Focus on the breath:

Brain decoding reveals internal states of attention during meditation

Helen Y Weng[1,2,3]*, Jarrod A Lewis-Peacock[5], Frederick M Hecht[1], Melina R Uncapher[2], David A Ziegler[2], Norman AS Farb[6], Veronica Goldman[1], Sasha Skinner[1,2], Larissa G Duncan[1,7], Maria T Chao[1,4], Adam Gazzaley[2]

[1]Osher Center for Integrative Medicine, [2]Neuroscape Center, [3]Department of Psychiatry, [4]Division of General Internal Medicine, University of California, San Francisco, San Francisco, California, USA. [5]Department of Psychology, University of Texas at Austin, Austin, Texas, USA. [6]Department of Psychology, University of Toronto, Mississauga, Ontario, Canada. [7]School of Human Ecology and Center for Healthy Minds, University of Wisconsin-Madison, Madison, Wisconsin, USA.

* Correspondence should be addressed to H.Y.W. (helen.weng@ucsf.edu)

Abstract

Evidence suggests meditation may improve health and well-being. However, understanding how meditation practices impact therapeutic outcomes is poorly characterized, in part because existing measures cannot track internal attentional states during meditation. To address this, we applied machine learning to track fMRI brain activity patterns associated with distinct mental states during meditation. Individualized brain patterns were distinguished for different forms of internal attention (breath attention, mind wandering, and self-referential processing) during a directed internal attention task. Next, these brain patterns were used to track the internal focus of attention, from moment to moment, for meditators and matched controls

during breath-focused meditation. We observed that while all participants spent the majority of time attending to breath (vs. mind wandering or self-referential processing), meditators with more lifetime practice demonstrated greater overall breath attention. This new framework holds promise for elucidating therapeutic mechanisms of meditation and furthering precision medicine approaches to health.

Meditation practices, or mental exercises that train qualities of attention, are increasingly used to improve health and well-being in clinical populations as well as the general public. Meditation and mindfulness-based interventions may cultivate sustained attention[1–3], compassion and prosocial behavior[4–6], creativity[7], improved brain structure and function[8–11], less implicit bias[12], reduced stress[13], and decreased symptoms in clinical populations with pain[14], depression[13,15], anxiety[13], and cancer[16]. Based on these promising results, meditation practices are being implemented in a variety of fields such as medicine, psychology, education, business, law, and politics[17]. Collectively, meditation practices may strengthen interoception (awareness of internal bodily sensations)[18,19], cognitive processes (including sustained attention, cognitive monitoring, and meta-awareness)[3,20,21], and emotion regulation (less judgment and more equanimity with internal experiences)[22,23]. With practice, these skills may lead to better monitoring and regulation of physical, emotional, and social processes, contributing to improved health decision-making and behaviors[18,19]. However, the mechanisms through which meditation improves health and well-being are poorly specified, in part because there is currently no precise way to assess the quality of meditation practice[17].

Mental states during meditation are challenging to measure because they are often internal, diverse, and fluctuating. For example, in a core practice of focused attention to the

breath, meditators focus their attention on sensations of the breath, until they notice distraction by other internal or external stimuli, and then nonjudgmentally return their focus to the breath. Even in this simple practice, distinct mental states may occur and are dynamically fluctuating over time: the object of attention (breath or distractions), level of meta-awareness (awareness of object of attention), as well as attitudinal qualities such as nonjudgment, kindness, and curiosity[24–26]. Further, each meditator, and indeed each meditation session, is unique in the content and fluctuation of mental states. This may partly explain why outcomes of meditation-based interventions often vary by individual, suggesting that participants may be employing different cognitive processes even when given similar instructions for formal practice[17,27]. Without the means to characterize mental states during meditation, scientists cannot quantify the internal attention states present during practice, and thus cannot predict or observe how meditation may change attentional qualities over time and contribute to clinical outcomes.

Currently, there is no validated, objective way to measure and quantify these mental states that occur dynamically during meditation practice. Prior work has implicated several brain networks in breath meditation practice, which suggests that diverse mental states may be present during meditation. Studies using functional magnetic resonance imaging (fMRI) show increased activation in networks involved in focused attention and cognitive control[10,28] (i.e., the Executive Function Network [EFN] including prefrontal cortex [PFC], anterior cingulate cortex, premotor cortex) as well as interoception or attention to internal bodily sensations[10,29] (including the insula[18,30]). In addition, decreased activation is seen in regions associated with mind wandering and self-referential processing[3,10,21] (i.e., the Default Mode Network [DMN including the anterior medial PFC, posterior cingulate cortex, and posterior inferior parietal lobule[31,32]]). Consistent with psychological and contemplative theories of how attention is altered by meditation[24,25],

these findings suggest that meditators may focus on the breath via increased engagement of neural networks associated with cognitive control and interoception, and disengage from mind wandering and self-referential processing via decreased DMN activation[3,10]. Disengaging from self-related thought[31] is particularly important from contemplative and clinical perspectives, due to the cognitive and emotional flexibility it cultivates[13,21]. However, the individual differences and fluctuation of mental states during meditation are obscured in traditional univariate analyses of brain data (used for most of these prior studies), in which spatial and temporal information are averaged within and across individuals to highlight consistent results at the group level. In other words, such approaches lack the temporal specificity of when each mental state and associated neural network are most engaged, and group-level averaging may fail to distinguish between different mental states that occur over time during meditation practice.

Initial research suggests that meditation may involve temporally distinct cognitive processes and neural networks. For example, by indicating moments of meta-awareness of mind wandering via button presses, experienced meditators' practice has been characterized as a cycle between four attention states: focused attention, mind wandering, meta-awareness, and refocusing of attention, with each activating a different neural network[33]. Meditators have also been provided continuous real-time neurofeedback[34] from the posterior cingulate cortex, a major hub of the DMN, which helped them maintain focused attention during meditation[35]. While these findings suggest the presence of distinct neural processes during meditation, these group-level characterizations may not be universally generalized because individuals differ in structure-function relationships. For example, across a sample of experienced meditators, activation levels in attention-related regions of the EFN during practice differed in their association with lifetime

meditation practice, such that middle levels of expertise were associated with the greatest activation[28].

Together, existing research suggests that some elements of neural function during meditation may be specific and individualized to each meditator, and may represent different mental states depending on practice history and other contextual factors. Further supporting the view that brain structure and function are unique for each individual, machine learning methods such as multi-voxel pattern analysis (MVPA)[36] can be used to identify and track diverse mental states *within* each person, and can distinguish unique brain patterns associated with visual perception of different objects[37], short- and long-term memory[38], and social vs. physical pain[39]. In addition to being able to distinguish neural patterns associated with attending to many types of *external* stimuli, this approach can also distinguish *internal* states of attention[40,41], which are more difficult to objectively validate. Recent work also suggests that neural signals associated with breath-focus can be distinguished from self-referential processing[42].

Based on these analytic issues, we developed the EMBODY framework (**E**valuating **M**ultivariate Maps of **Body** Awareness) which applies MVPA to fMRI data to classify on-going internal mental states during meditation, which can then be used to quantify meditation skills **(Fig. 1)**. Here, we validated the EMBODY framework by measuring mental states during the core practice of breath meditation in 8 experienced meditators and 8 age and gender-matched novice control participants. The EMBODY Task involved three steps: in **Step 1**, neural patterns associated with internal mental states relevant for meditation were identified separately for each meditator using the Internal Attention (IA) task. In this task, participants were directed to attend to internal stimuli for short intervals (16-50s). With eyes closed, attention was directed to sensations of the breath, mind wandering, and self-referential processing, as well as two control

conditions (attention to feet, attention to ambient sounds; **Fig. 2a**). In **Step 2**, the individualized brain patterns learned from Step 1 were used to decode, from moment to moment, the unknown internal attentional states occurring during an independent 10-minute breath-focused meditation. In **Step 3**, the decoded meditation period was then quantified into attention metrics, such as the percentage of time engaged in breath attention or self-referential processing. To assess criterion validity of the EMBODY task, these brain-derived attention metrics were linked to 1) subjective reports of internal attention during the IA task and after the meditation session, 2) lifetime meditation practice, and 3) trait questionnaires of interoception and mindfulness. We included individuals from both groups because a) meditators are more likely to produce distinct brain patterns from consistent practice in directing and sustaining internal attention, and b) novices are the population most studied in clinical intervention studies. The framework was tested in each individual, while group-level statistics were computed for the entire sample to assess construct validity and inform future research. See **Online Methods** for details.

RESULTS

Step 1: Distinguishing neural patterns of internal attention

The first aim of the EMBODY framework was to test whether MVPA applied to fMRI data could recognize individualized neural patterns associated with internal attention states important for breath meditation. With eyes closed, participants were directed using auditory instructions to engage in short durations (16-50s) in three distinct attentional states relevant for breath-focused meditation (attention to breath, mind wandering, self-referential processing) and two control conditions of internal and external attention (attention to feet and attention to ambient sounds, respectively) (**Fig. 2a**). Confirming this ability to distinguish individual neural

patterns, across all participants, each attentional state yielded a distinct neural signature (all classification accuracies>41% vs. 20% chance for 5 categories, *p*s<0.001; **Fig. 2b**). Furthermore, each attentional state was distinguished at more than twice chance levels, including the brain patterns most relevant for breath meditation (breath=50.5%, mind wandering=41.2%, self-referential processing=49.0%; t_{15}>4.65, *p*s<0.001, Cohen's *d*s>1.16) and the control conditions (feet=43.2%, sounds=43.7%, t_{15}>5.67, *p*s<0.0001, Cohen's *d*s>1.41) (**Fig. 2b**; see **Table S1** for classifier confusion matrix and **Fig. S1** for classifier accuracies by group in **Supplemental Information [SI]**).

The breath meditation-relevant brain patterns were reliably classified in 14/16 participants (at least 2/3 *p*s<0.001 for breath, mind wandering, and self-referential processing, **SI-Table S2).** This included all 8 meditators and 6 of 8 novices, all of whom were used in subsequent analyses. Across participants, average trial-level classification accuracy was positively correlated with subjective ratings of internal attention (across all conditions except mind wandering, r_{13}=0.67, *p*=0.032; **Fig. 2c**), demonstrating that better classification accuracy of the different conditions in the IA task reflects better attention to internal mental states.

Distributed and common brain regions contributing to unique brain patterns. Classifier importance maps that identified the voxels most important in distinguishing the attentional states[43] were unique for each participant and distributed throughout the brain (**Fig. 3a**). For initial characterization of brain regions that supported classification and were common across individuals, a frequency map was computed representing the sum of individual importance maps (**Fig. 3b**). This indicated that no brain region was important for all 14 participants in any mental state (maximum frequency≤10, **Fig. 3b**, **SI-Fig. S2a**, **SI-Fig. S2b**), and frequency histograms

showed that most voxels were important for only 1-3 participants (**SI-Fig. S2c-e**). Voxels that were important for breath-focused attention in a higher frequency of participants (N≥5) were located in the medial PFC (extending to the perigenual anterior cingulate cortex and the anterior mid-cingulate cortex), left dorsolateral PFC, bilateral occipital pole, and right cerebellum. Higher-frequency importance voxels for mind wandering (N≥5) were located in bilateral superior temporal gyrus, right lateral frontal pole, left precentral and postcentral gyrus, and areas of the PFC (left ventromedial, right anterior frontal pole, and right dorsolateral). Higher-frequency importance voxels for self-referential processing (N≥6) were located in bilateral dorsomedial PFC, right orbitofrontal cortex, bilateral dorsolateral PFC, left anterior medial PFC, right supramarginal gyrus, and bilateral precentral gyrus (**Fig. 3b**, **SI-Fig. S2**, **SI-Table S3**).

Step 2: Decoding the focus of attention during breath meditation

Individualized brain patterns for each participant were used to decode the focus of attention during 10 minutes of breath meditation, producing a second-by-second readout of internal attention states of attending to the breath, mind wandering, or self-referential processing (**Fig. 4a-d**). Classifier decisions at each time point were based on the class with the highest classifier evidence values (**SI-Fig. S3**). From these data, "mental events" were defined whenever there were 3 or more consecutive time points that were classified as belonging to the same mental state (**Fig. 4b**).

Step 3: Quantifying metrics of internal attention during breath meditation

Based on MVPA decisions of mental states present during meditation from Step 2, we computed novel metrics of attention during meditation for each participant, including *percentage*

time spent engaged in each mental state, *number of mental events* (or discrete periods engaged in each mental state), the *duration of each mental event*, and the *variance of the durations* (SD; see **Online Methods** for data reduction). For breath-focused meditation, we hypothesized that participants would direct their attention more to the breath than engaging in mind wandering or self-referential processing. Therefore, compared to the other mental states, participants should show greater: 1) percentage time attending to the breath, 2) number of breath mental events, 3) mean duration of attention to the breath, and 4) variance in duration on the breath (greater inter-trial variability due to longer durations).

Mental state profiles during breath-focused meditation. Group-averaged attention metrics during the 10-min breath meditation period are shown in **Fig. 5**. Attention metrics differed in the percentage time engaged in each mental state ($F_{2,12}$=8.93, p=0.001), the mean duration of mental events ($F_{2,12}$=6.47, p=0.005), and the mean variance of event durations ($F_{2,12}$=4.20, p=0.026). Consistent with our hypotheses, we found that during meditation, participants spent more time paying attention to their breath compared to mind wandering or self-referential processing (t_{13}s>3.18, ps<0.007). On average, the 10-min meditation periods contained 56.4 mental events of at least 6-s each (SD=11.26). Although the mean number of events across mental states did not differ significantly (p=0.31), when participants attended to the breath, the mean duration of those events (10.9s [3.5]) was longer than for mind wandering events (8.1s [1.6], t_{13}=3.28, p=0.006) and marginally longer than self-referential processing events (9.0s [2.6], t_{13}=1.94, p=0.07). Similarly, the variability of event durations tended to be greater for attention to the breath compared to both mind wandering (t_{13}=1.92, p=0.08) and self-referential processing (t_{13} = 2.46, p=0.029). This greater variance was likely due to the longer

duration of breath events. See **SI-Table S4** for full statistics and **SI-Fig. S4** for exploratory tests between groups.

Distraction from breath and mental state fluctuations. When participants became distracted from their breath, the distraction period lasted for an average of 20.4s (SD=4.77), and was marginally more likely to be attributed to entering a state of mind wandering (mean count=8.8 [2.46]) than self-referential processing (mean count=7.0 [2.57]; $t_{1,13}$=1.90, p=0.08). When engaged in mind wandering, participants were equally likely to transition to mental states of breath (mean count=8.6 [2.77]), or self-referential processing (mean count=7.9 [3.85], p=0.68), and when engaged in self-referential processing, they were equally likely to transition to breath (mean count=7.5 [2.85], or mind wandering (mean count=7.4 [2.90], p=0.92).

Criterion validity of EMBODY Task metrics. To validate whether EMBODY Task metrics accurately reflect attentional states during meditation, we correlated percentage time spent in each mental state with 1) subjective ratings of attention from the meditation period, 2) reported lifetime meditation hours, and 3) trait measures of interoception[44] and mindfulness[45]. After the meditation period, participants rated the percentage time they paid attention to their breath and to their thoughts, and these ratings were not correlated with any attention metrics (ps>0.68). However, within meditators, total reported lifetime meditation hours were positively correlated with percentage time spent attending to the breath (rho_7=0.71, p=0.047; **Fig. 6a**) and negatively correlated with percentage time engaged in self-referential processing (rho_7=-0.71, p=0.047; **Fig. 6b**) during the meditation period. The difference between these correlations was significant (z=2.78, p=0.005). Lifetime practice and percentage time engaged in mind wandering

were not significantly correlated (rho_7=-0.17, p=0.69). Furthermore, demonstrating specificity of the task metrics, the amount of lifetime hours meditating *specifically* on breath sensations was positively correlated with percentage time attending to the breath during the meditation session (rho_7=0.74, p=0.037; **Fig. 6c**) and negatively correlated with self-referential processing (rho_7=-0.91, p=0.002; **Fig. 6d**; difference between correlations z=1.92, p=0.056). Lifetime hours meditating on other body sensations or other meditation practices were not associated with any mental state during meditation (**Fig. 6e-f**; body ps>0.35; other practice ps>0.38).

Self-reported attention regulation, an interoception subscale assessing sustained attention to the body[44], was negatively correlated with percentage time attending to breath (rho_{13}=-0.58, p=0.028; see **SI-Methods** and **SI-Table S5** for correlations with all subscales). Trait interoception and mindfulness subscales were not correlated with meditation period ratings or lifetime hours of meditation practice (ps>0.19). Finally, no other EMBODY metrics (number of events, mean duration, variance of duration, distraction from breath) were significantly correlated with meditation period ratings or lifetime meditation hours.

DISCUSSION

This study was the first to test the EMBODY framework, where neural data were used to identify participant-specific brain patterns associated with five types of internal attention. These unique neural patterns were then used to track moment-by-moment fluctuations in mental states during breath-focused meditation. To our knowledge, this study provides the first demonstration of an objective measure that enables continuous observation of the fluctuating mental states that occur during an individual meditation session. By making these invisible internal processes visible and quantifiable, we were able to compute novel profiles of attention during meditation,

including the percentage of time engaging in breath attention, mind wandering, or self-referential processing. Across all participants with distinguishable brain patterns (N=14/16), attention profiles indicated they engaged more with the breath vs. other states (greater percentage of time attending, mean duration of breath events, and variability of duration), which suggests they were able to implement the meditative goal of sustaining more attention to the breath. Further demonstrating criterion validity of task metrics, experienced meditators who practiced more hours of meditation in their lifetime were able to focus longer on the breath and less on self-referential processing, during the 10-minute meditation session. The attention metrics may also show specificity, as the hours of *breath*-focused meditation in particular (and not other areas of the body) predicted greater percentage time attending to the breath and decreased self-referential processing. These results support theories that meditation can decrease self-relevant thought[21], which is particularly relevant for clinical populations characterized by maladaptive self-focused emotions and thoughts (e.g., depression and anxiety)[13]. Together, these findings suggest the EMBODY framework can indeed track distinct and fluctuating mental states during meditation, which holds promise for elucidating the basic attentional mechanisms through which meditation may improve health.

Distinct fMRI brain patterns were distinguishable by MVPA for five internal attentional states, even without changes in the external visual environment while participants' eyes remained closed. That is, when given auditory instructions, participants could change their internal focus of attention and produce patterns of brain activity that were stable and distinct enough to be recognized by pattern classifiers for attending to sensations of the breath or feet, engaging in mind wandering or self-referential processing, and listening to ambient sounds. Importantly, this approach was feasible for nearly all participants including every experienced

meditator and most (75%) of the matched control participants. Notably, the neural patterns were unique for each mental state in each participant. Important voxels that contributed to accurate classification were distributed across many areas of the brain and tended to be unique for a given participant.

Although this framework emphasizes the individual differences in brain patterns for person-specific decoding of meditation, we initially characterized brain regions that may contribute to accurate classification in a higher proportion of participants. For breath-focused attention, a large cluster spanned the medial PFC (mPFC) and consisted of negative importance voxels, indicating that the average *z*-scored activation was lower than mind wandering and self-referential processing. This region included the anterior mPFC, which overlapped with a *positive* importance cluster for self-referential processing, and is a key hub of the DMN midline core that reflects self-relevant affective decisions[32]. Less mPFC activation may represent an important signal in distinguishing breath vs. self-referential processing, and is consistent with studies showing decreased mPFC activation after meditation training[29]. Surprisingly, we did not find a consistent cluster in the insula which is important for interoception[30], and this null result may stem from the lack of an external stimulus (often used as a comparison condition[29]). Self-referential processing also involved a large negative importance cluster in the dorsomedial PFC, which is typically active in the task-positive EFN. Finally, the most prevalent regions for mind wandering included bilateral superior temporal gyrus. The temporal regions may potentially be associated with the dorsal mPFC subsystem of the DMN which includes the lateral temporal cortex and temporal pole, and is thought to be involved in making judgments about present mental states[32]. Research with larger samples may further characterize group-level neural patterns that can be used to decode meditation.

Supporting construct validity of the IA task, classification accuracy of brain patterns (which indicates reliability and distinctiveness) was positively correlated with participants' subjective attention ratings, and suggests that greater classification accuracy indeed reflected better internal attention. Similar to previous research[46], these results demonstrated that neural signals may differentiate interoception to distinct areas of the body (breath vs. feet) . This approach can thus potentially be used to track attention during body-based practices such as the body scan[24,25], which may improve health through cultivating understanding of homeostatic information present throughout the body[18,19]. These findings also demonstrated that the brain pattern for mind wandering, or the "movement" from one mental state to another[47], was distinct from self-referential processing, or engaging in mental content that is personally significant[31]. With this distinction, the wandering nature of attention can be disentangled from the *contents* of what the mind wanders to, which is often self-relevant processing[31]. This framework thus allows us to decode mind wandering and self-referential states continuously through time, which improves the temporal resolution of measuring mind wandering (that typically relies on intermittent probes[33,48]), and can be a tool to decode the contents of clinically-relevant resting state activity[31,41].

The IA task yielded distinct brain patterns for every experienced meditator tested, which supports the idea that internal attention can be trained and stabilized through meditation practice. Distinct brain patterns were also found for most matched controls, which suggests that even people without formal training can sustain internal attentional focus for relatively short durations, although some may need further training to yield distinct neural patterns. Together, this holds promise for testing longitudinal effects of mindfulness-based interventions with novices, investigating group differences between meditators and novices, and eventually linking

changes in attention during meditation to clinical outcomes. The framework should be validated in larger and more diverse samples, including clinical populations that may benefit from mindfulness-based interventions. Although these early results are promising, the IA task should be developed to increase classifier performance, including optimizing trial conditions and durations, testing different MVPA algorithms, integrating psychophysiological and behavioral data, and using real-time neurofeedback[34] to communicate measurements of internal attentional states to the participants during the experiment. However, very high accuracy of these mental states is likely unrealistic due to the difficulty in maintaining focused attention (25-50% of waking moments are estimated to involve mind wandering[48]). The framework should also be validated in neuroimaging modalities where meditators can sit upright with less sound distraction, such as electroencephalography and magnetoencephalography[42,46].

Subjective ratings of attention after the meditation period did not correlate with EMBODY metrics. This may reflect that participants show less accurate introspection for longer durations (10 minutes) compared to the shorter IA task trials (16-50s). Ratings assessing longer periods of time may be influenced by peak and recent experiences[49], emphasizing the need for objective measures that continuously assess meditation practice. EMBODY metrics also did not correlate with trait questionnaires of interoception or mindfulness in the expected direction, showing a negative correlation between percentage time attending to the breath and attention regulation, or sustaining attention to the body. Trait measures may assess higher-level constructs that may not directly correspond to fluctuating attention states during a single meditation session, and participants may vary in their ability to accurately report attentional qualities[17]. Given the high cost and limited availability of fMRI-based measures, more self-report and behavioral

measures should be validated with EMBODY metrics to identify which ones can be more widely used as objectively validated research outcomes.

Overall, the initial EMBODY framework shows promising ability to distinguish unique brain patterns of internal attention, which can then be used to track mental states during meditation. Meditation practices are multi-faceted in the qualities of attention they train, what internal and external stimuli they are applied to, and how they are implemented in clinical and general populations. This may partly explain why clinical outcomes are heterogenous, and effect sizes of mindfulness-based interventions are moderate[17]. The framework can thus be adapted to measure other meditative attentional qualities (e.g., meta-awareness[33] and nonjudgment[24,25]), as well as different types of meditation practice (e.g., open monitoring[3,7] and compassion[4,6]), and be used to track person-specific skills and clinical outcomes. By developing measures to precisely assess the attentional qualities cultivated by meditation, we will gain the measurement power needed to rigorously test the mechanisms through which meditation may improve health and well-being. Finally, the EMBODY framework highlights that each individual's brain signatures and meditation practice are unique, which we hope will aid researchers and clinicians in applying precision medicine approaches[50] to design interventions that will maximally benefit individuals in targeted and specific ways.

ACKNOWLEDGEMENTS

Research was funded by the US National Institutes of Health (NIH): K08 AT009385 (H.Y.W.), T32 AT006956 (F.M.H. and S. Adler), K24 AT007827 (F.M.H.), R01 EY028746 (J.A.L.P), R01 AG049424 and R21 AG041071 (D.A.Z. and A.G.), the UCSF Mt. Zion Health Fund Pilot in Integrative Medicine Research (H.Y.W.), and Osher Center for Integrative Medicine (OCIM)

Jaswa Fund for Meditation Research (H.Y.W.). The authors thank Lara Stables, Chad Smiddy, Elizabeth Pierce, Peter Wais, Ryan Lopilato, and Sierra Niblett (UCSF Neuroscape) for MRI consultation and sequences, study management, and data collection; Patricia Moran and Stephanie Lee (UCSF OCIM) for study management, screening, and data collection; Mark Estefanos for aid in software development; and Tiffany Ho and Regina Lapate for consultation regarding study design, data analysis and interpretation.

AUTHOR CONTRIBUTIONS

H.Y.W., J.L.P, F.M.H., and N.A.S.F. conceptualized the study. All authors contributed to the data analytic strategy and interpretation. H.Y.W., S.S., and V.G. collected, processed, and analyzed data. H.Y.W., J.L.P., S.S., and V.G. developed unpublished data analysis tools. H.Y.W. wrote the manuscript, with contributions from J.L.P. and N.A.S.F. and comments from all other authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

REFERENCES

1. MacLean, K. A. *et al.* Intensive meditation training improves perceptual discrimination and sustained attention. *Psychol. Sci.* **21**, 829–839 (2010).
2. Levinson, D. B., Stoll, E. L., Kindy, S. D., Merry, H. L. & Davidson, R. J. A mind you can count on: validating breath counting as a behavioral measure of mindfulness. *Front. Psychol.* **5**, 1202 (2014).

3. Lutz, A., Slagter, H. A., Dunne, J. D. & Davidson, R. J. Attention regulation and monitoring in meditation. *Trends Cogn. Sci.* **12**, 163–169 (2008).
4. Weng, H. Y. *et al.* Compassion training alters altruism and neural responses to suffering. *Psychol. Sci.* **24**, 1171–1180 (2013).
5. Leiberg, S., Klimecki, O. & Singer, T. Short-Term Compassion Training Increases Prosocial Behavior in a Newly Developed Prosocial Game. *PLoS ONE* **6**, e17798 (2011).
6. Ashar, Y. K. *et al.* Effects of compassion meditation on a psychological model of charitable donation. *Emotion* **16**, 691–705 (2016).
7. Colzato, L. S., Szapora, A. & Hommel, B. Meditate to Create: The Impact of Focused-Attention and Open-Monitoring Training on Convergent and Divergent Thinking. *Front. Psychol.* **3**, (2012).
8. Lazar, S. W. *et al.* Meditation experience is associated with increased cortical thickness. *Neuroreport* **16**, 1893–1897 (2005).
9. Hölzel, B. K. *et al.* Mindfulness practice leads to increases in regional brain gray matter density. *Psychiatry Res. Neuroimaging* **191**, 36–43 (2011).
10. Fox, K. C. R. *et al.* Functional neuroanatomy of meditation: A review and meta-analysis of 78 functional neuroimaging investigations. *Neurosci. Biobehav. Rev.* **65**, 208–228 (2016).
11. Fox, K. C. R. *et al.* Is meditation associated with altered brain structure? A systematic review and meta-analysis of morphometric neuroimaging in meditation practitioners. *Neurosci. Biobehav. Rev.* **43**, 48–73 (2014).
12. Kang, Y., Gray, J. R. & Dovidio, J. F. The nondiscriminating heart: Lovingkindness meditation training decreases implicit intergroup bias. *J. Exp. Psychol. Gen.* **143**, 1306–1313 (2014).

13. Baer, R. A. Mindfulness Training as a Clinical Intervention: A Conceptual and Empirical Review. *Clin. Psychol. Sci. Pract.* **10**, 125–143 (2003).
14. Zeidan, F. *et al.* Mindfulness Meditation-Based Pain Relief Employs Different Neural Mechanisms Than Placebo and Sham Mindfulness Meditation-Induced Analgesia. *J. Neurosci.* **35**, 15307–15325 (2015).
15. Ma, S. H. & Teasdale, J. D. Mindfulness-based cognitive therapy for depression: replication and exploration of differential relapse prevention effects. *J. Consult. Clin. Psychol.* **72**, 31–40 (2004).
16. Carlson, L. E. *et al.* Mindfulness-based cancer recovery and supportive-expressive therapy maintain telomere length relative to controls in distressed breast cancer survivors. *Cancer* **121**, 476–484 (2015).
17. Dam, N. T. V. *et al.* Mind the Hype: A Critical Evaluation and Prescriptive Agenda for Research on Mindfulness and Meditation. *Perspect. Psychol. Sci.* **13**, 36–61 (2018).
18. Farb, N. *et al.* Interoception, contemplative practice, and health. *Conscious. Res.* 763 (2015). doi:10.3389/fpsyg.2015.00763
19. Khalsa, S. S. *et al.* Interoception and Mental Health: A Roadmap. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* (2017). doi:10.1016/j.bpsc.2017.12.004
20. Tang, Y.-Y., Hölzel, B. K. & Posner, M. I. The neuroscience of mindfulness meditation. *Nat. Rev. Neurosci.* **16**, 213–225 (2015).
21. Dahl, C. J., Lutz, A. & Davidson, R. J. Reconstructing and deconstructing the self: cognitive mechanisms in meditation practice. *Trends Cogn. Sci.* **19**, 515–523 (2015).
22. Chambers, R., Gullone, E. & Allen, N. B. Mindful emotion regulation: An integrative review. *Clin. Psychol. Rev.* **29**, 560–572 (2009).

23. Desbordes, G. *et al.* Moving Beyond Mindfulness: Defining Equanimity as an Outcome Measure in Meditation and Contemplative Research. *Mindfulness* **6**, 356–372 (2015).

24. Kabat-Zinn, J. *Full Catastrophe Living: Using the Wisdom of Your Body and Mind to Face Stress, Pain, and Illness*. (Random House Audio, 2008).

25. Gunaratana, H. *Mindfulness in Plain English*. (ReadHowYouWant.com, 2010).

26. Lutz, A., Jha, A. P., Dunne, J. D. & Saron, C. D. Investigating the phenomenological matrix of mindfulness-related practices from a neurocognitive perspective. *Am. Psychol.* **70**, 632–658 (2015).

27. Lindahl, J. R., Fisher, N. E., Cooper, D. J., Rosen, R. K. & Britton, W. B. The varieties of contemplative experience: A mixed-methods study of meditation-related challenges in Western Buddhists. *PLOS ONE* **12**, e0176239 (2017).

28. Brefczynski-Lewis, J. A., Lutz, A., Schaefer, H. S., Levinson, D. B. & Davidson, R. J. Neural correlates of attentional expertise in long-term meditation practitioners. *Proc. Natl. Acad. Sci.* **104**, 11483–11488 (2007).

29. Farb, N. A. S., Segal, Z. V. & Anderson, A. K. Mindfulness meditation training alters cortical representations of interoceptive attention. *Soc. Cogn. Affect. Neurosci.* **8**, 15–26 (2013).

30. Craig, A. D. B. How do you feel--now? The anterior insula and human awareness. *Nat. Rev. Neurosci.* **10**, 59–70 (2009).

31. Andrews-Hanna, J. R., Smallwood, J. & Spreng, R. N. The default network and self-generated thought: component processes, dynamic control, and clinical relevance. *Ann. N. Y. Acad. Sci.* **1316**, 29–52 (2014).

32. Andrews-Hanna, J. R., Reidler, J. S., Sepulcre, J., Poulin, R. & Buckner, R. L. Functional-anatomic fractionation of the brain's default network. *Neuron* **65**, 550–562 (2010).

33. Hasenkamp, W., Wilson-Mendenhall, C. D., Duncan, E. & Barsalou, L. W. Mind wandering and attention during focused meditation: a fine-grained temporal analysis of fluctuating cognitive states. *NeuroImage* **59**, 750–760 (2012).

34. Sitaram, R. *et al.* Closed-loop brain training: the science of neurofeedback. *Nat. Rev. Neurosci.* **18**, 86–100 (2017).

35. Garrison, K. A. *et al.* Real-time fMRI links subjective experience with brain activity during focused attention. *NeuroImage* **81**, 110–118 (2013).

36. Norman, K. A., Polyn, S. M., Detre, G. J. & Haxby, J. V. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn. Sci.* **10**, 424–430 (2006).

37. Haxby, J. V. *et al.* Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* **293**, 2425–2430 (2001).

38. Lewis-Peacock, J. A. & Postle, B. R. Temporary activation of long-term memory supports working memory. *J. Neurosci. Off. J. Soc. Neurosci.* **28**, 8765–8771 (2008).

39. Woo, C.-W. *et al.* Separate neural representations for physical pain and social rejection. *Nat. Commun.* **5**, (2014).

40. Lewis-Peacock, J. A. & Postle, B. R. Decoding the internal focus of attention. *Neuropsychologia* **50**, 470–478 (2012).

41. Tusche, A., Smallwood, J., Bernhardt, B. C. & Singer, T. Classifying the wandering mind: Revealing the affective content of thoughts during task-free rest periods. *NeuroImage* **97**, 107–116 (2014).

42. Zhigalov, A., Heinilä, E., Parviainen, T., Parkkonen, L. & Hyvärinen, A. Decoding attentional states for neurofeedback: Mindfulness vs. wandering thoughts. *NeuroImage* (2018). doi:10.1016/j.neuroimage.2018.10.014

43. McDuff, S. G. R., Frankel, H. C. & Norman, K. A. Multivoxel Pattern Analysis Reveals Increased Memory Targeting and Reduced Use of Retrieved Details during Single-Agenda Source Monitoring. *J. Neurosci.* **29**, 508–516 (2009).

44. Mehling, W. E. *et al.* The Multidimensional Assessment of Interoceptive Awareness (MAIA). *PloS One* **7**, e48230 (2012).

45. Baer, R. A., Smith, G. T., Hopkins, J., Krietemeyer, J. & Toney, L. Using self-report assessment methods to explore facets of mindfulness. *Assessment* **13**, 27–45 (2006).

46. Kerr, C. E. *et al.* Effects of mindfulness meditation training on anticipatory alpha modulation in primary somatosensory cortex. *Brain Res. Bull.* **85**, 96–103 (2011).

47. Christoff, K., Irving, Z. C., Fox, K. C. R., Spreng, R. N. & Andrews-Hanna, J. R. Mind-wandering as spontaneous thought: a dynamic framework. *Nat. Rev. Neurosci.* **17**, 718–731 (2016).

48. Smallwood, J. & Schooler, J. W. The science of mind wandering: empirically navigating the stream of consciousness. *Annu. Rev. Psychol.* **66**, 487–518 (2015).

49. Redelmeier, D. A. & Kahneman, D. Patients' memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures. *Pain* **66**, 3–8 (1996).

50. Collins, F. S. & Varmus, H. A New Initiative on Precision Medicine. *N. Engl. J. Med.* **372**, 793–795 (2015).

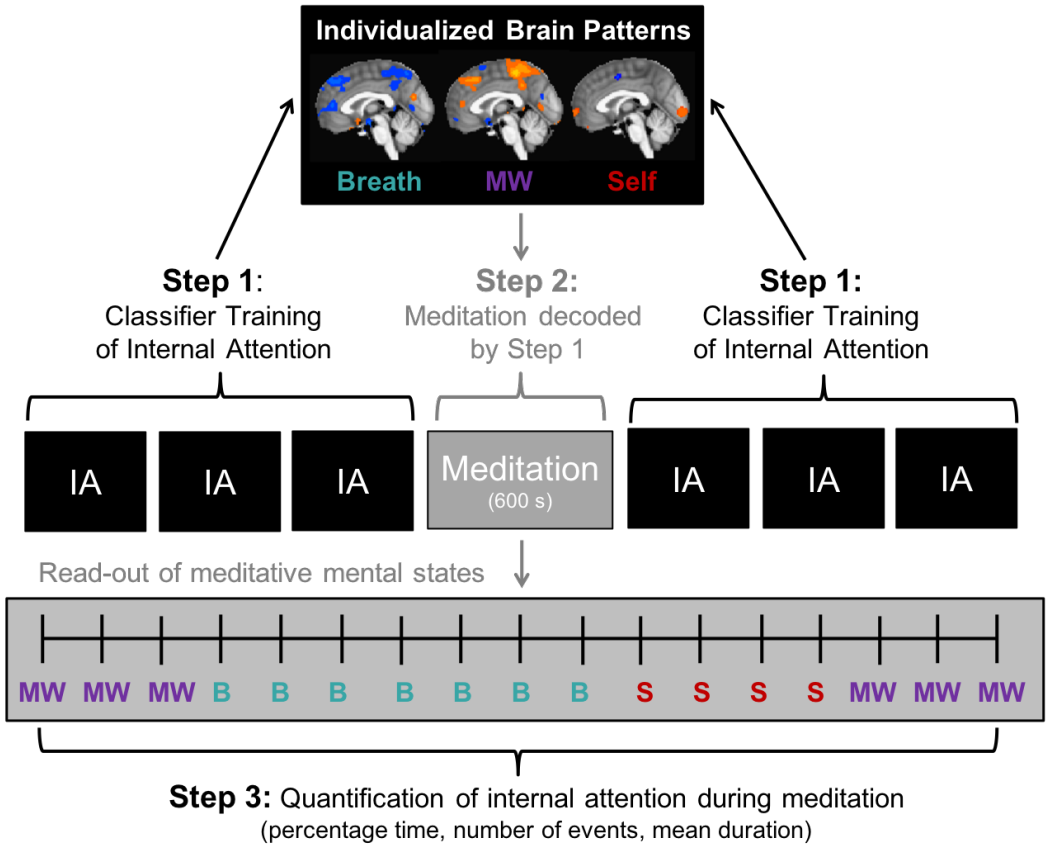

Figure 1. EMBODY Framework: Evaluating **M**ultivariate Maps of **Body** Awareness to measure internal attention states during meditation.

Step 1. Brain pattern classifier training. Machine learning algorithms are trained in fMRI neural patterns associated with internal mental states in the Internal Attention (IA) task. IA is directed via auditory instructions to pay attention with eyes closed to the breath, mind wandering, self-referential processing, and control conditions of attention to the feet and ambient sounds (see **Fig. 2**). Unique individualized brain patterns for each participant are learned using n-1 cross-validation with 6 blocks of the IA task.

Step 2. Meditation period classification. Neural patterns are collected during a 10-min meditation period (in this case, focused attention to the breath; administered in the middle of 6 IA blocks), and are decoded by multi-voxel pattern analysis (MVPA)[37] using the unique brain

patterns learned in Step 1. Meditation is characterized second-by-second into mental states of attention to breath (B), mind wandering (MW), or self-referential processing (S), producing a read-out of distinct and fluctuating mental states during meditation.

Step 3. Quantification of internal attention during meditation. From the temporal read-out of meditative mental states in Step 2, novel attention metrics during meditation can be quantified including percentage time spent in each mental state, number of times engaged in each mental state ("events"), and mean duration spent in each mental state. See **Online Methods** for details.

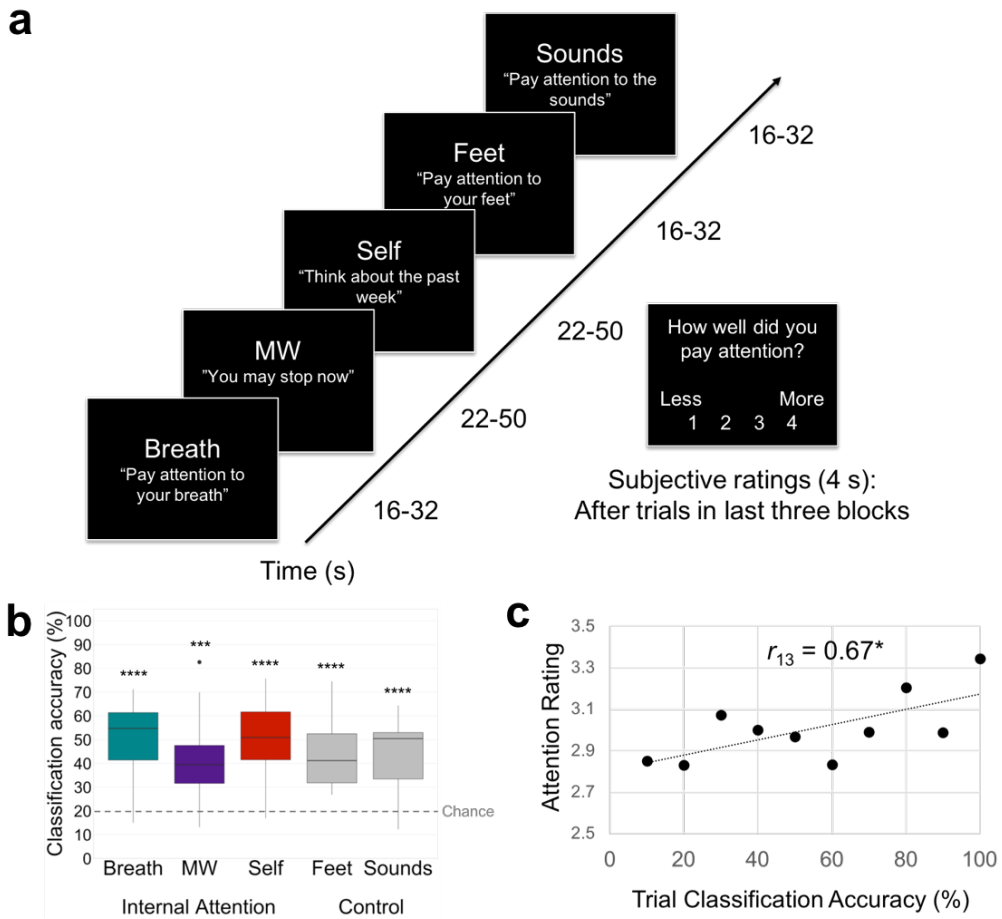

Figure 2. EMBODY Step 1: Classifier training of internal mental states. (a) Internal Attention (IA) task. With eyes closed, participants were directed via 2-s auditory instructions to pay attention to five internal mental states for brief time periods (16-50s). The IA task directed attention to three mental states relevant for breath meditation (Breath, mind wandering [MW], and self-referential processing [Self]), and to two control mental states (attention to the Feet [another area of the body] and ambient MRI Sounds [consistent external distractor]). Example auditory instructions are displayed in quotes. MW was induced by instructing participants to stop paying attention and let their minds attend to whatever they wanted. Conditions were randomized over six IA blocks in four orders, with 72s of data collected from each condition in each block. For the last half of IA task trials, subjective ratings of attention were collected after each trial

(except MW) using a button box (1 = less, 4 = more). See **Online Methods** and **SI** for full details. **(b)** From the IA task, the prediction accuracy of the classifier for identifying internal states of attending to the Breath, MW, and Self, and control conditions of attending to the Feet and Sounds. Box–whisker plots depict the median (black line), ±1.5 interquartile range (box), and minimal and maximal values (whiskers) of the mean prediction accuracy for all data in each condition (432s, TR=1s) across all subjects. Dots represent values >±1.5 interquartile range. Statistical significance was determined by a one-sample two-sided *t*-test against theoretical chance-level for classification of 5 categories (20%, denoted by dashed line). *** t_{15}=4.65, p<0.001, **** t_{S15}>5.67, ps<0.0001. **(c)** Classifier training accuracy at the trial-level was positively correlated with subjective ratings of attention (administered during the last half of IA task trials, collapsed across all conditions except MW). * p<0.05.

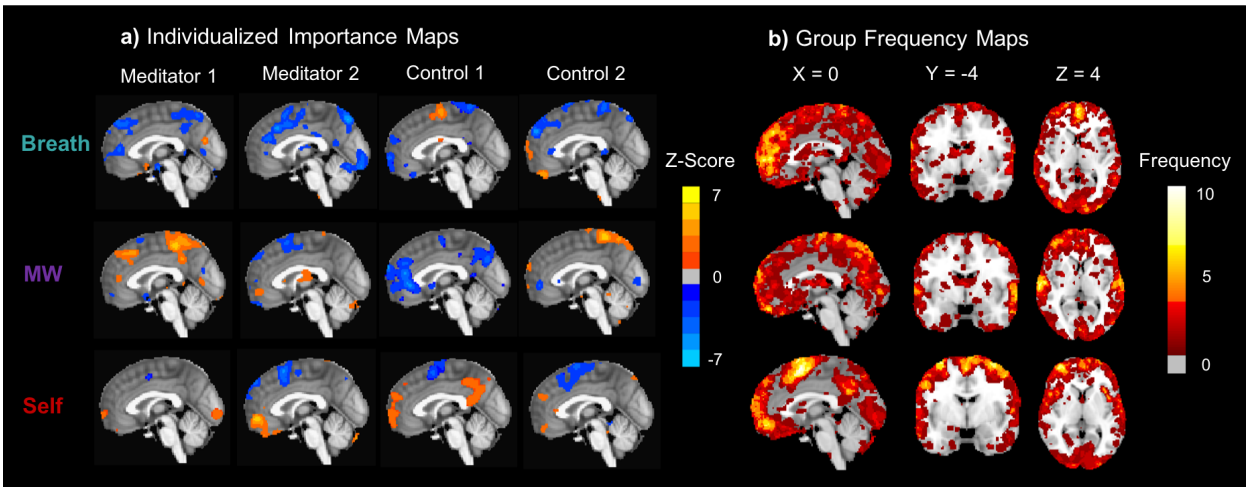

Figure 3. Classifier importance maps representing voxels that accurately distinguish internal mental states. **a)** Subject-level importance maps showing individualized brain patterns representing voxels that are important for distinguishing neural signatures of attention to the Breath, MW, and Self. For each task condition, importance values were computed by multiplying each voxel's classifier weight for predicting the condition and the average activation during the condition[43]. The maps were thresholded at ±2 SD and displayed on the MNI152 template to identify the most important voxels for each participant. Orange importance voxel indicate positive *z*-scored average activation values, and blue importance voxels indicate negative *z*-scored average activation values. **b)** For initial characterization of brain regions that supported classification and were common across individuals, group importance frequency maps indicate the number of participants for which the voxel accurately distinguished each mental state. All importance voxels were summed, irrespective of average positive or negative *z*-scored activation. Frequency maps were also computed that independently summed positive (**SI-Fig. S3a**) and negative (**SI-Fig. S3b**) *z*-scored activation voxels, as well as histograms of frequency counts (**SI-Fig. S3c-e**). Note that the maximum frequency for any importance map was 10/14.

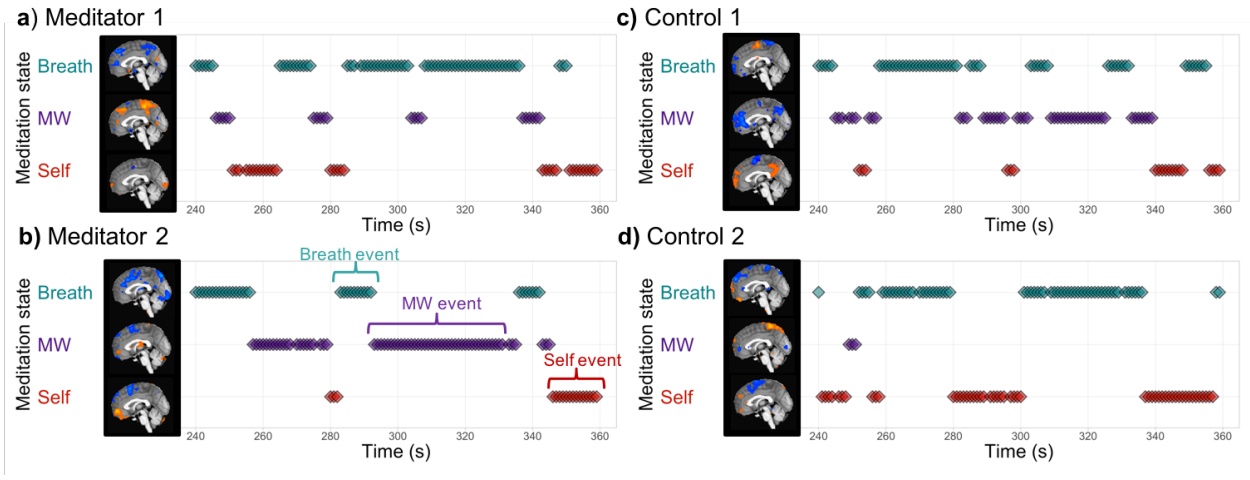

Figure 4. EMBODY Step 2: Decoding the internal focus of attention during breath-focused meditation using individualized brain patterns. Based on each participant's unique brain signatures for Breath, MW, and Self, classifier decisions were made for each time point of fMRI data (TR=1s), producing a read-out of attention states during breath meditation. The middle of the meditation period is displayed for two meditators **(a, b)** and their matched controls **(c, d)**. Mental events were quantified as 3 or more consecutive decisions from the same mental state **(b)**, and were used to compute metrics of attention during meditation in Step 3. See **Online Methods** for details.

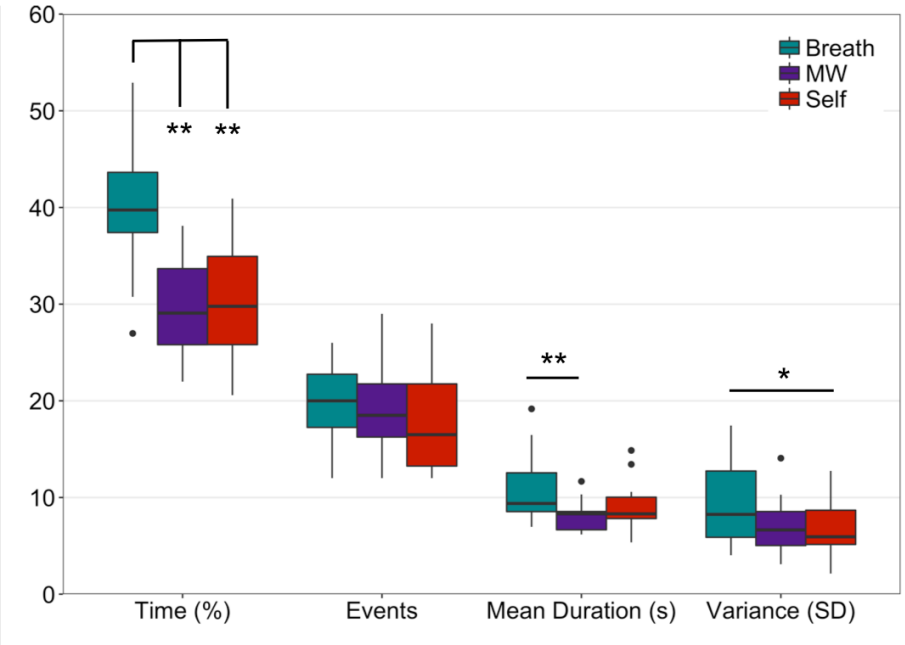

Figure 5. EMBODY Step 3: Quantification and mental state profiles of internal attention during meditation. Based on the read-out of mental states and event specification from Step 2, metrics of attention during breath meditation were quantified for each mental state: percentage time spent in each mental state (Breath, MW, or Self), the number of events, mean duration of events, and variability (standard deviation or SD) of duration of events. Box–whisker plots present the median (black line), ±1.5 interquartile range (box), and minimal and maximal values (whiskers). Dots represent values >±1.5 interquartile range. See **SI-Table S4** for full metric statistics.

* paired t_{13}=2.46 , p=0.029, after one-way ANOVA $F_{2,12}$ = 4.20, p=0.026

** paired ts$_{13}$≥3.18, ps≤0.007, after one-way ANOVA Fs$_{2,12}$≥6.47, ps≤0.005

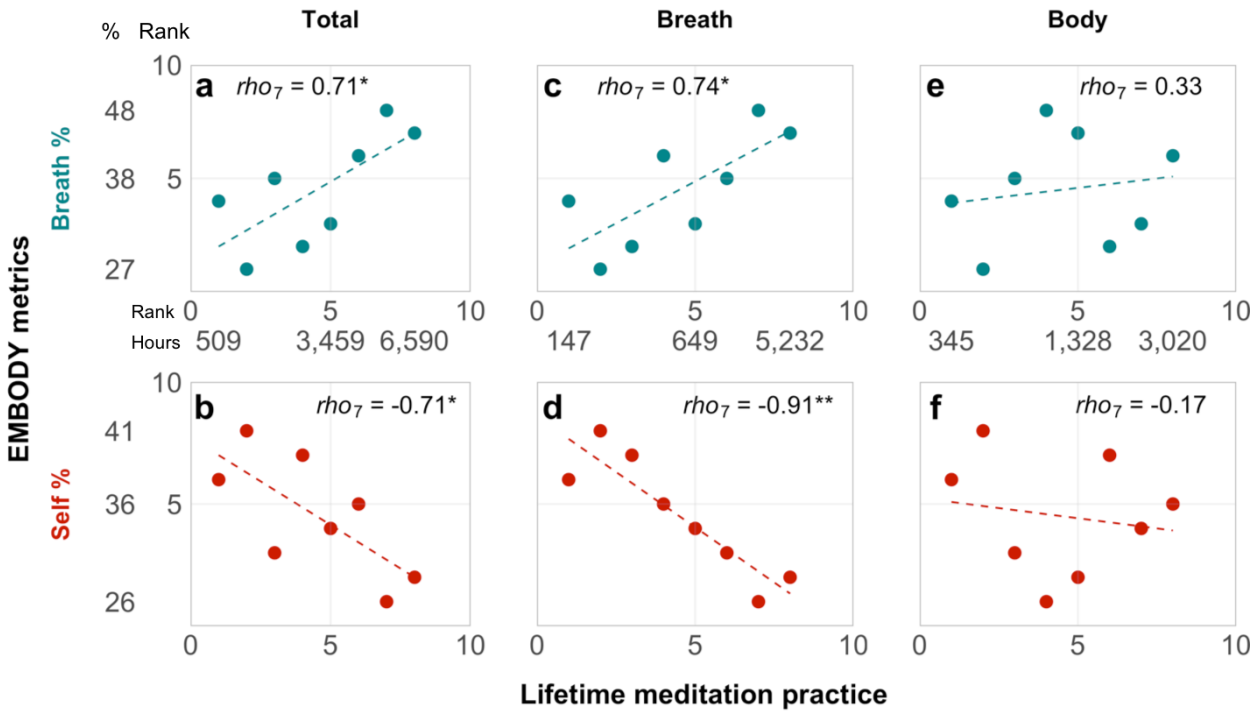

Figure 6. Mental states during meditation and lifetime meditation practice. Meditation metrics from the EMBODY Task were significantly predicted by lifetime hours of meditation practice. Total lifetime hours of meditation practice in general were associated with **(a)** greater percentage time attending to the breath and **(b)** less percentage time engaging in self-referential processing during meditation. Demonstrating specificity, lifetime hours of meditating particularly on breath sensations were associated with **(c)** greater percentage time attending to the breath and **(d)** less percentage time engaged in self-referential processing during meditation, while hours meditating on other bodily sensations were not associated with either **(e)** attending to breath or **(f)** self-referential processing. Ranks of both lifetime meditation hours and EMBODY metrics are displayed. Numerical values associated with ranks 1, 5, and 8 of each variable are displayed to aid interpretation of data.

* $p<0.05$, ** $p<0.01$

ONLINE METHODS

General framework and design rationale. The EMBODY framework was designed for decoding mental states during meditation practice at the *individual* level, that could also produce attention metrics for analysis at the *group* level (Fig. 1). We chose a number of design features to fit these purposes. Our main goal was to test the general framework in the entire sample to see whether: 1) unique brain patterns of internal attention states could be identified, 2) internal attention profiles would differ during meditation, and 3) attention metrics during meditation would correlate with subjective measures. Individuals were recruited from two distinct groups to test the framework. First, meditators were included because their experience in directing and sustaining internal attention would increase the likelihood of producing distinct neural patterns. Second, most mindfulness-based interventions study novice participants with limited meditation experience, so novice control participants were inlcuded to examine whether this framework could inform future clinical studies of meditation. Therefore, the approach was tested in individuals from both meditator and novice groups, while group-level statistics were computed for the entire sample to test construct validity and inform future research.

Participants. Participants recruited were healthy adults age 25-65 (with no medical, neurological, or psychiatric illness) who were not currently taking psychotropic medications, non-smokers, and MRI-compatible. Meditators were recruited from Bay Area meditation centers based in the Vipassana or Zen traditions (which train attention to bodily sensations) through flyers, online postings, and word of mouth. They were included if they had a consistent meditation practice (≥90 min/week) in the past 5 years, with at least 14 days of silent retreat practice, and at least half of practice included attention to breath and bodily sensations. Control

participants were recruited through flyers and online postings, and were included if they had not engaged in regular meditation practice or courses, yoga, or other mind-body practices (which was defined as >20 min at least twice weekly). Controls were age (within 5 years) and gender-matched to meditators.

Participants included 8 meditation practitioners (1 female, 1 non-binary person, 6 male, mean age = 38.4 [range 28-61], race/ethnicity: 6 White, 2 multiracial [African American/White and Asian/White]) and 8 matched novice control participants (mean age = 38.3 [range 25-63], race/ethnicity: 6 White, 1 Asian, 1 Latinx/Hispanic). Average lifetime meditation practice in meditators was 3495 hours (range 509-6590). See **SI-Table S6** for full demographics and **SI-Table S7** for meditator practice information. Two additional novices were excluded from the inability to align images due to excessive movement, and incorrect gender-matching to a meditator. All participants provided written informed consent in accordance with a protocol approved by the Institutional Review Board of the University of California, San Francisco. The study was registered at clinicaltrials.gov (NCT03344081).

Procedure. Eligibility was assessed by online questionnaire and phone interview, and eligible meditators completed the lifetime meditation practice interview over the phone. Surveys were administered online and completed within one week of the experiment. Participants were consented, trained in MRI task procedures, and then completed a 2-hour MRI protocol. They practiced the Internal Attention (IA) task, learning to direct attention to five internal states (sensations of the breath, feet, mind wandering, self-referential processing, and sounds; **Fig. 2a**). They practiced one short block of the IA task, one block with rating attention after each trial using a button box, and were given instructions for the breath-focused meditation session (~30

min total). They identified the area of the body where they felt the breath the most strongly (e.g., nose, throat, chest), and were instructed to keep their attention in that location for the remainder of the experiment. To engage in self-referential processing, participants generated 5 events from the past week, and 5 events that would occur in the next week during the training session. Experimenters ensured both meditators and novices fully understood the instructions. They were paid $65 for participation and ≤$20 for travel expenses after the experiment.

fMRI Paradigm. *Overall Framework.* The EMBODY Framework used multi-voxel pattern analysis (MVPA)[36] with fMRI data to decode the focus of internal attention during meditation in 3 steps: 1) individualized and distributed patterns of fMRI activity were identified for internal attention states relevant for breath meditation, 2) unique brain patterns from Step 1 were applied to a period of breath meditation to decode the focus of internal attention for each data point (600s), and 3) metrics of attention during meditation were computed from the decoded brain states (**Fig. 1**).

Step 1 data: Internal Attention (IA) task. fMRI data from the IA task were used to train a machine learning classifier to learn neural patterns associated with five internal mental states. To create training data that most closely resembled brain activity during meditation, participants' eyes remained closed the entire time, so the only stimulus change was their internal focus of attention. Neural patterns associated with breath, mind wandering, and self-referential processing were chosen to be most relevant for decoding the meditation period, which modeled the intended focus of attention during meditation (breath), and two common distractors from the breath (mind wandering and self-referential processing)[3,47]. Neural patterns associated with attention to feet

and awareness of ambient sounds (consistent sounds from the MRI) were chosen as control conditions to improve the classification specificity of the desired brain states in the IA task.

In the IA task, participants kept their eyes closed and received randomized 2-s auditory instructions to pay attention to 1) sensations of the breath (Breath), 2) sensations of the (Feet), 3) to stop paying attention and let their minds go wherever they would like (mind wandering or MW), 4) self-referential processing regarding the past, present, and future (Self), and 5) ambient sounds in the MRI scanner (Sounds; **Fig. 2a**). Instructions were also presented visually, which participants could briefly view if they forgot what to attend to. Six blocks of the IA task were administered, and participants were randomized to one of 4 stimulus order sets.

Each block contained 20s of baseline period (black screen) at the beginning and end, and consisted of 13 trials per block, resulting in 72s/condition within each block (balanced across 5 conditions). This yielded 432 total training data points for each condition over the experiment. Trial durations in the IA task ranged from 16-32s for attending to breath, feet, and sounds (3 trials/block each; every even-numbered trial length was randomized and administered twice/condition across the experiment), and 22-50s in MW and Self (allowing more time for MW and self-relevant modes to occur, 2 trials/block each, covering most of the duration range across the experiment). In the last three IA blocks, participants subjectively rated how well they paid attention after each trial using a button box (*How well did you pay attention?* 1 = less attention, 4 = more attention), and were encouraged to use the full range of responses. Participants also completed 1-2 blocks of a visual search task[51], which was not analyzed for this paper.

Step 2 data: Meditation Period. Participants engaged in 10 min of focused attention to the breath meditation in 2 blocks, which were in the middle of the 6 IA blocks. The meditation

period was split into two blocks (4 and 6 min) to help control participants stay engaged in the task. Participants were instructed to keep their eyes closed and pay attention to the sensations of the breath, and if their minds wandered, to return attention to the breath. For each block, they received a 6-s instruction at the beginning, and a 2-s reminder to pay attention one min before the end. After the meditation period, participants verbally rated the percentage time they paid attention to the breath and thoughts for each block.

Data acquisition. Experiments were run using E-Prime (Psychology Software Tools). Neuroimaging data were acquired with a 3 T MRI scanner (Siemens Prisma) using a 64-channel head and neck coil. We first collected scout images to align axial functional slices to the anterior commissure–posterior commissure line. A high-resolution 1×1×1 mm MPRAGE T1 anatomical scan was acquired for offline spatial registration. Functional images were acquired using a multiband gradient-echo EPI sequence[52] (2.4×2.4×2.4 mm, TE/TR = 30.2 ms/1 s, FOV=220 mm, 92×92 matrix, 56 slices, multiband acceleration=4) that covered most of the brain.

EMBODY fMRI data analyses: machine learning.

fMRI preprocessing. Data were preprocessed in AFNI[53], and were slice time corrected, aligned and motion-corrected to the first volume of the first EPI scan, and linearly de-trended in native space, respectively using 3dTshift, 3dAllineate, 3dvolreg, 3dDetrend. Subsequent pattern classification analyses were conducted using MVPA[36] (The Princeton Multi-Voxel Pattern Analysis Toolbox https://github.com/PrincetonUniversity/princeton-mvpa-toolbox), in conjunction with in-house software using Matlab (MathWorks) and Python (for post-processing of meditation period classifications in Steps 2-3).

Step 1 machine learning: Distinguishing neural patterns of internal attention. Using

preprocessed fMRI signal in native space, a pattern classifier was trained separately for each participant for trial periods from each condition (Breath, MW, Self, Feet, and Sounds; TR=1.0s, 432s/condition) using penalized logistic regression with L2 regularization and a penalty parameter of 0.01[54]. Regularization prevents over-fitting by punishing large weights during classifier training[54]. Condition labels for all classification analyses were shifted in time by 6s to account for hemodynamic lag. A binary logistic regression (1 vs. the others) was run for each of the 5 conditions, resulting in continuous classifier evidence values for each condition at each time point in the experiment (**SI-Fig. S3**). The condition that was assigned the highest evidence value yielded the categorical decision from the classifier. We evaluated classification accuracy by training on five blocks of data (fMRI task runs) and testing on the novel sixth block. The blocks used for training were then rotated, and a new block of data was tested until all six blocks of data had been classified.

Classification accuracy for each condition was computed for each participant (the percentage of accurate decisions output from the machine learning classifier). Group-level accuracy for each condition was tested with a one-sample *t*-test vs. 20% (theoretical chance level for 5 conditions), and the effect size was estimated with Cohen's *D*. Individual-level accuracy was tested with a Chi-square test determining whether the number of accurate vs. inaccurate decisions in each condition were significantly above chance levels (chance distribution is 87 accurate vs. 345 inaccurate decisions). Individuals that showed above-chance accuracy in 2/3 categories for Breath, MW, and Self conditions were used for subsequent analyses including decoding meditation states (all 8 meditators and 6 of 8 controls; **Table S2**).

IA ratings and classifier accuracy. Attention ratings were collected for the second half of trials (33/39 trials, excluding MW trials where participants were instructed to stop paying

attention and therefore no rating was administered). Within each subject, classifier accuracy was computed at the trial-level, and mean ratings were computed for each accuracy bin (10 bins from 0-100%). The subject-level mean ratings were then averaged across all subjects within each bin, and a Pearson's correlation was computed between classifier accuracy bin and mean ratings.

Individualized brain pattern importance maps. Classifier importance maps were computed for each participant using classifier weight information which identifies which voxels were most important in distinguishing neural patterns of Breath, MW, and Self[43]. This approach identified voxels with "positive importance" that had a positive weight and a positive *z*-scored average activation value (indicating that it was more active on average), and voxels with "negative importance" that had a negative weight and a negative *z*-scored average activation value (indicating that it was less active on average)[43]. For display purposes, each individual's importance values were non-linearly warped to the MNI152 2mm template using FSL[55] (FNIRT), smoothed with an 8mm Gaussian kernel, converted to *z*-scores (across voxels), and thresholded at ±2 SD to identify the most important voxels for each condition.

Group-level importance frequency maps. To identify common brain regions that contribute to accurate identifying attention to Breath, MW, and Self, individual importance maps normed to the MNI152 template were summed to produce frequency maps and displayed with FSL. Each voxel indicates the number of participants for which the voxel is important in distinguishing each mental state. Frequency maps were computed for all importance voxels, as well as positive and negative importance voxels, and frequency histograms for each map were created. Regions were identified with frequencies ≥5 or 6 (at least half of the maximum frequency found in each mental state) and a cluster extent threshold of 20 contiguous voxels (160 mm^3).

Step 2 machine learning: Decoding the internal focus of attention during breath meditation. Individualized brain patterns learned from Step 1 were applied to the 10-min meditation period to decode the internal focus of attention. The classifier was trained with all 5 mental states from the 6 blocks of the IA task, and decoded with the 3 states that were most relevant for breath-focused meditation: Breath, MW, and Self. For each data point during meditation (TR=1s, n=600, excluding data points from instruction periods), the classifier output a categorical decision of whether internal focus was on the Breath, MW, or Self (as well as continuous evidence values for each mental state). This produced a read-out of mental states during meditation over time.

To ignore spurious measurements of brain states that may fluctuate from one time point to the next, we focused our analyses on relatively stable periods. We defined a "mental event" for a given category as the classification of 3 or more consecutive time points for that category. To facilitate this, we smoothed the data such that a single incongruous decision between two events of the same type (e.g., MW event – Self decision – MW event) were relabeled according to the category of the surrounding events (e.g., Self => MW; average data points smoothed = 1.3%, SD = 0.41). Events were then quantified as 3 or more consecutive decisions of the same category, excluding any data that did not meet these criteria (average data excluded = 15.7%, SD = 4.82).

*Step 3: Quantify internal attention metrics during meditation***.** From the cleaned temporal readout of mental states during meditation from Step 2, novel metrics of internal attention during meditation were computed for each participant. For each mental state, *percentage time engaged, number of events, mean duration of events*, and *variability* (SD) of event duration were computed. The total number of mental events and mean duration of distraction from breath were

also calculated. Data were analyzed at the group level by testing for differences in metrics between conditions (Breath, MW, Self) with a one-way ANOVA. To test our main hypotheses that breath-focused meditation would result in significant differences between Breath metrics vs. the other mental states, significant results were followed up with planned pair-wise *t*-tests of Breath vs. MW and Breath vs. Self. Fluctuations between mental states were quantified by counting the number of transitions from each mental event (Breath, MW, Self) to the next mental event (Breath, MW, Self). For each mental state type, we tested the difference in the mean counts of subsequent transitions to the other mental state types (e.g., after Breath events, the difference in mean count between transitions to MW and Self), using pair-wise *t*-tests. Data were analyzed in SPSS (v. 24), figures were created with R, and brain maps were displayed using AFNI or FSLview.

Construct validity with subjective measures. To assess construct validity, EMBODY Task metrics were correlated with subjective measures of meditation period ratings, lifetime meditation hours, and trait questionnaires of interoception and mindfulness. *Meditation period ratings.* After the two meditation blocks, participants rated for each block the percentage of time attending to breath and to thoughts, and average ratings were then computed. *Lifetime meditation hours.* Through phone interview, meditators reported the years of consistent practice, the average weekly minutes of practice (individual and group), and meditation hours during retreat and monastic practice. They reported an estimate of how much practice was primarily focused on breath sensations, other bodily sensations, and other meditation practices such as lovingkindness or mantra practice (see **SI-Methods**). Based on this report, meditation practice hours were computed for total lifetime, breath, body sensations, and other practices. Due to the small sample

size, correlations with EMBODY metrics were conducted with non-parametric Spearman's *rho*. Differences between correlations were tested with the test of difference between two dependent correlations with one variable in common[56]. *Trait Questionnaires.* Trait interoception and mindfulness were measured with the Multidimensional Assessment of Interoceptive Awareness (MAIA)[44] and Five Facet Mindfulness Questionnaire (FFMQ)[45], respectively. *A priori* hypotheses were strongest for subscales that assessed sustained attention, particularly to the body. These included the MAIA Noticing and Attention Regulation subscales, and the FFMQ Observing and Acting with Awareness subscales. All other subscales were correlated with EMBODY metrics for exploratory purposes only (see **SI-Methods** and **SI-Table S5**).

Data sharing and code availability. MRI data will be available for participants who consented to share data at neurovault.org (will upload when paper is published). Code for the EMBODY Task, MVPA analysis, and post-processing are available upon request.

51. Wolfe, J. M., Cave, K. R. & Franzel, S. L. Guided search: An alternative to the feature integration model for visual search. *J. Exp. Psychol. Hum. Percept. Perform.* **15**, 419–433 (1989).
52. Auerbach, E. J., Xu, J., Yacoub, E., Moeller, S. & Uğurbil, K. Multiband accelerated spin-echo echo planar imaging with reduced peak RF power using time-shifted RF pulses. *Magn. Reson. Med.* **69**, 1261–1267 (2013).
53. Cox, R. W. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res. Int. J.* **29**, 162–173 (1996).
54. Duda, R. O., Hart, P. E. & Stork, D. G. *Pattern Classification*. (Wiley-Interscience, 2000).

55. Smith, S. M. *et al.* Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* **23 Suppl 1**, S208-219 (2004).

56. Steiger, J. H. Tests for comparing elements of a correlation matrix. *Psychol. Bull.* **87**, 245–251 (1980).