cingulate cortex (PCC). Putatively, each region is conceptualized to maintain a distinct but interrelated subprocess: (1) the OMPFC generates a representation of stimuli as self-referential; (2) the ACC is involved in the online monitoring of self-referential stimuli; (3) the MPFC is engaged during self-referential evaluation; (4) the PCC is involved in the integration of self-referential stimuli in the broader context of personal experience.

Another neural network that has been implicated in self-referential processing is the DMN. The DMN refers to a network of brain regions, comprised of the MPFC, PCC, ACC, and the left and right inferior parietal lobules (IPLs), that is functionally defined by decreased activation during goal oriented or attention demanding activity (Buckner et al., 2008; Buckner and Vincent, 2007; Raichle et al., 2001). Given the anatomical overlap between the DMN and the CMS (i.e., MPFC, PCC, ACC; see Qin and Northoff, 2011 for a review), the DMN has been implicated in self-referential processing—leading some researchers to propose the notion of a “default self”, arguing that the continuous sense of self may be synonymous with resting state DMN activity (Christoff et al., 2003; Beer, 2007; Boly et al., 2008).

Importantly, both CMS and DMN have been associated with complex self-related processes such as recalling the past, planning for the future, and forming beliefs and attitudes (Northoff et al., 2006; Qin and Northoff, 2011; Buckner et al., 2008; Raichle and Snyder, 2007; Murray et al., 2015)—all functions which inherently involve self-referential processing. Moreover, because self-reference tasks appear to elic
overlapping activation in CMS regions as resting state brain activity (e.g., MPFC, PCC), it is widely postulated that people engage in spontaneous self-referential processing (e.g., self-reflection) while at rest (Whitfield-Gabrieli and Ford, 2012). Indeed, what appears to be shared across tasks that activate the DMN is a focus on self—one’s emotions, personality, memories, and goals. Given that excessive self-focus is a transdiagnostic process that cuts across internalizing disorders, examining CMS and DMN activity as neural measures of self-referential processing may yield important insights into the nature of internalizing pathology.

3. Aberrant self-referential processing in internalizing disorders

3.1. Major depressive disorder

Major depressive disorder is the second leading cause of disability worldwide and is responsible for substantial socio-economic burden (Becker and Kleinman, 2013; Ferrari et al., 2013). In an effort to understand this pervasive and debilitating condition, various theoretical perspectives, cutting across time and culture, have implicated the self (Chentsova-Dutton et al., 2014; Kleinman, 1985). More specifically, cognitive models embedded in modern psychological science have emphasized the maladaptive role of excessive self-focus (Pyszczynski and Greenberg, 1987; Scheier and Carver, 1977; Mor and Winquist, 1988), which all are thought to produce marked increases in negative emotionality (Ingram, 1990). Importantly, self-focus inherently involves self-referential processing and thus should reliably activate CMS regions and the DMN.

Below, we review the literature investigating self-referential processing in major depressive disorder. These studies are discussed to demonstrate that self-referential processing in depression is characterized by a unique interplay among self-referential, limbic, and prefrontal control regions.

3.1.1. The interplay of CMS, limbic, and prefrontal regions

Lemogne et al. (2009) employed a standard fMRI self-referential task (Fossati et al., 2003) requiring participants to judge whether personality traits described them (‘self’ condition) or a socially desirable trait (‘general’ condition). Comparing the self-general contrast between depressed patients and healthy controls, they found that depressed patients had increased activation in the dorsal area of the MPFC and in the left dorsolateral prefrontal cortex (DLPFC). Moreover, depressed patients exhibited increased functional connectivity between the MPFC, DLPFC and DACC that was not observed in healthy subjects. Together, these findings suggest that depression may be characterized by a compensatory attempt to implement cognitive control in response to excessive negative self-evaluation.

Not only is depression associated with a general increase in self-focus, researchers have also demonstrated a depression related bias in the self-referential processing of negative emotional stimuli. For example, a number of studies have shown that depressed patients remember negative self-referential stimuli better than positive stimuli (Bradley and Mathews, 1983; Derry and Kuiper, 1981; Dobson and Shaw, 1987; Banos et al., 2001). Building off the notion that negative self-referential biases may be a hallmark feature of depression, Yoshimura et al. (2010) found that depressed patients showed increased activation in the MPFC and rostral ACC (RACC) during processing of negative words relative to controls. Because the RACC has been previously linked to emotion processing (Bush et al., 2000; Habel et al., 2005; Phan et al., 2002), the enhanced RACC activity observed in depressed patients may reflect an inincreased propensity to experience negative emotion. Consistent with Lemogne’s et al. (2009) interpretation, the hyperactivation of the MPFC may reflect a reflexive tendency to engage in self-evaluation after exposure to negative stimuli. Critically, the activity of the RACC and MPFC during processing of negative stimuli was correlated with depressive symptom severity, demonstrating that the interaction between self-referential and emotional processes relate to the subjective experience of depressive symptoms. Lastly, increased connectivity among the RACC, MPFC, and amygdala was observed for the depressed patients relative to controls, indicating the presence of an overactive neural network that may induce automatic emotional responding when depressed patients engage in processing negative stimuli.

Together, these results support the maladaptive role of excessive self-focus and show that depression may be characterized by increased susceptibility to negative self-focus and the subsequent experience of negative affect. This notion is further supported by Ramel et al.’s (2007) finding that amygdala response to self-relevant emotional stimuli during a negative mood induction correlated with increased recall of...
negative self-referential material.

Rumination is a prototypic feature of depression that involves repetitive thinking and focus on distress (Nolen-Hoeksema and Morrow (1993)). Rumination has been shown to increase when negative emotions are up-regulated (Ray et al., 2005) and has been associated with depressive severity, duration, and risk of relapse (Nolen-Hoeksema et al., 2008; Roberts et al., 1998). On the basis that rumination is a “higher order” form of self-referential processing, Cooney et al. (2010) compared neural activation during rumination between depressed patients and healthy controls. As expected, depressed patients showed increased activation in CMS regions during rumination relative to healthy controls, exhibiting increased activity in the orbitofrontal cortex (OFC), MPFC, and PCC—indicating that the content of depressed patients’ ruminations may be more self-centric, evaluative, and possibly autobiographical. Depressed patients also exhibited more activation in the RACC and subgenual ACC (SACC), areas that has been implicated in mood regulation (Drevets, 2000; Gotlib and Hamilton, 2008). Consistent with the aforementioned study by Yoshimura et al. (2010), this pattern of ACC activation may reflect an increased tendency for depressed patients to orient internally and experience increased emotionality during rumination relative to controls. Lastly, in line with Lemogne et al.’s (2009) finding, depressed patients also exhibited greater DLPFC activation. It is unclear, however, whether the hyperactivation of the DLPFC during rumination reflects an impaired regulatory mechanism or—similar to what Lemogne et al. (2009) proposed—represents the compensatory recruitment of cognitive control in response to internal conflict. Nonetheless, there is substantial evidence to suggest that difficulties in cognitive control and attentional disengagement underlie much of the deleterious effects of depressive rumination (see Köster et al., 2011 for a review).

In an innovative study focused on maximizing ecological validity, Kessler et al. (2011) contrasted neural activity elicited from individualized statements pertaining to core interpersonal difficulties from control statements. Results revealed that the presentation of individualized statements produced more activation in the ACC and MPFC, relative to the control condition across both groups—highlighting the increased self-relevance of the personal condition. Importantly, depressed patients demonstrated increased activity in limbic and subcortical regions relative to healthy controls, suggesting that depressed patients experience greater emotionality in response to negative personally relevant stimuli.

Indeed, the idea that depressed patients experience greater emotionality during processing of negative self-relevant stimuli receives further support from Siegle et al. (2002) Consistent with the notion that depressed patients exhibit increased self-referential processing in response to negative stimuli (Yoshimura et al., 2010), increased activity in the MPFC and PCC was observed during personally generated negative vs. positive valenced word contrasts in depressed patients relative to controls. Furthermore, depressed patients exhibited sustained amygdala processing in response to personally relevant negative words relative to controls. Importantly, the sustained amygdala activity in depressed patients was correlated with self-reported rumination, further elucidating a relationship between exaggerated affect and self-referential processing. Interestingly, amygdala activity was also inversely correlated with DLPFC activity, suggesting that depression may involve decreased inhibition of the amygdala via prefrontal control regions.

In a follow up study, Siegle et al. (2007) found that the ACC may be mediating the relationship between sustained amygdala activity and reduced DLPFC activity in depression. Specifically, increase in ACC activity was observed in response to negative words and accounted for the majority of the variance observed between sustained amygdala activity and reduced DLPFC activity. Furthermore, the functional connectivity among the amygdala, ACC, and DLPFC was reduced in depressed patients relative to controls. Given that the ACC is a core region within the CMS unit, the impaired functional relationship between the DLPFC and amygdala may be in part mediated by self-referential mechanisms such as the pervasive monitoring of current self versus an idealized standard (Lemogne et al., 2009). Indeed, animal models suggest that the DLPFC’s influence is mediated through projections from CMS regions—the MPFC, ACC, and OFC (Ghashghaei and Barbas, 2002; Ray and Price, 1993).

Interestingly, a relationship between prefrontal regions and the ACC was also observed in a study by Wagner et al. (2012). Participants were presented with negative, positive, and neutral self-relevant statements prior to completing a stroop task during fMRI. RACC activation increased during the negative statements relative to the other conditions, and was correlated with depressive symptoms. Critically, RACC activity during the negative condition was positively correlated with prefrontal activity during the stroop task, illustrating a possible compensatory mechanism for the increased demands on cognitive control as a function of interference elicited by the negative self-relevant statements.

Similarly, Hooley et al. (2009) presented remitted depressed patients and healthy controls with recordings of their mothers either praising or criticizing them, or discussing a neutral topic. Again, unique activation in the DLPFC, ACC, and amygdala was observed in the remitted depressed patients but not the healthy controls. Specifically, formerly depressed patients exhibited hypoactivity in the DLPFC and ACC in all conditions relative to healthy controls; but hyperactivity in the amygdala only during criticism. The overall decrease in prefrontal and ACC involvement across the conditions coupled with the specific increase in amygdala activity during criticism is suggestive of a general impairment in cognitive control during emotional responsivity to negative events. This interpretation is in line with the emerging theme that the increased coupling between DLPFC and CMS regions represents a compensatory but ineffective attempt to regulate the emotional consequences of excessive self-focus.

Together, these results suggest that the emotion regulatory deficits associated with depression may be characterized by a unique interplay among CMS, prefrontal control, and subcortical affective regions. The aforementioned studies clearly provide neural evidence in support of the long-standing theoretical postulation that depression involves excessive self-focus and a bias towards negative self-relevant thoughts. Moreover, such aberrant self-referential processing may be modulating the functional relationship between the amygdala and PFC, impairing the “top-down” inhibition of limbic activity and emotion regulation more generally (Davidson et al., 2003; Ochner et al., 2002, 2004).

3.1.2. DMN task suppression

Studies involving depressed patients have reported reduced DMN task suppression (i.e., increased DMN activity during task engagement) during both emotion and attention demanding tasks. For example, Sheline et al. (2009) found that relative to healthy controls, depressed patients exhibited reduced DMN suppression during both passive viewing and active reappraisal of negative pictures, suggesting that hyperactive self-referential processing may correspond to a general automatic emotion regulatory deficit rather than a specific impairment with voluntary “top-down” cognitive regulation (i.e., reappraisal).

Similarly, Rose and Ebmeier (2006) and Rodríguez-Cano et al. (2014) found that although depressed patients and healthy controls performed equally well on a working memory task, depressed patients exhibited reduced suppression in the medial OFC (MOPFC) and ACC during task completion. Remitted depressed patients have also exhibited reduced DMN suppression that mirror symptomatic MDD patients (Bartova et al., 2015). Interestingly, reduced DMN suppression was not identified in all remitted patients but only those who had adolescent-onset depression, suggesting that more chronic and severe depression correspond to more pronounced DMN suppression. In line with the previous studies, the reduction of DMN suppression in remitted patients was related to increased rumination—providing more direct evidence linking DMN activity with self-referential processing.

The DMN has been associated with self-referential processes such as autobiographical recall, self-evaluation, and planning for the future.
3.1.3. Resting state DMN connectivity

A growing number of studies have identified aberrant resting state activity in individuals with depression. Before reviewing these studies however, it is important to reiterate that resting state activity does not involve experimental parameters. Consequently, activity fluctuations are observed in relation to other regions in the brain with the assumption that similar brain activity patterns reflect the degree of functional connectivity between the regions measured. With that in mind, the resting state connectivity of the DMN has been found to be highly discriminant in patients with depression relative to healthy controls (Zeng et al., 2012; Berman et al., 2014), suggesting that depressive-related differences in self-referential processing extends beyond active task conditions to periods of inactivity. Interestingly, another study comparing whole-brain functional connectivity to a default mode network node between depressed patients and healthy controls found differences only during rest epochs but not active task conditions (Berman et al., 2010). Indeed, MDD patients have been observed to exhibit increased connectivity in the DMN, but decreased global whole-brain connectivity (Berman et al., 2014)—illustrating a specific neural profile that may prove fruitful as a diagnostic indicator and treatment outcome measure.

In general, depressed patients exhibit hyperconnectivity both within the anterior regions of DMN and between the DMN and the affect salience network—a paralimbic emotion processing network involving the frontal-insular cortex, ACC, amygdala, and temporal poles. Greicius et al. (2007) found increased resting state functional connectivity of the SACC, precuneus, and the thalamus to the rest of the DMN in depressed patients. Importantly, functional connectivity between the DMN and the SACC correlated with the duration of depression. Having connections to the hypothalamus, amygdala, and other limbic structures, the SACC is part of a broader affective network that is speculated to play a critical role in emotional processing (Öngür et al., 2003; Phillips et al., 2003; Price and Drevets, 2010)—with multiple studies finding increased connectivity between the amygdala and SACC, in addition to its relation to changes in negative affect and onset of depression (see Davey et al., 2012 for a review). Structurally, reduced SACC volume has been linked to depression (Rodríguez-Cano et al., 2014) and that stimulation of white matter tracts leading to the SACC has been associated with remission of depression. Together these studies illustrate that even during rest, an aberrant interaction between self-referential and affect networks can be observed in relation to depressive symptoms.

Recent work has uncovered a specific relationship between default mode connectivity and rumination. Zhu et al. (2012) found increased default mode connectivity in the MPFC and ACC in depressed patients relative to healthy controls, with the magnitude of connectivity positively correlating with self-reported rumination—further reinforcing the characterization of the MPFC and ACC as key regions of self-referential processing. Using a novel approach, Hamilton et al. (2011) examined DMN dominance—calculated as relative resting-state levels of DMN over TPN activity—in depressed and healthy controls and correlated DMN dominance with measures of maladaptive rumination and adaptive reflection. Although depressed patients and healthy controls did not differ on mean levels of DMN dominance, greater dominance of DMN in depression was associated with higher levels of rumination and lower levels of reflection. Together, these findings underscore a specific relationship between the DMN and rumination, and support the broader notion that reduced DMN task suppression in depression reflects over engagement in self-referential processing.

There is also evidence showing that depression involves hyperconnectivity between the DMN and cognitive control regions. Sheline et al. (2010) found that the dorsal MPFC (DMPFC) exhibited increased connectivity to seed regions of cognitive and affective networks, implicating the anterior cortical midline in maintaining the dysfunction of these networks. Davey et al. (2012) also found that depressed patients exhibited enhanced DMPFC and DLPPC connectivity with the ACC. As discussed in the previous section, DLPPC dysfunction has been identified in depression and linked with hyperactive cognitive activity (Leomoge et al., 2009, 2010; Siegel et al., 2002). Consequently, the increased resting state connectivity between the DLPPC, DMPFC and ACC further elucidate an aberrant relationship between cognitive control and self-referential processing in depression. Zuo et al. (2012) found structural evidence to further support this relationship, reporting that the integrity of the frontoparietal fiber tract was compromised in depressive patients and that the degree of white matter integrity was negatively correlated with rumination scores.

Consistent with the notion that depression involves impaired cognitive control, altered functional connectivity has also been observed between the DMN and salience network (SN), a key neural substrate of stimuli detection and orientation. Specifically, increased SN activation in response to negative stimuli, as well as aberrant resting state connectivity have been linked to repetitive negative thinking and depression more broadly (Strigo et al., 2008; Sliz and Hayley, 2012; Diener et al., 2012; Manoliu et al., 2014; Burrows et al., 2017). Moreover, the SN has been implicated in the switching between DMN self-referential activity and cognitive control (Sridharan et al., 2008; Uddin, 2015). Together, these studies suggest that SN dysfunction may contribute to the negativit bias of depression by increasing attentional sensitivity toward negative self-referential stimuli.

Overall, these studies strongly demonstrate that depressed patients have trouble disengaging from themselves, exhibiting excessive task-irrelevant self-referential processing which may impair cognitive control and affect regulation (Siegle et al., 2002; Joormann and Gotlib, 2008; Koster et al., 2011).

3.2. Anxiety disorders

Sharing overlapping symptomology and risk factors, anxiety disorders often co-occur with depression (Kendler, 1996). Moreover, anxiety disorders are some of the most common and debilitating psychological disorders—contributing to billions of dollars in productivity loss and economic burden (Hoffman et al., 2008; Baxter et al., 2013). Approximately 85% of patients with depression also experience symptoms of anxiety and likewise, comorbid depression occurs in up to 90% of patients with anxiety disorders (Gorman, 1996; Simon, 2009). Anxiety and depressive disorders also share similar cognitive, affective, interpersonal, and behavioral maintenance factors, with similarities outweighing differences (Harvey, 2004). With respect to self-referential processing in particular, anxiety disorders, like depression, are associated with excessive self-focus and self-related cognitive biases (Clark, 1999). Indeed, theoretical models of anxiety and depression both posit that self-referential processes (i.e., rumination or worry) play a substantial role in the perpetuation of negative affect and impaired cognitive functioning (Watkins, 2008). Despite the diagnostic, etiological, and theoretical overlap between anxiety and depression, there is noticeably less research examining self-referential processing in anxiety disorders. The majority of the studies conducted in this area have been specific to social anxiety—possibly reflecting the challenges of
managing the diagnostic and associated symptomatic heterogeneity within the broader label of anxiety. Moreover, social anxiety may lend itself more naturally to studies of self-referential processing given the inherent centrality of self and other. Indeed, theoretical models of social anxiety have focused heavily on the self and the consequences of negative self-referential evaluation (Schlenker and Leary, 1982; Rapee and Heimberg, 1997; Heimberg et al., 2010).

3.2.1. CMS & limbic interaction

A pioneering study by Blair et al. (2008) presented socially anxious patients and healthy controls with self vs. other statements containing praise or criticism, finding that relative to healthy controls, socially anxious patients exhibited increased activity in the MPFC and amygdala during self-directed criticism. The fMRI finding was complemented by behavioral data indicating that socially anxious patients rated criticism to be significantly more unpleasant than healthy controls. Interestingly, socially anxious patients exhibited more elevated responses to self-criticism in the DMPFC. Given that the DMPFC has been associated with the ability to mentalize others (Mitchell et al., 2006; Moran et al., 2006), these results suggest that socially anxious patients may have an enhanced representation of other people’s mental states, perhaps reflecting their concern with how others view them.

This hypothesis was supported in a follow-up study by Blair et al. (2011), in which they replicated their previous findings and found that, opposite healthy controls, socially anxious patients exhibited increased activation in the ventral MPFC (vMPFC) in response to 2nd person statements relative to 1st person statements. Given that the vMPFC is typically associated with processing 1st person statements (Denny et al., 2012), this finding suggests that socially anxious individuals may reflect on themselves through the lens of others instead of 1st person self-evaluation. This valuable insight may lead to more refined cognitive treatments and introduces a unique set of implications centered on the difference between 1st and 2nd person self-relevant processing.

To further clarify the role of self-referential processing in social anxiety, Boehme et al. (2015a) compared the neural activity of low and high socially anxious individuals while viewing stimulated social situations. During the task, participants were instructed to direct their attention inward (i.e., on bodily states, thoughts, emotions) or outward (i.e., to the person in the video clip). Results revealed that highly socially anxious individuals demonstrated increased activation of the MPFC, PCC, temporal parietal junction, and temporal poles during the inward attention vs. outward condition relative to their low socially anxious counterparts. Moreover, activation in these regions in the high socially anxious group was positively correlated with a trait measure of self-focused attention. Importantly, these findings provide direct evidence linking self-focus with increased activation of the MPFC and support the more general notion that excessive self-focus in social anxiety corresponds to hyperactivity in CMS regions. Theoretically, this study clearly demonstrates that the maladaptive influence of excessive self-focus is not specific to depression but rather supports Ingram’s (1990) postulation that excessive self-focus is a common factor of internalizing disorders.

More recent work has examined the intersection of self-referential processing, emotion, and cognitive functioning by comparing the performance and neural activity of socially anxious patients to healthy controls during an emotional Stroop task containing “disorder-related” words (e.g., ‘speech’, ‘blush’) and neutral words (Boehme et al., 2015b). As predicted, socially anxious patients exhibited emotional Stroop interference to disorder-related words. Interestingly, fMRI analyses revealed that socially anxious patients showed stronger activation in the amygdala, bilateral insula, MPFC, dorsal ACC, and language areas to disorder-related words relative to healthy controls. Notably, the strength of the interference effect was positively correlated with activation in self-referential regions (i.e., MPFC, dorsal ACC). In line with Blair et al. (2008, 2011) conclusions, these results suggest that socially anxious patients are more likely to engage in negative self-referential evaluation in response to socially relevant stimuli—which may in turn, disrupt executive functioning and impede emotion regulation.

In other studies not involving explicit self-referential stimuli, similar patterns of neural activity involving hyperactivation of the amygdala and self-referential regions have been observed in socially anxious participants responding to words (Schmidt et al., 2010), faces (Labuschagne et al., 2012; Phan et al., 2013), anticipation (Boehme et al., 2013), photographic scenes (Heitmann et al., 2016), and videos (Pujol et al., 2013). Given that theoretical models of social anxiety have emphasized the role of post-event processing (e.g., retrospective rumination) in generating and maintaining social anxiety (Clark and Wells, 1995; Rapee and Heimberg, 1997; Heimberg et al., 2010), these studies suggest that abnormal amygdala and MPFC activity may contribute to maladaptive post-event processing of non-explicit self-relevant stimuli. This idea is supported from a study conducted by Heitmann et al. (2016), in which they compared the functional connectivity between socially anxious patients and healthy controls during an ecologically valid picture viewing task involving disorder related (e.g., giving a speech, job interview) and neutral images. Consistent with depression, the analyses revealed that socially anxious patients exhibited hyperconnectivity among the amygdala, MPFC, and ACC, suggesting that internalizing disorders more broadly may be characterized by a dysfunctional interaction between self-referential and limbic networks.

Collectively, these findings share considerable overlap with the depression literature insofar as both forms of internalizing disorders appear to be characterized by hyperactivity in CMS and limbic regions. This conclusion, however, remains speculative given the small number of studies that used explicit self-referential task paradigms. Moreover, few studies have specifically sought to compare self-referential processes between depression and anxiety. A broader limitation is that research has focused primarily on social anxiety—future research should determine whether the aberrant self-referential processing identified in social anxiety generalizes to other anxiety disorders.

3.2.2. DMN activity & connectivity

Zhao et al. (2007) found evidence that DMN activity in anxiety may be characterized by an anterior-posterior dissociation. In comparing DMN activity in anxious patients and healthy controls who listened to emotionally neutral words alternating with rest, anxiety patients exhibited reduced MPFC deactivation (i.e., increased MPFC activity) but greater PCC deactivation during the listening vs. rest condition. Given that past studies have linked reduced MPFC suppression with higher levels of subjective anxiety (Simpson et al., 2001), the hyperactivity of the MPFC could reflect difficulties with attentional disengagement from internal distress during the task. On the other hand, the greater deactivation of the PCC is in line with previous studies associating lower PCC activity with sadness, anxiety, and general neuroticism (Liotti et al., 2000; Sampaio et al., 2014), suggesting that the abnormal deactivation of the PCC in the posterior regions of the DMN may reflect impaired emotion inhibition. Importantly, this anterior-posterior DMN dichotomy received further support from a recent study by Coutinho et al. (2015), showing that resting state functional connectivity of the anterior portions of the DMN was positively correlated with anxiety, whereas connectivity in the posterior regions was negatively correlated.

Notably, these conclusions are remarkably consistent with a study conducted by Liao et al. (2010), in which they found that relative to healthy controls, socially anxious patients exhibited decreased resting state connectivity in the PCC but increased connectivity in the dorsal MPFC. Critically, the increased functional connectivity in the MPFC positively correlated with self-reported fear of being evaluated by others, suggesting that increased self-focus may be a response to enhanced fear sensitivity of social evaluation. Together, these findings are not only consistent with the anterior-posterior dissociation, but also add to a growing body of evidence pointing to the central role of the MPFC as a “self-referential hub” of social anxiety.

Importantly, aberrant resting state connectivity in social anxiety is
not limited within the DMN. For example, Hahn et al. (2011) found reduced functional connectivity between the amygdala and MOFC as well as the PCC in socially anxious patients compared to healthy controls. Interestingly, both the strength of the PCC-amygdala and OFC-amygdala connectivity was negatively correlated with anxiety scores, suggesting that the magnitude of connectivity between self-referential and subcortical regions might reflect the degree of subjective anxiety.

3.3. Summary

Collectively, these studies convincingly demonstrate that aberrant self-referential processing underlies both depression and anxiety. Consistent across studies using a variety of self-referential task paradigms, depressed and anxious patients exhibited hyperactivity in self-referential and limbic regions. Moreover, the majority of the reviewed research also reported abnormal connectivity within the anterior regions of the DMN and between the DMN and affective brain regions. This pattern of neural activity suggests that internalizing disorders may be broadly characterized by a dynamic interplay between excessive self-referential processing and emotion dysregulation. Although more specific to depression, many of the studies also reported abnormal prefrontal activation and reduced connectivity between frontal control regions and the amygdala, suggesting that overactive self-referential processing may impair top-down emotion regulation (Etkin and Schatzberg, 2011). Consequently, interventions targeted to reduce rumination, worry, self-evaluation, and other forms of self-focus may be particularly promising in treating internalizing disorders.

4. Mindfulness training & self-referential processing

Given that we have established the premise that internalizing problems are linked to maladaptive self-referential processing, converging lines of research have shown that mindfulness training: a) reduces symptoms of depression and anxiety, and b) modulates self-referential processing. The aim of this section, therefore, is to draw from these studies to construct the hypothesis that the therapeutic properties of mindfulness training in treating depression and anxiety are accompanied, and possibly mediated, by changes in self-referential processing. Prior to reviewing the literature, however, it is necessary to define mindfulness training and highlight its place within the complex, heterogeneous construct of mindfulness. This way, we aim to appropriately contextualize mindfulness training, a contemporary practice with diverse historical, religious, and philosophical roots, within the broader milieu of clinical psychological science.

4.1. What is mindfulness training?

Mindfulness training originates from the Buddhist contemplative traditions, and is considered a derivative of older soteriological oriented meditative practices (Ahir, 1999). These practices involve applying concentrated equanimous attention to present-moment sensory phenomena with the goal of eradicating mental suffering (e.g., craving, aversion; Gunaratana, 2002). In the late 20th century, the construct of mindfulness arose out of its Buddhist roots, shedding its soteriological and mystical connotations and into contemporary western medicine (Kabat-Zinn, 1982; Teasdale et al., 1995). Although the term ‘mindfulness’ varies in its definition and usage, mindfulness training is often defined in psychological terms as the practice of attending to the present moment with non-judgmental awareness (Kabat-Zinn, 1990). For example, mindfulness meditation, a common form of mindfulness training, typically involves sitting in a quiet environment and applying focused attention to ongoing sensory phenomena. Importantly, mindfulness training cultivates a blend of concentrated attention with open nondiscriminatory awareness of the present moment experience. These two elements can be traced back to Theravada Buddhism, wherein two meditative techniques are prescribed: Samatha, intended to cultivate mental concentration and Vipassana, used to develop equanimous insight into the nature of experience. Samatha and Vipassana have been considered complimentary sides of the same coin and are not easily differentiable (Gunaratana, 2002). It is perhaps unsurprising that these subtle but different meditative techniques have often been subsumed together in contemporary mindfulness training.

Recently, these technical differences have received increased scientific attention because of the likelihood that they engender different effects on mind and body. Specifically, Lutz et al. (2008) have bisected mindfulness training as: (1) Focused Attention (FA), which involves directing and sustaining attention on a selected object; and (2) Open Monitoring (OM), involving nonreactive, meta-cognitive monitoring of the present moment experience. Although FA and OM training involve many overlapping elements, each form is conceptualized to develop distinct cognitive abilities. Specifically, OM training is thought to contain more emotion regulatory properties than FA training because of its focus on fostering nonreactive awareness and reducing elaborative processing (Lutz et al., 2008). Critically, both techniques are involved in clinical adaptations of mindfulness training, but rarely explicitly distinguished or practiced separately.

In addition to the technical variation within mindfulness training, there is considerable heterogeneity associated with the broader construct of mindfulness. Both Buddhist scholars and contemplative scientists have expounded on the continuously evolving, polythetic nature of mindfulness (Bishop et al., 2004; Gethin, 2011; Williams and Kabat-Zinn, 2011; Vago and Silbersweig, 2012). Vago and Silbersweig (2012) separated mindfulness as a state, trait, training practice, and a psychological intervention. Such “construct heterogeneity” introduces significant theoretical and design challenges against conducting rigorous experimental studies. For example, mindfulness training could produce differential effects as a function of individual differences in dispositional mindfulness. Further, designs must account for differences in the propensity to engage in state mindfulness during task completion (e.g., completing a fMRI scan with or without the intent to be mindful). In addition, the development of mindfulness and its associated benefits is not exclusive to deliberate training modalities. Langer’s work (2014) exemplifies cultivating mindfulness through daily life, demonstrating that it is both practically feasible and beneficial to engage the world mindfully during everyday activities. Lastly, MBTs differ in the extent to which they utilize mindfulness meditation, mindfulness-based exercises (e.g., raisin exercise from Williams et al., 2007), or promote mindful engagement in daily life. In sum, variation within and between meditative and non-meditative mindfulness training modalities underscore the importance for researchers to recognize and skillfully account for these differences within the context of the research question.

Consideration of construct heterogeneity is relatively new to contemplative science (Lutz et al., 2015), and past research often involved studies in which the different facets of mindfulness are subsumed under its namesake. Consequently, the literature reviewed in this section will comprise of investigations in which mindfulness training and its associated principles are broadly implemented either as part of a MBT or contemplative practice (e.g., meditation course).

4.2. Mindfulness training as a complementary treatment for internalizing disorders

In a comprehensive meta-analysis, Sedlmeier et al. (2012) found robust effect sizes supporting the salutary effects of mindfulness training on a variety of psychological health variables. Interestingly, the largest effects were related to reducing negative emotionality, anxiety, neuroticism and interpersonal problems—variables that are central to internalizing disorders and psychopathology more broadly. Indeed, mindfulness training and its core principles have been incorporated into a number of clinical interventions such as Mindfulness-Based Stress Reduction (MBSR; Kabat-Zinn, 1990, Kabat-Zinn, 2003), Mindfulness
Based Cognitive Therapy (MBCT; Teasdale et al., 1995), Acceptance and Commitment Therapy (ACT; Hayes et al., 1999) and Dialectical Behavioral Therapy (DBT; Linehan, 1993).

Importantly, the efficacy of these MBTs in treating internalizing disorders has received substantial support from a series of recent meta-analyses. In a review of 39 studies totaling 1140 participants, Hoffman et al. (2008) found that MBTs yielded moderate effect size estimates for reducing anxiety and depressive symptoms. Interestingly, large effect sizes were detected after filtering for patients diagnosed with anxiety disorders and depression. Moderate effect sizes were detected for patients with non-internalizing disorders but nonetheless exhibited elevated symptoms of anxiety and depression, demonstrating that MBTs ameliorate internalizing symptoms across diagnostic classification. Indeed, meta-analyses aimed at examining MBTs in patient populations with physical medical conditions show similar effects in reducing depression and anxiety (Cramer et al., 2012; Bohlemeier et al., 2010). Focusing on depression, two meta-analyses found that MBCT was effective in reducing depressive symptoms and risk of relapse in patients with recurrent depression (Chiesa and Serretti, 2011; Piet and Hougaard, 2011). Interestingly, MBTs exhibit similar efficacy as antidepressants in preventing depressive relapse, demonstrating that MBTs are a viable alternative for remitted patients who are unwilling or intolerant of using antidepressants (Kuyken et al., 2015).

These conclusions were further supported from the largest meta-analysis to date conducted by Khoury et al. (2013). Across 209 studies involving 12,145 participants, MBTs exhibited large and clinically significant effects in treating anxiety and depression, including maintenance of treatment gains at follow-up. Although MBTs did not differ from CBT in treatment efficacy, attrition rates for MBT were significantly lower than CBT, suggesting higher commitment among MBT patients. In order to test the commonly held assumption that mindfulness is the central mechanism of MBT, Khoury et al. (2013) also assessed change in self-reported mindfulness as a moderating variable, finding a strong positive correlation between increased mindfulness and clinical outcomes.

Despite the importance of establishing an association between increased mindfulness and clinical outcomes, there remains much to be uncovered regarding the underlying therapeutic mechanisms. One inherent limitation in conducting mechanistic studies of mindfulness training within the broader context of MBTs is that MBTs involve a variety of treatment components that are unrelated to mindfulness training (e.g., psychoeducation, therapeutic alliance, etc). Consequently, MBTs only serve as proxies for mindfulness training insofar that mindfulness training comprises a core component of the intervention. In other words, discerning the therapeutic mechanisms of mindfulness training is a complicated endeavor, and it is difficult, and arguably unreasonable, to divorce mindfulness training from its implementation as part of a broader clinical intervention. Despite recognizing that mindfulness training cannot be equated to MBTs, it is for this reason that we elected to draw upon studies from both efficacy research on MBTs and more “insular” experimental work involving the neuroscience of mindfulness meditation.

Responding to the challenges associated with delineating the therapeutic mechanisms of mindfulness training, Sedlmeier et al. (2012) emphasized the value of forming theories informed by Buddhist philosophy. Not only would an intergrationalist approach make use of the vast canonical literature on meditative mindfulness practices, it would also facilitate the bridging of two highly comprehensive epistemological systems aimed at understanding and reducing psychological suffering. Embracing this approach, contemplative scientists have begun to examine the effects of mindfulness training on self-referential processing. Indeed, Buddhist teachings ascribe meditative mindfulness practices as a way to reduce self-focus and personal identification—processes that, on the basis of the current review, appear central to the development and maintenance of internalizing disorders.

4.3. The effects of mindfulness training on self-referential processing

Farb et al. (2007) conducted one of the first studies examining mindfulness training and self-referential processing by comparing the neural activity of participants who had completed an 8-week mindfulness training course (MT) relative to novice participants. Participants were instructed to assume two distinct modes of self-reference during processing of trait adjectives: (1) narrative focus, aimed at elaborating on what the words mean to the participant, and (2) experiential focus, aimed at attending to present moment sensory experiences without purpose or attachment. Together, narrative and experiential focus served as proxies for cognitive rumination and mindful awareness (i.e., the mode of experience cultivated through mindfulness training) respectively. Results revealed that both novices and MT participants exhibited similar neural activity during the narrative focus condition, with expected activation in CMS and left hemisphere language regions. Experiential focus, however, was associated with significant group differences. Novices exhibited relatively small reductions in CMS regions and increased activation in left lateralized prefrontal regions. MT participants on the other hand, exhibited a pervasive shift away from CMS activation towards a right lateralized network comprised of the ventral and dorsal lateral PFC and insula. Notably, MT participants also exhibited reduced amygdala activity during experiential focus.

Together, these results show that the neural substrates of mindful awareness are malleable as a function of mindfulness training. Importantly, the shift away from CMS regions toward lateral prefrontal regions in MT participants may be indicative of a self-detached experience in which emotion processing (as evidenced by the reduced activation of the amygdala) is less coupled with the “self-relevance” of the stimuli. In addition, functional connectivity analyses revealed a strong coupling between the right insula and MPFC in novices, whereas the right insula was coupled with the DLPFC in MT participants. This differential pattern of functional connectivity suggests that mindfulness training may promote the shifting of interoceptive experience away from habitual self-referential processes toward a more deliberate mode of sensory awareness maintained by lateral prefrontal regions. In line with Buddhist teachings, these findings suggest that mindfulness training promotes a unique mode of self-reference wherein “default” CMS networks are quieted, facilitating the experience of sensory phenomenon through the lens of self-detached observation rather than self-centric narratives.

Similarly, Grant et al. (2011) studied Zen meditators and found decreased activation in the MPFC, OFC, DLPFC, and amygdala compared to controls during a pain induction. Interestingly, pain induction studies have shown a reliable increase in DLPFC activity thought to reflect cognitive stimulus evaluation (Strigo et al., 2003). Not only did the advanced practitioners exhibit lower DLPFC activity, larger reductions in the DLPFC were associated with lower subject ratings of pain, suggesting that reduced cognitive evaluation may correspond to less pain perception. Together, these results are consistent with Farb et al. (2007), suggesting that focused attention to the present moment dampens self-referential processing. This in turn, appears to reduce the subjectivity intensity of negative experience associated with what is typically identified as aversive stimuli. Although there are technical discrepancies between Zen meditation and mindfulness training, there is an overlapping emphasis on the cultivation of non-elaborative attention—an emerging mechanism that appears to reduce self-referential and limbic activity.

To further investigate the relationship between self-referential professing and emotion, Taylor et al. (2011) compared neural activity between experienced meditators and beginning meditators while they viewed negative, positive, or neutral pictures in mindful and natural states. Interestingly, all participants self-reported reduced emotional intensity across valence categories during mindful viewing, suggesting a global attenuation of emotional reactivity as a function of mindful picture viewing. However, the fMRI data revealed significant
differences between experienced meditators and novices such that the advanced practitioners exhibited deactivation of the MPFC and PCC across all valence types during mindful picture viewing. On the other hand, novices exhibited down-regulation of the left amygdala via increased prefrontal activation during negative and positive pictures. These findings strongly suggest that although voluntary engagement in mindfulness attenuates subjective emotional reactivity in both experienced meditators and beginners, these effects are maintained by markedly different neural mechanisms. Consistent with Farb et al. (2007) and Grant et al. (2011), extensive mindfulness training appears to promote emotional stability via direct attenuation of self-referential processing, whereas novice trainees engage in more effortful top-down regulation of the limbic system. Interestingly, that experienced meditators exhibited deactivation in the MPFC and PCC—key regions of the DMN—introduces the possibility that prolonged training practice may alter DMN activity.

Altered DMN activity has, indeed, been associated with mindfulness training. In the first study to examine self-referential processing during meditation, Brewer et al. (2011) compared self-reported mind wandering, DMN activity, and functional connectivity in highly experienced meditators relative to matched controls as they performed FA, OM, and loving-kindness (directing well-wishing towards others) meditation. Results revealed that across all meditations, the DMN was less active in meditators than controls, supporting the notion that meditative practice attenuates self-referential processing. Furthermore, meditators reported less mind wandering and exhibited increased functional connectivity among the PCC, DACC, and DLPFC relative to controls. Meditators also exhibited a unique pattern of functional connectivity among the MPFC, temporal lobes, and insula. The increased connectivity between the PCC and “task-positive” brain regions (e.g., DACC and DLPFC) may suggest an interaction between self-referential processing and present-focused attention such that frontal control regions are repeatedly engaged to attenuate “self-referential interference” (e.g., mind wandering). The increased connectivity between the MPFC and interoceptive brain regions (i.e., temporal lobes and insula) is consistent with other studies demonstrating that mindfulness training enhances bodily awareness. Interestingly, there were no significant differences in these patterns of functional connectivity during rest relative to meditation, suggesting that with greater meditative experience, the waking “default” state may be transformed to resemble the states cultivated during training.

Collectively, these studies illustrate an emerging pattern that mindfulness training modulates self-referential brain regions in ways suggestive of reduced narrative self-focus. This notion, however, is not unchallenged. A recent study by Lutz et al. (2016) found increased MPFC activation in long-term meditators relative to controls during processing of self-praise and self-criticism. Moreover, MPFC activation correlated positively with self-reported non-reactivity and showed decreased functional connectivity to posterior DMN regions. Self-report affect ratings indicated that meditators experienced less difference between self-praise and criticism relative to controls, supporting the equanimous effects of extensive mindfulness training. Interestingly, the fMRI analysis revealed that meditators exhibited increased activation in emotion generative regions during both conditions relative to controls. This constellation of findings is inconsistent with the literature reviewed above insofar that the neuroimaging data suggests that meditators were engaged in more self-referential and emotional processing. One intriguing explanation is that the meditators were more aware and accepting of their immediate response to the self-relevant stimuli instead of elaborating or regulating their experience. Consequently, this approach could have enhanced present-moment emotional processing, which in turn reduced affective ratings after the actual experience. A related methodological implication is that the study employed explicitly self-relevant stimuli (e.g., statements about the participant). In other words, meditators may be less self-focused during processing of non-explicitly self-relevant stimuli (as is the case in previously reviewed studies) but more self-focused during processing of explicitly self-relevant stimuli. Although speculative, this interpretation is consistent with the notion that mindfulness training fosters active engagement in whatever is occurring in the present moment.

5. Conclusion

Internalizing disorders are characterized by excessive self-focus and emotion dysregulation, which have been evidenced neurally through aberrant activity in self-referential and emotion generative brain regions. Meta-analytic reviews demonstrate that MBTs are effective in reducing internalizing symptoms. A growing body of research also shows that mindfulness training modulates overlapping self-referential and limbic brain regions implicated in internalizing disorders. Together, these three separate but complimentary literatures converge...
to form the hypothesis that mindfulness training ameliorates internalizing symptoms through modulating the self-referential system. Fig. 2 integrates these findings into a process model, illustrating potential pathways through which mindfulness training shifts the tendency to engage in narrative self-elaboration toward experiential awareness.

Interestingly, mindfulness training exhibits both similarities and differences to cognitive-behavioral treatment models of internalizing disorders. Specifically, mindfulness training shares conceptual similarities to CBT insofar that both seek to attenuate the bi-directional link between maladaptive cognitions and deregulated emotionality. Unlike CBT approaches however, mindfulness training is not necessarily predicated on intervening at the level of cognition (e.g., cognitive restructuring, thought challenging). Rather, mindfulness training promotes shifting the frame of self-reference through which maladaptive cognitions arise. Consequently, to the extent that the enhanced CMS activity observed in internalizing patients corresponds to an increased propensity to engage in excessive narrative-based self-focus, it is reasonable to expect that mindfulness-induced shifts from such self-centrism to open experiential awareness will be accompanied by reduced CMS brain activity.

5.1. Evidence for self-referential processing as a therapeutic mechanism of mindfulness training

Importantly, there is limited but emerging evidence to support the notion that mindfulness training ameliorates internalizing symptoms through modulating self-referential processing. Way et al. (2010) measured neural activity during rest and viewing of emotional faces in meditation naive college students. Consistent with studies involving experienced meditators, Way and colleagues found that dispositional mindfulness was negatively correlated with resting activation in the MPFC, parietal self-referential regions, and amygdala. In contrast, depressive symptoms were positively correlated with activity in the same areas. Together, these results establish the expected relationships among mindfulness, self-referential processing, amygdala activity, and depression symptoms. Although these findings appear promising, experimental studies involving mindfulness training are needed to draw causal inferences and corroborate the hypothesized mechanism.

Addressing the need for experimental evidence, Goldin et al. (2009) examined the effects of an 8-week MBSR course on self-referential processing in social anxiety patients. Self-report measures revealed decreased symptoms of social anxiety, depression, rumination, and state anxiety across the intervention. To examine self-referential processing, participants underwent fMRI while viewing and responding to positive and negative social trait adjectives—with some trials requiring participants to indicate whether the adjective was self-descriptive. Behaviorally, participants endorsed more positive traits and fewer negative traits after MBSR, indicating less negatively oriented self-perception. Interestingly, differential patterns of neural activity were observed for the negative and positive adjective conditions as a function of mindfulness training. When responding to negative adjectives, participants exhibited increased activation in the left IPL and medial prefrontal cortex regions implicated in attention allocation. In response to positive adjectives, MBSR resulted in decreased activation in self-referential (i.e., MPFC) and language (i.e., left inferior frontal gyrus) regions. Notably, the reductions in self-referential and linguistic regions were only observed for the positive but not negative adjectives. This contrast between positive and negative adjectives suggests that it might be “easier” for socially anxious training novices to experientially respond to positive words because the narrative conceptual link between self and positive traits is more tenuous than self and negative traits. It follows then, that the increased attentional engagement to negative words may reflect the beginning of an effortful process to overturn the deeply rooted habitual tendency to link negative traits with the narrative “I”.

In a separate study involving the same patients, Goldin and Gross (2010) examined MBSR-related changes in emotional reactivity and regulation of negative self-beliefs. Specifically, Goldin and colleagues employed an emotion regulation task that required patients to engage in either breath-focused attention or distraction-focused attention in response to negative self-beliefs (e.g., ‘people always judge me’). Behaviorally, patients reported reduced negative emotionality on trials involving breath-focused attention but not during distraction. Consistent with the first study, the neuroimaging data revealed that MBSR corresponded to increased activation in attention related brain regions during breath-focused attention (i.e., precuneus and IPL). Furthermore, BOLD time series analysis of the amygdala showed that relative to baseline, post-MBSR patients exhibited an initial spike in amygdala activity before decreasing rapidly, indicating that mindfulness training may increase the detection or initial emotional salience of negative stimuli. Critically, the reduction in amygdala activity occurred well before the cue to shift attention to the breath, suggesting that the emotion regulatory benefits of focused breathing may become more automatic and less effortful as a function of training.

Finally, Hözel et al. (2013) conducted a longitudinal study examining the effects of MBSR on GAD patients. Participants were either randomized to MBSR or a general stress management course and completed an fMRI facial affect labeling task before and after the intervention. Relative to controls, MBSR participants showed greater pre-post activation in the ventral lateral PFC (VLPFC)—a region implicated in emotion regulation and GAD symptom expression (Maslowsky et al., 2010). Further supporting the clinical relevance of this change, the pre-post change in VLPFC activity correlated negatively with pre-post changes in self-reported anxiety. MBSR also produced increased connectivity among the limbic, prefrontal, and CMS regions. Interestingly, the strength of the amygdala-prefrontal connectivity was negatively correlated with both anxiety symptoms during post intervention and pre-post changes in self-reported anxiety. Together, the findings suggest that mindfulness training produced meaningful changes in emotion regulation and anxiety symptoms. However, the MBSR-related increase in connectivity between amygdala and prefrontal regions run counter to previous studies showing that down-regulation of limbic activity is associated with negative connectivity between the amygdala and frontal regions (Lee et al., 2012). Given that mindfulness involves openly experiencing sensory phenomena, as opposed to voluntary cognitive down-regulation (e.g., cognitive reappraisal), the increased amygdala-prefrontal connectivity may signal a unique form of emotion regulation involving effortless self-detachment.

In what may be the most direct test of mindfulness training on depression-related self-referential processing, Farb et al. (2007) compared the neural correlates of sadness provocation (i.e., contrasting viewing of sad films vs. neutral films) between MBSR participants and waitlist controls. Consistent with the reviewed literature, sadness provocation in the control group was associated with increased neural activity in the CMS and language regions, and decreased activity in attention and somatosensory regions. In contrast, sadness provocation in the MBSR group elicited less activation in CMS regions and left-lateralized language regions, and greater activation in the insula. Interestingly, right insula activity in the MBSR group was negatively correlated with self-reported depression scores. Not only does this work suggest that mindfulness training reduces narrative-based self-referential processing in response to negative stimuli, it supports the notion that changes in experiential interoceptive awareness may reduce depressive symptoms.

Together, these studies demonstrate that mindfulness training can alter self-referential processing and produce measurable differences in cognition, emotion regulation, interoception, and internalizing symptoms. However, on the basis of the present review, it is apparent that there is a dearth of clinical neuroscience research on mindfulness training. That is, few research studies have specifically examined self-referential processing as a function of mindfulness training in the clinical domain, thereby precluding the direct testing that change in
self-referential processing is a key therapeutic mechanism of mindfulness training in treating internalizing disorders.

### 5.2. Limitations & future directions

The majority of the reviewed studies employed neuroimaging methods to compare between-group neural activity (e.g., patients vs. healthy controls; mindfulness experts vs. novices), a common way to discern unique neural processes associated with a phenomenon or condition of interest. There are, however, several methodological issues in conducting this type of analysis. First, because activation is calculated as a contrast between two conditions, it can be difficult to ascertain whether group differences are attributable to the control or active condition. This challenge is particularly relevant to mindfulness research because experienced practitioners have been shown to exhibit baseline activity that more closely resemble the meditative state (Brewer et al., 2011), thereby reducing the contrast between the comparison conditions. One possible solution is to use imaging methods that do not rely on BOLD contrasts such as arterial spin labeling (Tang et al., 2015), during which paramagnetic tracers are used to calculate absolute measures of cerebral blood flow within the same task condition (see Petcharunpaisan et al., 2010 for a technical overview). Second, the functional complexity associated with a single brain region impedes the ability to draw definitive conclusions regarding the functional significance of observed activation differences. Consequently, functional interpretations, including the ones advanced in our model, should be carefully weighed against methodological limitations and formulated with thorough consideration of the task design and related literature.

Furthermore, many studies were exploratory as opposed to explanatory—exploring whether a theoretical postulation, such as the impairing effects of excessive self-focus, can be evidenced at the neural level. The exploratory nature of these designs in conjunction with the functional diversity inherent in interpreting brain-level data raises significant analytical challenges and increases the susceptibility to circular reasoning (e.g., constructing a study based on a well-established theoretical model of depression, then interpreting the results as evidence for that model; Poldrack, 2006, 2011). It may, therefore, be prudent to shift toward tightly controlled experimental designs that involve isolating specific brain regions within a larger network to generate more “insular” hypotheses regarding functional significance or mechanisms of change. For instance, breaking down specific regions of the CMS (e.g., MPFC as a proxy of self-evaluation) and DMN may be helpful for uncovering a mechanistic understanding of how mindfulness training influences specific forms of self-referential processing and symptomatology in internalizing patient populations.

It is, however, critical to recognize that the validity of our model and efforts to discern mechanisms of change more broadly are challenged by the high likelihood that the effects of mindfulness training differ as a function of experience. Based on the literature (Farb et al., 2007; Grant et al., 2011; Taylor et al., 2011), beginning mindfulness practitioners seem to down-regulate emotion via recruitment of pre-frontal control (possibly reflecting effortful decentering and voluntary engagement of experiential awareness), whereas experienced practitioners appear to achieve emotional regulation through more trait-like changes in self-referential processing (see the comparison of the purple and red lines in Fig. 2). Indeed, it may be fruitful to conduct studies with the knowledge that emotion regulation and other therapeutic processes will likely differ as a function of prior training experience.

The caveat is, however, that no studies to date have observed the speculated “shifts” in emotion regulation, self-referential processing, and attention that are thought to occur over prolonged mindfulness training. The notable absence of empirical evidence supporting these changes not only signifies that a core aspect of our model remains untested but also speaks to the ambiguity surrounding what constitutes “short” vs. “long” term practice, and challenges the operationalization of training experience (e.g., frequency, duration, intensity) more broadly. It is probable that the relationship between mindfulness training and outcome changes is not uniform across individuals. Consequently, repeated assessment across a tightly controlled training period may be a reasonable first step of many toward understanding practice effects.

An important corollary is that individual differences likely play a significant role in shaping the effects of mindfulness training over time. Despite the advantages of the nomothetic approach, conclusions drawn from group-level data do not typically generalize to the individual level (Molenar, 2004), and are ill-suited to capture important idiographic changes over time. Moreover, nomothetic brain-level data cannot capture the variety of phenomenological experiences (e.g., mind wandering, open monitoring) that occur during meditation or other forms of mindfulness training. This inability to discriminate between mind states during training not only precludes the investigation of many phenomenally oriented research questions but also maintains the assumption that participants are always engaged in the induced state of interest (e.g., focused attention on breath). Indeed, the strength and validity of our model rests on the assumption that observable changes in CMS brain activity correspond to changes in the phenomenological experience of self.

Critically, many of these limitations reflect the broader challenge of collecting time-intensive measures of mindfulness training. Thus, it may be worthwhile to explore methodological options that are amenable to repeated recording of neural activity during mindfulness training. Given the well-documented cost and feasibility constraints associated with fMRI, newer neurophysiological recording technologies such as mobile EEG may prove to be useful.

To address these limitations, and begin testing the hypothesis that self-referential processing is a key therapeutic mechanism of mindfulness training, future studies should consider: (a) employing a longitudinal randomized design with at least one active control group in order to delineate changes that are specific to mindfulness training and account for potential confounds such as placebo effects and habituation; (b) varying experimental paradigms to include a variety of self-referential and emotional stimuli across sensory modalities (e.g., verbal statements, images, pre-selected personal experiences) to distinguish the expected self-referential related mechanisms from task-specific confounds; (c) collecting data across multiple time points (as opposed to pre-post design) to draw time series and idiographic inferences; (d) utilizing functional connectivity analysis to determine how mindfulness training influences the inter- and intra-connectivity in networks implicated in internalizing disorders (e.g., anterior-posterior DMN dissociation, relationship between self-referential and emotion brain regions, etc.); (e) employing non-contrast dependent imaging methods such as arterial spin labeling; and (f) maximizing ecological validity by recruiting depressed and anxious patients with no prior mindfulness training experience.

In carrying out this line of research, it is imperative to select task paradigms and recruit samples that are consistent with relevant prior studies as standardization is needed to minimize confounds, enhance validity, and increase cross-study generalizability. Additionally, it may be prudent to consider how broader sociocultural differences influence mode of self-reference. For example, self-referential processing may vary as a function of religiosity (e.g., Christianity vs. Buddhism vs. Atheism) and or cultural values (e.g., individualism vs. collectivism). Therefore, it may be fruitful to account for these factors as potential confounds within the context of this work. Given the breadth of these suggestions in conjunction with feasibility constraints related to time, financial resources, patient access, and equipment capital, it is likely that a long systemic line of research involving many studies will be needed to thoroughly examine this hypothesis.

An initial goal may be to obtain longitudinal evidence establishing the expected relationships among mindfulness training, self-referential processing, and internalizing symptoms in internalizing patient populations. Using the process model delineated in Fig. 2 as a guiding
framework, if reduced self-referential processing is a core mechanism underlying the therapeutic benefits of mindfulness training, then mindfulness-naive internalizing patients randomly assigned to mindfulness training should exhibit initial increases in prefrontal control regions signifying effortful decentering and application of mindful (i.e., experiential) awareness. Over prolonged practice, evidence for reduced self-referential processing is expected to emerge across self-report (e.g., rumination, worry scores), behavioral (e.g., mind wandering probes), and neural (e.g., reduced CMS, amygdala, task-positive DMN activity) measures. Symptomatically, this process should be accompanied by reductions in self-reported internalizing symptoms, with symptom reduction relating to the aforementioned changes in prefrontal (increase in short-term) and CMS (reduction in long-term) activity.

Given that there is no conclusive longitudinal evidence demonstrating the proposed experience-driven shift from PFC to CMS recruitment, the temporal demarcation of what constitutes “short-” and “long-term” remains unclear and in need of additional study. Furthermore, to the extent that overactive self-referential brain activity is sensitive to effortful engagement of mindful awareness, it is plausible that both increases in prefrontal and decreases in CMS activity could be observed within a relatively short (e.g., 8 weeks) mindfulness training intervention. Our primary contention, however, is that irrespective of the timing, directionality, and broader nature of the PFC-CMS relationship, mindfulness-related activation changes in these brain regions are predicted to relate to symptom change. This would provide compelling initial evidence to support the proposed role of self-reference in the maintenance and amelioration of internalizing pathology. Later studies may benefit from investigating: (a) how mindfulness training alters functional connectivity among self-referential, frontal control, and affective networks in relation to internalizing symptoms and emotion regulation more broadly; (b) whether mindfulness training uniquely modulates different forms of self-referential processing such as rumination, worry, and self-criticism; (c) the possible differential effects of mindfulness associated with increased training experience.

5.3. Concluding comments

This review bridges neuroscience, contemplative studies, and clinical psychology to understand the self in relation to internalizing disorders. An emerging idea from this review is that it may be as useful to “let go” of the self as it is to understand it—endeavors that are, from a contemplative perspective, more complimentary than disparate. Indeed, eastern contemplative traditions and contemporary western psychology each offer meaningful ways to understand the self and address suffering. The present review has revealed remarkable parallels between these two epistemological systems, demonstrating that excessive attachment to the self as a fixed narrative entity may lie at the heart of depression and anxiety.

Mindfulness training represents a promising solution to this problem by promoting a shift away from such self-centrism—rife with rumination of the past, and worry for the future—toward an open non-elaborative embracement of the present. There is, however, much to be done in understanding its therapeutic properties. Depression and anxiety are common conditions that have afflicted human minds and captured collective wonderment for centuries. In this light, examining the intersection of self-referential processing, mindfulness training, and internalizing pathology may simply reflect an extension of the perennial prerogative of “knowing thyself.” Retaining an expansive multidisciplinary perspective and integrating the rich knowledge and methodologies of east and west will likely accelerate progress in this collective endeavor. Consistent with the ethos of mindfulness, there is no better (or other) time to start than now.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declarations of interest

None.

Acknowledgements

The authors wish to thank the Mind & Life Institute for their inspiration and support.

References


