PLOS ONE

Still facial photographs of long-term meditators are perceived by naïve observers as less neurotic, more conscientious and more mindful than non-meditating controls

--Manuscript Draft--

Manuscript Number:	PONE-D-19-08913R1
Article Type:	Research Article
Full Title:	Still facial photographs of long-term meditators are perceived by naïve observers as less neurotic, more conscientious and more mindful than non-meditating controls
Short Title:	Long-term meditators perceived as less neurotic, more conscientious and more mindful than controls
Corresponding Author:	Simon B. Goldberg University of Wisconsin Madison UNITED STATES
Keywords:	Mindfulness; meditation; observer ratings; Personality; thin slices
Abstract:	The impact of meditation training on self-report psychological variables is well-established. Although meditation training is purported to have interpersonal impacts, whether naïve observers perceive differences associated with long- and short-term meditation training is largely unknown. The current study provided a stringent test of this possibility through observer ratings of a very thin slice of expressive behavior: still photographs. Photographs were drawn from a larger study investigating differences between long-term meditators (LTM) and meditation naïve participants (MNP) who were exposed to one of three experimental conditions. Photographs of ninety-nine targets (16 LTMs, 83 MNPs) were taken at baseline, prior to the randomization of MNPs to an eight-week mindfulness meditation course (mindfulness-based stress reduction; n=27), an active control comparison condition (health enhancement program; n=29), or a waitlist control group (n=27) and again after the training period. Pre- and post-intervention photographs were then rated by 25 meditation teachers and 86 undergraduate raters on five domains theoretically linked to meditation training. Results indicated that relative to MNPs, LTMs were rated as less neurotic and more conscientious, mindful, and “comfortable in their own skin” at baseline (ds=0.61 to 0.70, ps<.050), although not more agreeable or attractive. Results were largely unchanged when controlling for five observable confounds (age, gender, race/ethnicity, body mass index, attractiveness). No evidence was found supporting experimental effects of short-term meditation training on observer ratings. Thus, it seems that if meditation is associated with observable differences in facial behavior, effects may be limited to long-term training.
Order of Authors:	Simon B. Goldberg
	Matthew Hirshberg
	Lawrence Tello
	Helen Weng
	Lisa Flook
	Richard Davidson
Opposed Reviewers:	
Response to Reviewers:	Editor: 1. The ms states that, “Our study was powered to detect large to moderate differences between LTM and MNP groups and large differences between the three randomized conditions (Cohen, 1992).” Please provide more details on the a priori power analysis (effect size chosen, power level chosen, etc). I assume the power analysis was conducted for the parent study and not for this subsidiary study, with different aims and a smaller sample size. Please discuss implications of this for interpretation of the results.

RESPONSE: We have added discussion of the a priori power analysis which was conducted for the parent study on p. 10-11.

“Our study was powered to detect large to moderate differences between LTM and MNP groups and large differences between the three randomized conditions. However, a priori power calculations were for the larger trial from which these data are drawn. Specifically, the original study was powered to detect large to moderate (d = 0.74) between group effects based on prior fMRI studies. The larger trial proposed samples of 36 participants per group (total n = 144 across four groups) to allow for attrition.”

We discuss the implications of this for interpretation of results on p. 16.

“The available sample was below that for which the power analysis was conducted, which had assumed a moderate-to-large effect size (d = 0.74). Thus, it is very likely that the current study was underpowered to detect more modest effects.”

2. R1 indicates some important concerns about the use of undergraduate raters. Additionally, can undergraduates naive to mindfulness be expected to know what “mindful” means and thereby make accurate ratings on this? More broadly, since as you state neither group of raters were experts in personality, it can’t be assumed that the results say anything about what is understood scientifically to be neurotic, conscientious, etc. This limitation should be acknowledged.

RESPONSE: We appreciate these concerns and agree that our use of naïve raters is a limitation in some respects. We discuss this on p. 17.

“Neither group of raters were experts in personality, so our results merely indicate a lay interpretation of the items that were rated, rather than a scientifically-based understanding of the specific personality dimensions assessed (although the Big Five items were drawn from self-report measures used to assess personality in the general population). Further, the majority of our sample of raters were young, undergraduate students (mean age = 19.07). This group may be prone to a variety of perceiver effects that introduced bias into the ratings (e.g., associated with rating targets who are on average 30 years older), raising questions about the degree to which the ratings of photographs in the current study may generalize to interpersonal perceptions made by individuals of a wider variety of ages in the context of daily life.”

We have also clarified our rating procedure in this revision, noting that the “mindful” item was only rated by the meditation instructors (pp. 10 and 24). On p. 10:

“Across two separate samples, meditation teacher raters rated a single item for neuroticism and agreeableness, along with “comfortable in their own skin” and “mindful.””

3. Given that target group comparisons are being made on 6 outcomes, a family-wise error rate correction is indicated. Additionally, please provide exact p values for all main analyses.

RESPONSE: We now include exact p-values for all main analyses on pp. 25 and 26. We have employed Benjamini and Hochberg’s (1995) multiple comparison false discovery rate control method.

4. Please indicate whether only those MNPs providing photos both pre and post were included in analyses or whether any MNPs were lost to post-test (i.e., did not provide photos after the training period). If the latter, multilevel modeling of the data could eke out a bit more statistical power.

RESPONSE: We have clarified on p. 7 that only those MNPs providing photos both pre and post were included in analyses.

5. The LTMs reported a very wide range of practice experience (1,439 to 32,612 hrs). Given the premise that meditation experience may change what the face says about personal traits, it would be interesting to know whether accumulated practice time

correlates with the observer ratings.

RESPONSE: We agree that this is an interesting question. However, our sample of LTMs is very small (n = 16) to conduct and report this analysis. We have mentioned this possibility as a future direction (p. 16), but feel it is less appropriate to include this test in the current study.

“A future study with a larger sample of LTMs could also explore potential dose effects of meditation on observer ratings of personality across LTMs (i.e., is more training associated with more positive perceptions). The small sample of LTMs in the current study (n = 16) prohibited a proper assessment of this possibility.”

Reviewer #1:

1. Please consider tempering some of the underlying assumptions in this paper that thin slices measure reality. The introduction suggests that thin slices might tap into actual personality and I think this is a problematic assumption in this context. For example, I’m concerned about this phrase: “Observer ratings of brief excerpts of expressive behavior (i.e., “thin slices”) 16 is one method used by social psychologists to assess interpersonal perception within laboratory contexts. These observations have been shown to predict dimensions of personality, internal states, and outcomes including teaching and negotiation effectiveness using minimal samples of behavior (e.g., less than five minutes).” First, the effect sizes for most traits based on thin slices is quite small. Second, the most robust effect across all thin slices, especially for photographs, is for extraversion which isn’t assessed here. Third, impressions change over time. For example, narcissists are seen in very positive ways right off the bat but after a few interactions, people have had enough and rate them fairly negatively. Further, people's initial judgments are often based on stereotypes and perceiver effects but over time, are based more on what the person is actually like. Thus, snap judgments are made but are not rigid. Taken together, thin slices might predict an outcome or “actual” personality for some traits greater than chance but the effects are generally tiny and observed only for a narrow set of traits (e.g., observable ones like extraversion) and are not as consequential in the long run as they might seem. As such, I hope the authors consider the possibility that the thin slices are perceptions rather than reality and may or may not continue past the first impression.

RESPONSE: Thank you for these insightful comments. We have significantly modified our discussion of the thin slices paradigm along these lines in the Abstract, Introduction, and Discussion sections of our manuscript (pp. 2, 4-7, 13-15).

For example, the Introduction now includes this paragraph (p. 5):

“It is important to note that while observer rating paradigms involving thin slices of behavior at zero acquaintance are intriguing and potentially informative, they should not be interpreted as necessarily measuring reality. Such ratings show only modest correlations with self-ratings (especially for less observable traits like neuroticism, r = .08)23 and can be heavily influenced by stereotypes and other perceiver effects (i.e., rating biases within the individual making the rating).30 Further, these judgments are not necessarily stable over time, tending to become more accurate as acquaintanceship increases.31 Nonetheless, ratings at zero acquaintance do provide information about how an individual is perceived by strangers. The incremental validity of such ratings over and above more conventional self-report measures is an issue that still warrants further study.”

And the Discussion section includes this paragraph (p. 14):

“It is important to highlight that perceived differences between LTMs and MNPs were evident using a very small sample of expressive behavior (i.e., still photographs). While this may reflect the potency of potential differences (i.e., they could be detected from minimal information), the observed effect should be contextualized within the broader thin slices and zero acquaintance ratings literature. Meta-analytic evidence has highlighted the modest reliability for ratings by strangers of internal states based on still visual cues (e.g., rrr = .25 for emotional stability) as well as the low self-other agreement for ratings by strangers of internal states (e.g., r = .08 for emotional stability,

r = .12 for openness to experience). Therefore, perceived differences must be interpreted as simply that – differences in perception – and may or may not reflect differences in LTMs' and MNPs' actual internal states. This is particularly so given the small amount of information available to observers (e.g., still photographs). Again, still photographs are clearly not analogous to how interpersonal perception typically occurs within daily life – we almost always have access to considerably more information when interacting with others. However, our results nonetheless suggest that observers somehow perceived LTMs more favorably across several domains highly relevant to meditation training, despite the limited information provided."

2. Most thin slice work goes one step further than just assessing impressions – specifically, researchers code cues like facial expressions, clothing, hair, so on and then determine which cues predict specific impressions (e.g., smiling predicts perceptions of extraversion). This type of coding would help us better understand the effects. Which cues are LTM giving off that naïve participants are not? Ideally, participants provided self-perceptions so that valid cues can be identified, but I'm guessing those data are not available. In sum, I think a PLOS ONE paper should identify mechanisms for the observed impressions and finding cues that are associated with impressions is one way to find them.

RESPONSE: We appreciate this suggestion and agree that it would be quite interesting to determine specific cues that predicted the observer ratings and specifically those which differentiated the LTMs and MNPs. However, we feel this is really another study and ideally one that involves implementation of highly sophisticated methods that go well beyond the study as designed (e.g., machine learning; Mast et al., 2015, Current Directions in Psychological Science). We now include mention of this important future direction in the Discussion section (p. 15).

"It could also be illuminating to determine which aspects of behavior raters are using as cues to differentiate LTMs and MNPs or the short-term effects of training. This could be done using machine learning and social sensing technologies."

In our view, our current study seems sufficiently rigorous for PLOS ONE given our recruitment of both LTMs and MNPs, random assignment of MNPs to eight-week behavioral interventions, and assessment of images by both naïve and meditation teacher raters.

3. The authors mention why they picked the traits they did, but I was hoping for more reasoning behind those selections. For example, why were some of the Big Five but not all of them selected when trait mindfulness has been shown to be associated with openness as well?

RESPONSE: Due to the large number of images to be rated and our interest in having multiple items assessing key personality dimensions, we limited the number of personality traits we assessed. We have provided additional discussion of our rationale on pp. 9-10.

"Due to the large number of photographs and dimensions being assessed, six separate surveys were created (see Table 2). Across four separate samples, undergraduate raters rated two items for each of the three Big Five personality traits (conscientiousness, agreeableness, neuroticism) as well as attractiveness and "comfortable in their own skin." We anticipated greater difficulty recruiting meditation teachers, so planned to have them rate a subset of items in order to increase reliability of available ratings. Across two separate samples, meditation teacher raters rated a single item for neuroticism and agreeableness, along with "comfortable in their own skin" and "mindful.""

We agree that assessing openness could have been appropriate, although in our view openness is more likely a predictor of engaging in meditation practice rather than an outcome of practice. Nonetheless, we now mention our lack of assessment of extraversion and openness in the Discussion section (p. 17).

"While we focused on dimensions theoretically linked to meditation training and embodiment, we did not assess two of the five Big Five dimensions (i.e., extraversion,

openness)."

4. Going further, given that the authors are especially interested in perceptions of internal states/attributes, it would be important to note the large body of work showing that internal attributes are difficult for others to accurately observe. For example, self-perceptions of neuroticism are quite accurate (i.e., they predict objective outcomes) whereas others' impressions (new and close acquaintances) are less so (e.g., Vazire, 2010). As such, why would one expect to see internal attributes in photos, which represents one of the thinnest of all thin slice paradigms? I also encourage the researchers to consider how raters used the rating scale as a clue as to whether people were simply guessing because they had so little information – for example, if on a 1-7 scale, most of the attributes were near the midpoint (4), suggesting some ambivalence. In sum, please explain why the given traits were selected with respect to past work and theory.

RESPONSE: Thank you for this comment. Per the previous suggestion, we have expanded our discussion of thin slices paradigms and ratings made at zero acquaintance, particularly for less visible traits such as neuroticism (p. 4-6). We fully agree that we could have picked traits with higher visibility than the ones we chose. As the Reviewer noted, our selection was based on a specific interest in detecting aspects of personality and other constructs related to meditation training, rather than traits that were simply more likely to be detected by observers. However, we agree that the difficulty in rating internal states (e.g., neuroticism) at zero acquaintance was not clear in our original submission. We have now clarified this point on p. 5, including reference to both the Realistic Accuracy Model and the Self-Other Knowledge Asymmetry model.

"Within zero-acquaintance ratings, a variety of factors appear to influence the accuracy of ratings, including the type of information available (e.g., audio plus visual vs. text/electronic communication) and the trait being rated (e.g., extraversion vs. neuroticism). The Realistic Accuracy Model (RAM) provides a framework for understanding this variability, noting that accurate judgments of personality require relevant information to be available and both detected and utilized by raters in making judgments. This model helps understand why personality traits with higher visibility (e.g., extraversion) are more accurately rated at zero acquaintance than less visible traits (e.g., neuroticism), as well as why accuracy generally increases with the quantity and quality of available information (e.g., audio plus visual information yields more reliable ratings than text/electronic communication). Similarly, the Self-Other Knowledge Asymmetry (SOKA) model highlights how the accuracy of ratings varies across domains, with self-ratings showing higher accuracy than observer ratings for traits low in visibility (e.g., neuroticism)."

In our view, while the difficulty in rating internal states is important to acknowledge, it still seemed worthwhile determining whether naïve observers perceived LTM or MNP following meditation training differently on these dimensions. From our perspective, the difficulty rating these aspects of personality exerts a conservative rather than liberal bias on our ability to detect between-group differences (assuming such differences exist to be detected). However, we agree with the Reviewer's previous point that these ratings should still not be interpreted as reflecting reality, although they may reflect a difference in interpersonal perceptions of internal states.

5. It wasn't clear to me why photographs were selected or why photographs are somehow superior or address limitations of short video clips (i.e., past work). Further, the paper assumes that these impressions have important interpersonal effects, but I think this might go too far. In everyday life, we rarely ever get a photograph without some other additional information about someone. For example, online networks (e.g., Facebook, online dating apps) all come with extra information about a person, so when are photos alone helpful? I'm a fan of thin slice work, but in many ways, the experiment here reflects a situation that rarely happens in real life. Please consider highlighting this point.

RESPONSE: We agree that most interpersonal impressions are based on much more information than that provided in a photograph. Photographs were selected for both logistical reasons as well as an interest in pushing the lower bounds of interpersonal perception by examining ratings of very thin slices of behavior. Again, we believe that

severely limiting the amount of information available would have exerted a conservative rather than liberal bias in terms of our ability to detect between-group differences. At once, we agree our use of limited information raises questions regarding the validity of ratings as measures of reality, rather than simply interpersonal perceptions. We include discussion of the points raised here on pp. 5-6, and 14-15.

As stated on p. 5:

"It is important to note that while observer rating paradigms involving thin slices of behavior at zero acquaintance are intriguing and potentially informative, they should not be interpreted as necessarily measuring reality. Such ratings show only modest correlations with self-ratings (especially for less observable traits like neuroticism, r = .08) and can be heavily influenced by stereotypes and other perceiver effects (i.e., rating biases within the individual making the rating). Further, these judgments are not necessarily stable over time, tending to become more accurate as acquaintanceship increases. Nonetheless, ratings at zero acquaintance do provide information about how an individual is perceived by strangers. The incremental validity of such ratings over and above more conventional self-report measures is an issue that still warrants further study."

On p. 6:

"The current study sought to address these limitations and explore the boundaries of information required to detect meditation-related differences through second-person observation. We employed one of the thinnest slices of expressive behavior – a still photograph – which provided a more stringent test of the lower limits of information for detecting correlates of meditation training. Relative to videos, photographs are less unlikely to provide potentially confounding contextual cues as to whether an individual engages in meditation practice (e.g., based on the language they use or the topics they discuss)."

On pp. 14-15:

"It is important to highlight that perceived differences between LTMs and MNPs were evident using a very small sample of expressive behavior (i.e., still photographs). While this may reflect the potency of potential differences (i.e., they could be detected from minimal information), the observed effect should be contextualized within the broader thin slices and zero acquaintance ratings literature. Meta-analytic evidence has highlighted the modest reliability for ratings by strangers of internal states based on still visual cues (e.g., rrr = .25 for emotional stability) as well as the low self-other agreement for ratings by strangers of internal states (e.g., r = .08 for emotional stability, r = .12 for openness to experience). Therefore, perceived differences must be interpreted as simply that – differences in perception – and may or may not reflect differences in LTMs' and MNPs' actual internal states. This is particularly so given the small amount of information available to observers (e.g., still photographs). Again, still photographs are clearly not analogous to how interpersonal perception typically occurs within daily life – we almost always have access to considerably more information when interacting with others. However, our results nonetheless suggest that observers somehow perceived LTMs more favorably across several domains highly relevant to meditation training, despite the limited information provided."

6. There were a few aspects of the methodology that I think need additional clarification. First, please explain the rater questionnaires and procedures in more detail. The ratings were collapsed across types (undergrad, meditators) but it looks like there was little overlap in terms of who rated what. Please be clearer about who rated what and why everyone didn't rate all traits. Second, what was the situation that preceded the photographs? It would be important to know any systematic differences in what procedurally came before the photos across groups. Third, the demographically matched language should be described further. Specifically, why were the key demographic variables selected? I think variables related to the DV are important – specifically variables that lead to different impressions that might be confounds. For example, gender is often a variable that is associated with different trait ratings (e.g., women are seen as more neurotic than men), so was gender a matching variable?

RESPONSE: Details regarding the rater questionnaires and procedures have been added on pp. 9-10.

Details regarding the photograph data collection procedure have been added on p. 9. Of note, both LTMs and MNPs had identical study visits conducted by research assistants blind to condition (i.e., LTM vs. MNP and random assignment for MNP) and study hypotheses.

“Prior to randomization, LTMs and MNPs were photographed by a research assistant blind to study condition (i.e., LTM vs. MNP). MNPs were photographed again post-intervention. Photographs were taken as part of study visits that were equivalent for LTMs and MNPs.”

We have clarified the demographic matching procedure on p. 7, including matching on gender.

“Demographic matching within the larger trial was focused on age and gender due to the association between these variables with a variety of neurobiological and psychological variables.”

7. Please eliminate all causal language. In several places, it is assumed that meditation caused differences in impressions, but because this wasn’t a true experiment, there could have been actual personality differences in who mediated (LTM) or other factors that influenced impressions. For example, LTMs might represent a group of people who are more committed to goals (i.e., their personality caused them to meditate more than other people who might have started meditation but stopped) and this is what led to differences in impressions rather than mediation in particular.

RESPONSE: We appreciate this point and fully agree that we cannot imply long-term meditation training caused differences in perceptions of LTMs vs. MNPs. We have modified our language in several places in the manuscript to reflect this.

For example, on p. 13 we now state:

“Results supported the possibility that long-term meditation training may be linked with perceived differences in neuroticism and conscientiousness by strangers despite minimal information. Significant differences in the moderate range were observed on ratings of these personality traits, as well as on novel items intended to assess constructs specific to meditation (comfortable in their own skin, mindful).”

8. Further, the discussion is misleading when discussing how robust results were across groups – please avoid trying to sell the paper and be humbler about the effects.
RESPONSE: We have modified the Discussion section in an effort to not sell our results. In particular, we have more clearly highlighted the limitations of our study design and have added additional context drawn from the broader thin slices and zero acquaintance literatures.

For example, p. 14 now includes:

“The ways in which these perceptions made from still photographs may relate to perception in daily life is not clear from the current study, of course. Nonetheless, it is possible that LTMs may experience a more welcoming interpersonal environment, vis-à-vis others’ initial impressions. Should that be the case, improved interpersonal interactions could be both an outcome of long-term practice as well as a mechanism through which long-term practice yields benefits in other domains (e.g., well-being, quality of life).”

9. Also, when discussing the impact of findings, please do not assume that, for example, impressions largely derived from 20 year olds would have a major impact on 50 year olds. As mentioned above, the experimental paradigm is one rarely experienced in daily life (i.e., photos without context), so it is unclear how these impressions alone would influence people’s lives. A complete aside, these raters saw people in this study as fairly unattractive (mean is well below the midpoint) – this is

likely because 20 year olds (the majority of raters) don't appreciate the effect of an additional 30 or so years on one's looks and style! Perhaps consider what it means when fairly young individuals are rating older individuals when discussing the results. Anyways, I think it is more appropriate to talk about the social cognition aspect of the work rather than real life consequences that are never measured and would likely be extremely small if they were. In sum, the project is interesting in its own right and it is not necessary to give overblown summaries or possible impacts to get people excited about it.

RESPONSE: This is an important point and we appreciate you raising it. We now discuss this specific limitation on p. 17:

"Further, the majority of our sample of raters were young, undergraduate students (mean age = 19.07). This group may be prone to a variety of perceiver effects that introduced bias into the ratings (e.g., associated with rating targets who are on average 30 years older), raising questions about the degree to which the ratings of photographs in the current study may generalize to interpersonal perceptions made by individuals of a wider variety of ages in the context of daily life."

10. When discussing power analyses, please indicate what you mean when you say a moderate effect (e.g., d = .20? .40?). People do not always agree about what a moderate effect might be. Also, be clear as to whether these analyses were considered after data collection – if the data were collected before this project was considered, simply stating that these analyses are exploratory and were not pre-registered allows the reader to take into account researcher degrees of freedom. This would not undermine the integrity of the authors or their work; rather, it increases the transparency of the project. Speaking of transparency, please report all measures that were assessed and any procedures about the larger project this study was part of. I am not recommending making data beyond the results here available, but it would help me/others evaluate the work if all measures and procedures are available.

RESPONSE: Additional details regarding the power analysis and the specific effect size used for power analysis for the larger study has been added on p. 10.

"However, a priori power calculations were for the larger trial from which these data are drawn. Specifically, the original study was powered to detect large to moderate (d = 0.74) between group effects based on prior fMRI studies. The larger trial proposed samples of 36 participants per group (total n = 144 across four groups) to allow for attrition."

We have added additional details regarding our planning for this study, our lack of pre-registration, and additional measures that were collected (p. 11). Of note, the between-group comparisons on personality and embodiment items were planned a priori.

"The larger trial from which these data were drawn was not pre-registered. At the time the larger trial was funded (2008), pre-registration was not widespread. A variety of self-report, behavioral, and neuroimaging data were collected as part of the larger trial and are not reported here. Similarly, the data collected via observer ratings reported here were not pre-registered. However, between-group comparisons (i.e., LTMs vs. MNPs, MBSR vs. HEP vs. waitlist) on personality dimensions and embodiment items were planned a priori. Along with rating attractiveness, raters also assessed "old" and "healthy" as two additional potential confounds, but those data are not reported here. An additional rating task was completed by a subsample of raters in which observers were shown a pre- and post-test photograph and were asked to choose which occurred following a well-being intervention. There was no indication that observers were able to predict post-intervention photographs above chance (overall, or for the active groups [MBSR, HEP] specifically). These data are not reported here."

Reviewer #2:

1. It is not clear why and how meditation would affect personality impressions based on just a picture. A couple of assumptions should be made to expect such effects: 1. meditation (or mindfulness training) should causally affect these personality traits. More

literature in support should be provided. 2. These personality traits should somehow be expressed in the face (again, any support?) 3. These personality traits could then be accurately perceived based on a picture. Is there any evidence for this? As far as I know, findings on this issue are mixed at best, with some studies finding low to zero correlations between personality ratings based on picture and personality of the target. Some research also indicates that if anything, ratings are most reliable for extraversion, but not for the other dimensions (see Walker & Vetter, JPSP, 2016, for a discussion of previous findings). Thus, although the question is interesting in an intuitive sense, I was not convinced by the theoretical rationale guiding the predictions – or better put, the study seems to lack a clear theoretical basis.

RESPONSE: We appreciate these concerns. We have modified our Introduction to provide a clearer rationale for our study, while at once noting the mixed literature on thin slices, particularly in regards to the personality dimensions assessed and the ability to assess these dimensions using a photograph (pp. 3-7). Towards this end, we have drawn more heavily from the theoretical and empirical literature related to thin slices paradigms.

For example, on p. 3:

“Meditation training has been linked via meta-analysis to lower neuroticism and higher attention and empathy. Further, dispositional mindfulness, a psychological construct purported to be cultivated in the context of various forms of meditation training (e.g., mindfulness meditation), is correlated with several aspects of personality including neuroticism (r = -.45), conscientiousness (r = .32), and agreeableness (r = .22).”

And on pp. 4-5:

“One widely used observer rating paradigm involves ratings made by strangers (i.e., at zero acquaintance), often through brief excerpts or “thin slices” of expressive behavior. Ratings made by individuals who have more information about targets (e.g., family members, coworkers) tend to outperform ratings at zero acquaintance (e.g., when predicting self-ratings of personality). However, zero-acquaintance ratings have high ecological validity: in theory, they reflect how an individual is perceived by strangers, a key aspect of interpersonal perception in daily life. Further, from the perspective of evolutionary psychology, the ability to detect personality traits at zero acquaintance has clear adaptive utility (e.g., detection of emotional instability). Within zero-acquaintance ratings, a variety of factors appear to influence the accuracy of ratings, including the type of information available (e.g., audio plus visual vs. text/electronic communication) and the trait being rated (e.g., extraversion vs. neuroticism). The Realistic Accuracy Model (RAM) provides a framework for understanding this variability, noting that accurate judgments of personality require relevant information to be available and both detected and utilized by raters in making judgments. This model helps understand why personality traits with higher visibility (e.g., extraversion) are more accurately rated at zero acquaintance than less visible traits (e.g., neuroticism), as well as why accuracy generally increases with the quantity and quality of available information (e.g., audio plus visual information yields more reliable ratings than text/electronic communication). Similarly, the Self-Other Knowledge Asymmetry (SOKA) model highlights how the accuracy of ratings varies across domains, with self-ratings showing higher accuracy than observer ratings for traits low in visibility (e.g., neuroticism).”

2. A related concern: Do the results reflect a general positive impression (e.g. faces of ltm’s looked more happy?), or do they actually reflect higher levels of mindfulness, comfort, and conscientiousness, and lower levels of neuroticism? If the authors have self-report (and/or partner-report) ratings of these big five traits in the target, it would become more interesting: for example, are long-term meditators and post-MBSR participants indeed more conscientious (etc.) and could observers accurately perceive this?

RESPONSE: We appreciate the Reviewer’s question and agree that it is an interesting possibility that the raters may have reflected a generally positive impression rather than actually higher levels of the constructs assessed. However, the fact that differences were not observed on agreeableness or attractiveness, and generally persisted after

controlling for attractiveness seems to us to argue against this possibility.

Including self-report measures is another interesting idea. Although, the primary rationale for the study was to determine if there were differences in interpersonal perception associated with meditation training (long- or short-term). From this standpoint and based on the notion that self-report is prone to a variety of biases (e.g., social desirability), it seems to us less relevant/appropriate to include self-report.

In response to this comment and several made by Reviewer 1, we have modified the interpretation of our results to highlight that they reflect differences in perception rather than differences in reality (p. 13).

“Results supported the possibility that long-term meditation training may be linked with perceived differences in neuroticism and conscientiousness by strangers despite minimal information.”

3. In the introduction, the authors discuss one previous relevant study that showed that long-term meditators, and meditators post-retreat, were rated as more happy based on a brief videoclip. One of the goals of the current research is to see whether ratings based on just a picture would yield similar effects. Again, the authors do not really give a very good reason why this would be interesting. If the argument is that the study has interpersonal implications: in real life, one usually does not have a static imagine of one’s interaction partner. I’m raising this issue because, in the discussion, the authors somewhat surprisingly propose that future research investigates larger samples of behavior by using videos. What the authors present in the introduction as a strength and unique contribution of the study (in comparison with previous work that used videos), is here identified as a potential limitation. This was a bit confusing.

RESPONSE: Thank you for highlighting this point. We have provided a clearer rationale for the study’s use of still photographs in the Introduction section (pp. 3-6). For example, on p. 6:
“The current study sought to address these limitations and explore the boundaries of information required to detect meditation-related differences through second-person observation. We employed one of the thinnest slices of expressive behavior – a still photograph – which provided a more stringent test of the lower limits of information for detecting correlates of meditation training. Relative to videos, photographs are less unlikely to provide potentially confounding contextual cues as to whether an individual engages in meditation practice (e.g., based on the language they use or the topics they discuss).”

We agree that still photographs are likely not the best representation of interpersonal perception in everyday life – this point is now made on p. 16.

“Our use of photographs may have obscured changes detectable through a thicker slice of behavior (e.g., video recordings) or behaviors with higher relevance to the constructs of interest.”
However, it is important to note here that our use of such a thin slice of behavior arguably exerts a conservative rather than liberal bias on the likelihood of detecting differences in perceptions associated with meditation training.

We agree that noting the use of videos as a future direction is confusing after having framed our motivation for the current study around determining the thinnest necessary slice for detecting between-group differences. We have modified this portion of the Discussion section to highlight several specific future directions based on the broader interpersonal perception literature (e.g., samples of behavior drawn from contexts with higher trait-relevance, identification of specific aspects of behavior that are linked to between-group differences). In our view, these are important future directions that do not overlap with the existing literature (e.g., Choi et al., 2012). These seem especially important to discuss given our study failed to find observer-rated differences after a short period of meditation training (the portion of the study that could have provided causal evidence). It seems plausible that a thicker slice of behavior, particularly if sampled in in a trait-relevant context, may be a promising approach to detecting the interpersonal effects of meditation training.

	4. In short: I believe that the paper could be strengthened by placing the current findings more strongly into the broader literature on face perception and accuracy. Without a better theoretical background framework, it is difficult to assess the value and meaning of the current findings. RESPONSE: Thank you for these suggestions. We have substantially expanded our discussion of the broader literature on interpersonal perception through thin slices of behavior on pp. 4-6 in the hopes of providing a clearer context and rationale for the current study.
Additional Information:	
Question	**Response**
Financial Disclosure Enter a financial disclosure statement that describes the sources of funding for the work included in this submission. Review the submission guidelines for detailed requirements. View published research articles from *PLOS ONE* for specific examples. This statement is required for submission and **will appear in the published article** if the submission is accepted. Please make sure it is accurate. **Unfunded studies** Enter: *The author(s) received no specific funding for this work.* **Funded studies** Enter a statement with the following details: • Initials of the authors who received each award • Grant numbers awarded to each author • The full name of each funder • URL of each funder website • Did the sponsors or funders play any role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript? • **NO** - Include this sentence at the end of your statement: *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.* • **YES** - Specify the role(s) played. * typeset	This work was supported by a grant from the National Center for Complementary and Integrative Health (NCCIH) P01AT004952 to Richard J. Davidson and a grant from the Mind & Life Institute (Varela Award) to Simon B. Goldberg. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.
Competing Interests Use the instructions below to enter a	Richard J. Davidson is the founder, president, and serves on the board of directors for the non-profit organization, Healthy Minds Innovations, Inc.

competing interest statement for this submission. On behalf of all authors, disclose any competing interests that could be perceived to bias this work—acknowledging all financial support and any other relevant financial or non-financial competing interests. This statement **will appear in the published article** if the submission is accepted. Please make sure it is accurate. View published research articles from *PLOS ONE* for specific examples. **NO authors have competing interests** Enter: *The authors have declared that no competing interests exist.* **Authors with competing interests** Enter competing interest details beginning with this statement: *I have read the journal's policy and the authors of this manuscript have the following competing interests: [insert competing interests here]* * typeset	
Ethics Statement Enter an ethics statement for this submission. This statement is required if the study involved: • Human participants • Human specimens or tissue • Vertebrate animals or cephalopods • Vertebrate embryos or tissues • Field research Write "N/A" if the submission does not require an ethics statement. General guidance is provided below. Consult the submission guidelines for detailed instructions. **Make sure that all information entered here is included in the**	This study was approved by the University of Wisconsin ED/SBS IRB (2015-1437). Written consent was obtained from participants.

Methods section of the manuscript. **Format for specific study types** **Human Subject Research (involving human participants and/or tissue)** • Give the name of the institutional review board or ethics committee that approved the study • Include the approval number and/or a statement indicating approval of this research • Indicate the form of consent obtained (written/oral) or the reason that consent was not obtained (e.g. the data were analyzed anonymously) **Animal Research (involving vertebrate animals, embryos or tissues)** • Provide the name of the Institutional Animal Care and Use Committee (IACUC) or other relevant ethics board that reviewed the study protocol, and indicate whether they approved this research or granted a formal waiver of ethical approval • Include an approval number if one was obtained • If the study involved *non-human primates*, add *additional details* about animal welfare and steps taken to ameliorate suffering • If anesthesia, euthanasia, or any kind of animal sacrifice is part of the study, include briefly which substances and/or methods were applied **Field Research** Include the following details if this study involves the collection of plant, animal, or other materials from a natural setting: • Field permit number • Name of the institution or relevant body that granted permission	
Data Availability Authors are required to make all data underlying the findings described fully available, without restriction, and from the time of publication. PLOS allows rare exceptions to address legal and ethical concerns. See the PLOS Data Policy and FAQ for detailed information.	Yes - all data are fully available without restriction

A Data Availability Statement describing where the data can be found is required at submission. Your answers to this question constitute the Data Availability Statement and **will be published in the article**, if accepted. **Important:** Stating 'data available on request from the author' is not sufficient. If your data are only available upon request, select 'No' for the first question and explain your exceptional situation in the text box. Do the authors confirm that all data underlying the findings described in their manuscript are fully available without restriction?	
Describe where the data may be found in full sentences. If you are copying our sample text, replace any instances of XXX with the appropriate details. • If the data are **held or will be held in a public repository**, include URLs, accession numbers or DOIs. If this information will only be available after acceptance, indicate this by ticking the box below. For example: *All XXX files are available from the XXX database (accession number(s) XXX, XXX.*). • If the data are all contained **within the manuscript and/or Supporting Information files**, enter the following: *All relevant data are within the manuscript and its Supporting Information files.* • If neither of these applies but you are able to provide **details of access elsewhere**, with or without limitations, please do so. For example: *Data cannot be shared publicly because of [XXX]. Data are available from the XXX Institutional Data Access / Ethics Committee (contact via XXX) for researchers who meet the criteria for access to confidential data.* *The data underlying the results*	All relevant data are within the manuscript and its Supporting Information files.

presented in the study are available from (include the name of the third party and contact information or URL). • This text is appropriate if the data are owned by a third party and authors do not have permission to share the data. * typeset	
Additional data availability information:	

University of Wisconsin--Madison
The School of Education

Department of Counseling Psychology

335 Education Building
1000 Bascom Mall
Madison, Wisconsin 53706-1326

Department Phone (608) 263-9503
Fax (608) 265-3347
E-Mail: counpsych@education.wisc.edu

June 17th, 2019

Dear Dr. Brown,

Thank you and the reviewers for your careful evaluation of our manuscript titled " Still facial photographs of long-term meditators are perceived by naïve observers as less neurotic, more conscientious and more mindful than non-meditating controls " (Manuscript PONE-D-19-08913). We are appreciative of the offer to address your and the Reviewers' concerns in a revision. We have carefully reviewed each of the comments provided and have accordingly made several changes that are outlined below. In particular, we have significantly modified the Introduction section of our manuscript, drawing more heavily on the thin slices literature, clarifying the limitations of this paradigm to infer information about internal states, and providing justification for our use of this methodology. We have added additional details regarding our sample and assessment procedures and have worked to more fully contextualize and acknowledge the limitations of our approach. We are grateful for the many helpful suggestions that you and the Reviewers provided.

Thank you for your continued consideration of this manuscript. We look forward to your decision.

Sincerely,

Simon Goldberg, PhD
Assistant Professor
Department of Counseling Psychology
University of Wisconsin - Madison
335 Education Building
1000 Bascom Mall
Madison, WI, 53703
sbgoldberg@wisc.edu

Still facial photographs of long-term meditators are perceived by naïve observers as less neurotic, more conscientious and more mindful than non-meditating controls

Simon B. Goldberg[1,2]*, Matthew Hirshberg[2], Lawrence Y. Tello[2,3], Helen Y. Weng[4], Lisa Flook[2,5], & Richard J. Davidson[2,6,7]

[1]Department of Counseling Psychology, University of Wisconsin – Madison, Madison, WI, USA

[2]Center for Healthy Minds, University of Wisconsin – Madison, Madison, WI, USA

[3]National Eye Institute, National Institutes of Health, Bethesda, MD, USA

[4]Osher Center for Integrative Medicine, University of California, San Francisco, San Francisco, CA, USA

[5]Learning Policy Institute, Palo Alto, CA, USA

[6]Department of Psychology, University of Wisconsin – Madison, Madison, USA

[7]Department of Psychiatry, University of Wisconsin – Madison, Madison, USA

*Correspondence to: Simon B. Goldberg, Department of Counseling Psychology, University of Wisconsin – Madison, 335 Education Building, 1000 Bascom Mall, Madison, WI, 53706

sbgoldberg@wisc.edu

Abstract

The impact of meditation training on self-report psychological variables is well-established. Although meditation training is purported to have interpersonal impacts, whether naïve observers perceive differences associated with long- and short-term meditation training is largely unknown. The current study provided a stringent test of this possibility through observer ratings of a very thin slice of expressive behavior: still photographs. Photographs were drawn from a larger study investigating differences between long-term meditators (LTM) and meditation naïve participants (MNP) who were exposed to one of three experimental conditions. Photographs of ninety-nine targets (16 LTMs, 83 MNPs) were taken at baseline, prior to the randomization of MNPs to an eight-week mindfulness meditation course (mindfulness-based stress reduction; n=27), an active control comparison condition (health enhancement program; n=29), or a waitlist control group (n=27) and again after the training period. Pre- and post-intervention photographs were then rated by 25 meditation teachers and 86 undergraduate raters on five domains theoretically linked to meditation training. Results indicated that relative to MNPs, LTMs were rated as less neurotic and more conscientious, mindful, and "comfortable in their own skin" at baseline (ds=0.61 to 0.70, ps<.050), although not more agreeable or attractive. Results were largely unchanged when controlling for five observable confounds (age, gender, race/ethnicity, body mass index, attractiveness). No evidence was found supporting experimental effects of short-term meditation training on observer ratings. Thus, it seems that if meditation is associated with observable differences in facial behavior, effects may be limited to long-term training.

Keywords: mindfulness; meditation; observer ratings; personality; thin slices

There has been a dramatic increase in interest in meditation over the past several decades. Recent data from the National Health Interview Survey showed a three-fold increase in past year meditation use between 2012 and 2017 in the United States (4.1% to 14.3%).[1] Scientific interest in the topic has also increased exponentially. Experimental data from randomized controlled trials (RCTs) have shown reductions in psychiatric symptoms and improvements in well-being in the context of relatively brief (e.g., eight-week) meditation interventions, in both clinical[2,3] and non-clinical populations.[4] A related body of research has examined differences between individuals with extensive meditation experience (i.e., long-term meditators [LTM]) and meditation naïve participants (MNP). Relative to MNPs, LTMs have often shown lower psychiatric symptoms and higher well-being[5] along with a host of biological differences indicating less reactive stress physiology and improved emotion and attention regulation,[6–9] although not always.[10] Meditation training has been linked via meta-analysis to lower neuroticism and higher attention and empathy.[11] Further, dispositional mindfulness, a psychological construct purported to be cultivated in the context of various forms of meditation training (e.g., mindfulness meditation),[12,13] is correlated with several aspects of personality including neuroticism (r = -.45), conscientiousness (r = .32), and agreeableness (r = .22).[14]

Theories on the effects of meditation practice claim that they will be embodied and impact both psychological and physiological variables.[15] Although the effects of meditation practice impacting psychological variables as reported by the practitioner have been well-established,[2,5] little research has examined how meditation practice may impact interpersonal perceptions by others, including close relationships and strangers. Theories suggest that meditation practice enhances a sense of interconnectedness, and may also impact others through greater interpersonal connection and prosocial behavior.[16–20] Most research has been conducted

on first-person reports of effects, and rarely report on second-person observer effects, to examine whether the effects of meditation practice can be detectible by others.[21] However, interpersonal perception through second-person observers is important, both as a potential indicator of internal states (e.g., mood, well-being) as well as a proxy for how an individual may be perceived in daily life. Moreover, being perceived more favorably by others (e.g., as happier) may lead to more positive expectations of social interactions,[22] which could in turn produce beneficial inter- and intrapersonal effects. Thus, the domain of interpersonal perception may be one *in* which and *through* which beneficial effects of meditation practice appear. We tested the possibility that meditation training may be associated with differences in second-person observer effects with a stringent criterion, using a thin slice of behavior from still photographs and observers with no familiarity with the participants.

A wide variety of observer rating paradigms have been employed in social psychology. In their review of 263 independent samples, Connelly and Ones[23] demonstrated that observer ratings show predictive validity (e.g., of academic and job performance), incremental to and at times better than self-ratings. Although, the accuracy of observer ratings (i.e., inter-rater agreement, self-other agreement) varied considerably depending on a variety of factors including who was providing the rating (e.g., family member vs. stranger) and the trait being rated (e.g., extraversion vs. neuroticism).

One widely used observer rating paradigm involves ratings made by strangers (i.e., at zero acquaintance), often through brief excerpts or "thin slices" of expressive behavior.[24] Ratings made by individuals who have more information about targets (e.g., family members, coworkers) tend to outperform ratings at zero acquaintance (e.g., when predicting self-ratings of personality[23]). However, zero-acquaintance ratings have high ecological validity: in theory, they

reflect how an individual is perceived by strangers, a key aspect of interpersonal perception in daily life.[25,26] Further, from the perspective of evolutionary psychology, the ability to detect personality traits at zero acquaintance has clear adaptive utility (e.g., detection of emotional instability).[27] Within zero-acquaintance ratings, a variety of factors appear to influence the accuracy of ratings, including the type of information available (e.g., audio plus visual vs. text/electronic communication) and the trait being rated (e.g., extraversion vs. neuroticism).[23] The Realistic Accuracy Model (RAM) provides a framework for understanding this variability, noting that accurate judgments of personality require relevant information to be available and both detected and utilized by raters in making judgments.[26,28] This model helps understand why personality traits with higher visibility (e.g., extraversion) are more accurately rated at zero acquaintance than less visible traits (e.g., neuroticism), as well as why accuracy generally increases with the quantity and quality of available information (e.g., audio plus visual information yields more reliable ratings than text/electronic communication).[23] Similarly, the Self-Other Knowledge Asymmetry (SOKA) model highlights how the accuracy of ratings varies across domains, with self-ratings showing higher accuracy than observer ratings for traits low in visibility (e.g., neuroticism).[29]

It is important to note that while observer rating paradigms involving thin slices of behavior at zero acquaintance are intriguing and potentially informative, they should not be interpreted as necessarily measuring reality. Such ratings show only modest correlations with self-ratings (especially for less observable traits like neuroticism, $r = .08$)[23] and can be heavily influenced by stereotypes and other perceiver effects (i.e., rating biases within the individual making the rating).[30] Further, these judgments are not necessarily stable over time, tending to become more accurate as acquaintanceship increases.[31] Nonetheless, ratings at zero

acquaintance do provide information about how an individual is perceived by strangers. The incremental validity of such ratings over and above more conventional self-report measures is an issue that still warrants further study.

To our knowledge, only one prior study examined the effect of meditation training on thin slices of behavior using ratings at zero acquaintance. In two small independent samples (*n*s=26 and 20), Choi, Karremans, and Barendregt[32] found that novice meditators were rated as looking happier after a meditation retreat and experienced meditators were rated as looking happier relative to control participants based on brief video snippets. While intriguing, this work included only one rating dimension, a modest sample size of targets (i.e., those being rated), and a lack of random assignment, limiting conclusions that can be drawn about the experimental effects of meditation training. Further, it is unclear whether an even thinner slice of behavior (e.g., a still photograph, which has been used in other thin slices research),[22] may allow detection of training-related correlates.

The current study sought to address these limitations and explore the boundaries of information required to detect meditation-related differences through second-person observation. We employed one of the thinnest slices of expressive behavior – a still photograph – which provided a more stringent test of the lower limits of information for detecting correlates of meditation training. Relative to videos, photographs are less unlikely to provide potentially confounding contextual cues as to whether an individual engages in meditation practice (e.g., based on the language they use or the topics they discuss). Photographs were obtained in the context of a larger study examining differences between LTMs and MNPs as well as the experimental effects of meditation training through an RCT. Observers included undergraduate raters as well as meditation teachers to represent a range of perspectives from naïve to expert.

Assessments were made across several dimensions theoretically linked to meditation practice and embodiment,[33] including three of the Big Five personality traits.[14] Of note, rating dimensions were selected based on their theoretical relevance to meditation training and included several constructs reflecting internal states despite their low visibility (e.g., neuroticism).

Materials and Method

Participants

This study was approved by the University of Wisconsin-Madison Institutional Review Board. Written consent was obtained from participants.

Targets. Targets were recruited as part of a larger study investigating the effects of long- and short-term meditation training.[9,34] Participants included as targets in the current study consented to have their photographs used in future research. A sample of 16 long-term meditators (LTM; age=50.62, *SD*=9.56 years, 8 female, 14 non-Hispanic white, 10 attended graduate school, see Table 1 for full demographics) were recruited at meditation centers and through mailing lists. To be included, LTMs had to have practiced Vipassana and compassion/loving-kindness meditation for at least three years, have a daily practice of 30 minutes or more, and have attended at least three residential meditation retreats lasting five days or more (see Rosenkranz et al., 2016). LTMs had an average of 8,774 lifetime hours of meditation practice (range=1,439 to 32,612; *SD*=7,041). A sample of 83 meditation naïve participants (MNP; age=48.79, *SD*=11.13 years, 53 female, 76 non-Hispanic white, 42 attended graduate school) matched on age and gender were recruited in the Madison, WI area using online and print media. Demographic matching within the larger trial was focused on age and gender due to the association between these variables with a variety of neurobiological and psychological variables. Only MNPs who provided images at both pre- and post-test were rated.

The LTM and MNP groups did not differ by age, gender, race/ethnicity, or education (*p*s>.050). Participants in both groups were excluded if they had used psychotropic medications, had a psychiatric diagnosis in the past year, or had a history of bipolar or schizophrenic disorders, brain damage, or seizures.

Raters. Two samples of raters were recruited. A sample of 25 Buddhist meditation teachers (age=52.80, *SD*=12.25 years, 9 female, 20 non-Hispanic white) were recruited through meditation centers. Inclusion criteria were self-identification as a meditation teacher and experience leading residential meditation retreats. A sample of 86 undergraduate raters (age=19.07, *SD*=3.37 years, 54 female, 55 non-Hispanic white) were recruited through psychology courses at the University of Wisconsin–Madison.

Procedure

Intervention. Following baseline assessment, which included collection of photographs of spontaneous emotion (see below), MNPs were randomly assigned to one of three conditions: mindfulness-based stress reduction (MBSR),[15] health enhancement program (HEP),[34] or a waitlist (WL) control condition. MBSR is a standardized, eight-week mindfulness intervention involving instruction in formal (e.g., sitting meditation) and informal (e.g., attentiveness during daily life) mindfulness practice. MBSR was delivered in the typical group format by experienced MBSR instructors. HEP is an active control condition designed specifically to match MBSR as closely as possible, while not including mindfulness content. HEP includes mild physical activity, functional movement, nutrition education, and music and imagery designed to enhance psychological health.[34] HEP was delivered in an eight-week, group format by instructors with expertise in HEP content but no background in mindfulness. Participants in the WL condition received no intervention. The current MNPs were a subset of a sample of 130

participants enrolled in the larger trial.[34] Among the 86 MNP who provided consent to have their photographs used in future research, 27 were assigned to MBSR, 29 to HEP, and 27 to WL.

Photographs. Prior to randomization, LTMs and MNPs were photographed by a research assistant blind to study condition (i.e., LTM vs. MNP). MNPs were photographed again post-intervention. Photographs were taken as part of study visits that were equivalent for LTMs and MNPs. Participants were told "we're just going to take your photo," allowing a spontaneous facial expression (i.e., not restricted to a neutral expression). Spontaneous facial expressions have been shown to yield more accurate observer ratings of personality.[35] Color photographs were subsequently cropped at the neck and resized so that participants' heads were approximately the same size.

Rating paradigm. Raters, blind to study condition, provided ratings of still photographs. Both undergraduate and meditation teacher raters completed ratings through online surveys. Undergraduate raters completed surveys in the laboratory and meditation teachers completed surveys remotely. Each item was phrased as follows, with the respective trait varied across items: "How [trait] is this person? Please respond from 1 (not at all) to 7 (very)." We chose traits to be rated due to their relationship to well-being and embodiment, both of which are theoretically cultivated through meditation practice.[4,14,33] Of note, these dimensions were selected based on their potential association with meditation training, rather than their visibility (e.g., extraversion was not rated). Six items were drawn from the Ten Item Personality Inventory[36] assessing the Big-Five personality dimensions of conscientiousness, agreeableness, and neuroticism (two items for each). Two novel items assessed mindfulness ("mindful") and embodiment ("comfortable in their own skin"). One item assessed attractiveness as a potential confound.[22]

Due to the large number of photographs and dimensions being assessed, six separate surveys were created (see Table 2). Across four separate samples, undergraduate raters rated two items for each of the three Big Five personality traits (conscientiousness, agreeableness, neuroticism) as well as attractiveness and "comfortable in their own skin." We anticipated greater difficulty recruiting meditation teachers, so planned to have them rate a subset of items in order to increase reliability of available ratings. Across two separate samples, meditation teacher raters rated a single item for neuroticism and agreeableness, along with "comfortable in their own skin" and "mindful."

Data Analysis

Inter-rater reliability of observer ratings was determined using Shrout and Fleiss's[37] intra-class correlation coefficient (ICC) for fixed judges (i.e., ICC3, in which each target is rated by all judges; see Supplemental Materials Table 1 for sample R code). Observer ratings were then aggregated across raters and within target, yielding a single rating per target for each item. For items assessing personality dimensions, a composite score was computed by averaging across the two items (reverse scored as appropriate). To compare LTMs and MNPs at baseline, regression models were constructed predicting observer ratings from LTM status. Subsequent models controlled for observable confounds (age, gender, race/ethnicity, observer-rated attractiveness, body mass index). Analysis of variance (ANOVA) models examined intervention-related changes in observer ratings across the three randomized conditions (MBSR, HEP, WL). Our study was powered to detect large to moderate differences between LTM and MNP groups and large differences between the three randomized conditions.[38] However, *a priori* power calculations were for the larger trial from which these data are drawn. Specifically, the original study was powered to detect large to moderate ($d = 0.74$) between group effects based on prior

fMRI studies. The larger trial proposed samples of 36 participants per group (total $n = 144$ across four groups) to allow for attrition. To control for potential Type I error due to conducting tests across six target dimensions in our primary analyses (i.e., LTM vs. MNP at baseline, intervention-related time X group effect for MNPs), we controlled for false discovery rate using Benjamini and Hochberg's[39] method implemented using the 'p.adjust' function in R.[40] This method provides a more powerful alternative to Bonferroni-type adjustments and has been recommended for theory-driven contexts specifically.[41]

The larger trial from which these data were drawn was not pre-registered. At the time the larger trial was funded (2008), pre-registration was not widespread. A variety of self-report, behavioral, and neuroimaging data were collected as part of the larger trial and are not reported here. Similarly, the data collected via observer ratings reported here were not pre-registered. However, between-group comparisons (i.e., LTMs vs. MNPs, MBSR vs. HEP vs. waitlist) on personality dimensions and embodiment items were planned *a priori*. Along with rating attractiveness, raters also assessed "old" and "healthy" as two additional potential confounds, but those data are not reported here. An additional rating task was completed by a subsample of raters in which observers were shown a pre- and post-test photograph and were asked to choose which occurred following a well-being intervention. There was no indication that observers were able to predict post-intervention photographs above chance (overall, or for the active groups [MBSR, HEP] specifically). These data are not reported here.

Results

All items showed adequate inter-rater reliability (ICC$\geq$.75).[42] The two item personality scales showed adequate internal consistency reliability ($\alpha \geq .80$). Ratings from meditation teachers and undergraduates were highly correlated ($rs \geq .66$, $ps < .001$) and rater type (i.e.,

meditation teacher vs. undergraduate) did not moderate the association between group (LTM vs. MNP at baseline, MBSR vs. HEP vs. WL in ANOVAs) and ratings (*p*s>.050). Ratings were therefore combined to increase reliability.[37]

Means and standard deviations for the six constructs assessed are included in Tables 3 and 4. At baseline, LTMs were rated more highly than MNPs on conscientiousness, comfortable in their own skin, and mindful and were rated lower than MNPs on neuroticism (absolute value of *d*s=0.61 to 0.70, *p*s<.050; Figure 1, Table 3). No differences were observed for ratings of agreeableness or attractiveness (*p*s>.050). Results were unchanged controlling for observable confounds, with the exception of two models (LTM status predicting comfortable and agreeableness) in which LTM status became a marginally significant predictor (*p*s=.085, .089, respectively) when confounds were statistically controlled. A sensitivity analysis separated ratings from undergraduates and meditation teachers. Using undergraduate ratings only, LTMs were rated as more conscientious and comfortable than MNPs, but no longer less neurotic (*p*=.052). Using meditation teacher ratings only, LTMs were rated as less anxious and more mindful, but no longer more comfortable (*p*=.161).

Time by group interaction terms were used to model changes in observer ratings for MNPs over the course of intervention (MBSR, HEP, WL; Table 4). No time by group effects were observed on any of the six dimensions assessed (*p*s>.050). Results remained unchanged controlling for observable confounds. Exploratory *post hoc* analyses examined time by group interactions for each pairing (i.e., MBSR vs. HEP, MBSR vs. WL, etc.) and pre-post within-group changes. No time by group effects or pre-post within-group changes were observed (*p*s>.050).

Discussion

The current study examined the association of long- and short-term meditation training with observer ratings of personality and related dimensions. Photographs were taken of LTMs and MNPs at baseline, and again following MNPs' completion of training based on randomization to a mindfulness meditation training (MBSR), an active control condition (HEP), or a waitlist. High inter-rater reliability was obtained across all rating dimensions. Results supported the possibility that long-term meditation training may be linked with perceived differences in neuroticism and conscientiousness by strangers despite minimal information. Significant differences in the moderate range were observed on ratings of these personality traits, as well as on novel items intended to assess constructs specific to meditation (comfortable in their own skin, mindful). While results do not imply these reflect actual differences between LTMs and MNPs on these traits (and the broader interpersonal perception literature suggests ratings of internal states by strangers may not reflect actual differences),[23,27] they also do not appear to reflect merely globally positive ratings of LTMs. In particular, LTMs and MNPs did not differ on ratings of agreeableness and attractiveness. Further, results did not appear to be primarily driven by differences in observable confounds (e.g., age, attractiveness). Results were most robust when ratings from undergraduates and meditation teachers were combined.

In contrast, no evidence was found suggesting effects of short-term meditation training on observer ratings. In fact, neither MBSR nor the active control condition (HEP) differed from the waitlist group in observer-rated changes over the course of the eight-week interventions. Thus, if meditation is associated with changes in observers' perception of personality (which our design cannot demonstrate for LTMs due to non-random assignment), these effects may be restricted to long- rather than short-term training.

The possibility that long-term meditation practice may be associated with observer's perception of neuroticism and conscientiousness is intriguing, albeit qualified by a lack of randomization of participants to the LTM and MNP conditions. Interpersonal perceptions matter; they show incremental validity beyond self-report for predicting a variety of outcomes (e.g., teacher effectiveness, academic performance)[23,24] and are not confounded with known biases in self-report (e.g., social desirability) that may be particularly pernicious when assessing internal processes impacted by meditation training (e.g., mindfulness).[13,43] Our results support the notion that long-term meditation practice is associated with more favorable perceptions by others in still photographs, on dimensions with potentially important intra- and interpersonal consequences.[44] The ways in which these perceptions made from still photographs may relate to perception in daily life is not clear from the current study, of course. Nonetheless, it is possible that LTMs may experience a more welcoming interpersonal environment, vis-à-vis others' initial impressions.[22] Should that be the case, improved interpersonal interactions could be both an outcome of long-term practice as well as a mechanism through which long-term practice yields benefits in other domains (e.g., well-being, quality of life).

It is important to highlight that perceived differences between LTMs and MNPs were evident using a very small sample of expressive behavior (i.e., still photographs). While this may reflect the potency of potential differences (i.e., they could be detected from minimal information), the observed effect should be contextualized within the broader thin slices and zero acquaintance ratings literature. Meta-analytic evidence has highlighted the modest reliability for ratings by strangers of internal states based on still visual cues (e.g., r_{rr} = .25 for emotional stability) as well as the low self-other agreement for ratings by strangers of internal states (e.g., r = .08 for emotional stability, r = .12 for openness to experience).[23] Therefore, perceived

differences must be interpreted as simply that – differences in perception – and may or may not reflect differences in LTMs' and MNPs' actual internal states. This is particularly so given the small amount of information available to observers (e.g., still photographs). Again, still photographs are clearly not analogous to how interpersonal perception typically occurs within daily life – we almost always have access to considerably more information when interacting with others. However, our results nonetheless suggest that observers somehow perceived LTMs more favorably across several domains highly relevant to meditation training, despite the limited information provided.

It would be worthwhile continuing to unpack these initial findings in future studies. One potentially fruitful future direction implied by the Realistic Accuracy Model would be obtaining samples of behavior in situations with higher relevance to the traits being assessed which may be more likely to provide available relevant information that could be detected and utilized by raters to make valid judgments.[26] Recent work has shown that ratings of neuroticism are more accurate in trait-relevant situations (e.g., socially stressful situations).[27] A potential future direction could be obtaining samples of behavior in such situations (e.g., Trier Social Stress Task),[9,45] ideally within the context of random assignment to short-term meditation training. Perhaps tasks could be used and/or developed that provide information about other aspects impacted by meditation training assessed here (e.g., mindfulness, embodiment, conscientiousness). It could also be illuminating to determine which aspects of behavior raters are using as cues to differentiate LTMs and MNPs or the short-term effects of training. This could be done using machine learning and social sensing technologies.[46] Future studies could also continue to explore potential signals detectable through still photographs, perhaps opening the door to measurement

strategies that have not been widely implemented but could hold promise for the detection of emotional signals in daily life (e.g., examining facial expressions obtained through smartphones). A future study with a larger sample of LTMs could also explore potential dose effects of meditation on observer ratings of personality across LTMs (i.e., is more training associated with more positive perceptions). The small sample of LTMs in the current study (n = 16) prohibited a proper assessment of this possibility.

The lack of short-term training effects is harder to interpret and could be due to a genuine lack of impact or an insensitivity of our observer-rated measures to these potentially more subtle changes. Pre-post analyses also relied on a subset of participants, which reduced our statistical power to detect effects (although very small pre-post effects suggest power is not the sole issue). It seems the most prudent conclusion is that short-term meditation training does not impact interpersonal perceptions of personality observable through photographs. As noted above, it could be valuable to assess the impact of short-term training in future studies through samples of behavior drawn from contexts more likely to provide cues reflecting target constructs of interest (e.g., socially stressful situations, tasks requiring conscientiousness).

Key limitations of the current study include a lack of random assignment to LTM and MNP conditions (introducing a risk of selection bias) and a relatively modest number of LTMs. The available sample was below that for which the power analysis was conducted, which had assumed a moderate-to-large effect size (d = 0.74). Thus, it is very likely that the current study was underpowered to detect more modest effects. Our use of photographs may have obscured changes detectable through a thicker slice of behavior (e.g., video recordings) or behaviors with higher relevance to the constructs of interest. We also relied primarily on personality dimensions which may be less sensitive to short-term training, given their trait-like nature. While we

focused on dimensions theoretically linked to meditation training and embodiment, we did not assess two of the five Big Five dimensions (i.e., extraversion, openness). Neither group of raters were experts in personality, so our results merely indicate a lay interpretation of the items that were rated, rather than a scientifically-based understanding of the specific personality dimensions assessed (although the Big Five items were drawn from self-report measures used to assess personality in the general population).[36] Further, the majority of our sample of raters were young, undergraduate students (mean age = 19.07). This group may be prone to a variety of perceiver effects that introduced bias into the ratings (e.g., associated with rating targets who are on average 30 years older), raising questions about the degree to which the ratings of photographs in the current study may generalize to interpersonal perceptions made by individuals of a wider variety of ages in the context of daily life.

These limitations notwithstanding, we believe this is the first study to demonstrate that naïve observers perceive LTMs differently than MNPs on several domains theoretically linked with mindfulness training, even through very minimal samples of behavior (i.e., still photographs). This work highlights the possibility that LTMs are perceived differently in daily life, which in theory could both represent an outcome of long-term training as well as a potential mechanism through which long-term training confers psychological and interpersonal benefits. Future work investigating the interpersonal effects of mindfulness training is warranted.

Acknowledgments

The authors would like to thank William Hoyt for his valuable input on design of the observer rating paradigm and comments on a previous draft of this manuscript and Donal MacCoon, David Bachhuber, Tammi Kral, Amy Cayo, and Melissa Rosenkranz who contributed to planning and execution of the larger study from which targets were drawn. This work was supported by a grant from the National Center for Complementary and Integrative Health (NCCIH) P01AT004952 to Richard J. Davidson and a grant from the Mind & Life Institute (Varela Award) to Simon B. Goldberg. Richard J. Davidson is the founder, president, and serves on the board of directors for the non-profit organization, Healthy Minds Innovations, Inc.

References

1. Clarke TC, Barnes PM, Black LI, Stussman BJ, Nahin RL. Use of yoga, meditation, and chiropractors among U.S. adults aged 18 and over. Hyattsville, MD: National Center for Health Statistics; 2018.

2. Goldberg SB, Tucker RP, Greene PA, Davidson RJ, Wampold BE, Kearney DJ, et al. Mindfulness-based interventions for psychiatric disorders: A systematic review and meta-analysis. Clin Psychol Rev. 2018 Feb;59:52–60.

3. Goyal M, Singh S, Sibinga EMS, Gould NF, Rowland-Seymour A, Sharma R, et al. Meditation Programs for Psychological Stress and Well-being: A Systematic Review and Meta-analysis. JAMA Intern Med. 2014 Mar 1;174(3):357–68.

4. Khoury B, Sharma M, Rush SE, Fournier C. Mindfulness-based stress reduction for healthy individuals: A meta-analysis. J Psychosom Res. 2015 Jun 1;78(6):519–28.

5. Lykins E, Baer RA. Psychological functioning in a sample of long-term practitioners of mindfulness meditation. J Cogn Psychother. 2009;23(3):226–41.

6. Fox KCR, Dixon ML, Nijeboer S, Girn M, Floman JL, Lifshitz M, et al. Functional neuroanatomy of meditation: A review and meta-analysis of 78 functional neuroimaging investigations. Neurosci Biobehav Rev. 2016 Jun 1;65:208–28.

7. Kaliman P, Álvarez-López MJ, Cosín-Tomás M, Rosenkranz MA, Lutz A, Davidson RJ. Rapid changes in histone deacetylases and inflammatory gene expression in expert meditators. Psychoneuroendocrinology. 2014 Feb 1;40:96–107.

8. Lutz A, Greischar LL, Rawlings NB, Ricard M, Davidson RJ. Long-term meditators self-induce high-amplitude gamma synchrony during mental practice. Proc Natl Acad Sci. 2004 Nov 16;101(46):16369–73.

9. Rosenkranz MA, Lutz A, Perlman DM, Bachhuber DRW, Schuyler BS, MacCoon DG, et al. Reduced stress and inflammatory responsiveness in experienced meditators compared to a matched healthy control group. Psychoneuroendocrinology. 2016 Jun 1;68:117–25.

10. Lykins ELB, Baer RA, Gottlob LR. Performance-Based Tests of Attention and Memory in Long-Term Mindfulness Meditators and Demographically Matched Nonmeditators. Cogn Ther Res. 2012 Feb 1;36(1):103–14.

11. Sedlmeier P, Eberth J, Schwarz M, Zimmermann D, Haarig F, Jaeger S, et al. The psychological effects of meditation: A meta-analysis. Psychol Bull. 2012;138(6):1139–71.

12. Brown KW, Ryan RM. The benefits of being present: Mindfulness and its role in psychological well-being. J Pers Soc Psychol. 2003 Apr;84(4):822–48.

13. Goldberg SB, Tucker RP, Greene PA, Simpson TL, Hoyt WT, Kearney DJ, et al. What can we learn from randomized clinical trials about the construct validity of self- report measures of mindfulness? A meta-analysis. Mindfulness. 2019;

14. Giluk TL. Mindfulness, Big Five personality, and affect: A meta-analysis. Personal Individ Differ. 2009 Dec 1;47(8):805–11.

15. Kabat-Zinn J. Full catastrophe living: Using the wisdom of your body and mind to face stress, pain, and illness (Revised ed.). New York: Random House; 2013.

16. Leiberg S, Klimecki O, Singer T. Short-Term Compassion Training Increases Prosocial Behavior in a Newly Developed Prosocial Game. PLOS ONE. 2011 Mar 9;6(3):e17798.

17. Weng HY, Fox AS, Shackman AJ, Stodola DE, Caldwell JZK, Olson MC, et al. Compassion Training Alters Altruism and Neural Responses to Suffering. Psychol Sci. 2013 Jul 1;24(7):1171–80.

18. Fredrickson BL, Cohn MA, Coffey KA, Pek J, Finkel SM. Open hearts build lives: Positive emotions, induced through loving-kindness meditation, build consequential personal resources. J Pers Soc Psychol. 2008 Nov;95(5):1045–62.

19. Hutcherson CA, Seppala EM, Gross JJ. Loving-kindness meditation increases social connectedness. Emotion. 2008;8(5):720–4.

20. Lindsay EK, Young S, Brown KW, Smyth JM, Creswell JD. Mindfulness training reduces loneliness and increases social contact in a randomized controlled trial. Proc Natl Acad Sci. 2019 Feb 26;116(9):3488–93.

21. Davidson RJ, Kaszniak AW. Conceptual and Methodological Issues in Research on Mindfulness and Meditation. Am Psychol. 2015 Oct;70(7):581–92.

22. Harker L, Keltner D. Expressions of positive emotion in women's college yearbook pictures and their relationship to personality and life outcomes across adulthood. J Pers Soc Psychol. 2001 Jan;80(1):112–24.

23. Connelly BS, Ones DS. An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. Psychol Bull. 2010;136(6):1092–122.

24. Ambady N, Bernieri FJ, Richeson JA. Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. In: Advances in Experimental Social Psychology [Internet]. Academic Press; 2000 [cited 2019 Mar 28]. p. 201–71. Available from: http://www.sciencedirect.com/science/article/pii/S0065260100800064

25. Hirschmüller S, Link to external site this link will open in a new window, Egloff B, Link to external site this link will open in a new window, Nestler S, Back MD. The dual lens model: A comprehensive framework for understanding self–other agreement of personality judgments at zero acquaintance. J Pers Soc Psychol. 2013 Feb;104(2):335–53.

26. Funder DC. Accurate Personality Judgment. Curr Dir Psychol Sci. 2012 Jun 1;21(3):177–82.

27. Hirschmüller S, Egloff B, Schmukle SC, Nestler S, Back MD. Accurate Judgments of Neuroticism at Zero Acquaintance: A Question of Relevance. J Pers. 2015;83(2):221–8.

28. Funder DC. On the accuracy of personality judgment: A realistic approach. Psychol Rev. 1995 Oct;102(4):652–70.

29. Vazire S. Who knows what about a person? The self–other knowledge asymmetry (SOKA) model. J Pers Soc Psychol. 2010 Feb;98(2):281–300.

30. Kenny DA. Interpersonal perception: A social relations analysis. New York: Guilford Press; 1994.

31. Lee K, Ashton MC. Acquaintanceship and self/observer agreement in personality judgment. J Res Personal. 2017 Oct 1;70:1–5.

32. Choi Y, Karremans JC, Barendregt H. The happy face of mindfulness: Mindfulness meditation is associated with perceptions of happiness as rated by outside observers. J Posit Psychol. 2012 Jan 1;7(1):30–5.

33. Farb N, Daubenmier JJ, Price CJ, Gard T, Kerr C, Dunn B, et al. Interoception, Contemplative Practice, and Health. Front Psychol [Internet]. 2015 Jun 1 [cited 2019 Mar 28];6. Available from: https://doaj.org

34. MacCoon DG, Imel ZE, Rosenkranz MA, Sheftel JG, Weng HY, Sullivan JC, et al. The validation of an active control intervention for Mindfulness Based Stress Reduction (MBSR). Behav Res Ther. 2012 Jan 1;50(1):3–12.

35. Naumann LP, Vazire S, Rentfrow PJ, Gosling SD. Personality Judgments Based on Physical Appearance. Pers Soc Psychol Bull. 2009 Dec 1;35(12):1661–71.

36. Gosling SD, Rentfrow PJ, Swann WB. A very brief measure of the Big-Five personality domains. J Res Personal. 2003 Dec 1;37(6):504–28.

37. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. Psychol Bull. 1979;86(2):420–8.

38. Cohen J. A power primer. Psychol Bull. 1992;112(1):155.

39. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B Methodol. 1995;57(1):289–300.

40. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2018.

41. Glickman ME, Rao SR, Schultz MR. False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies. J Clin Epidemiol. 2014 Aug 1;67(8):850–7.

42. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016 Jun 1;15(2):155–63.

43. Grossman P. Defining mindfulness by how poorly I think I pay attention during everyday awareness and other intractable problems for psychology's (re)invention of mindfulness: Comment on Brown et al. (2011). Psychol Assess. 2011 Dec;23(4):1034–40.

44. Friedman HS, Kern ML. Personality, Well-Being, and Health. Annu Rev Psychol. 2014 Jan 3;65(1):719–42.

45. Kirschbaum C, Pirke K-M, Hellhammer DH. The 'Trier Social Stress Test' – A Tool for Investigating Psychobiological Stress Responses in a Laboratory Setting. Neuropsychobiology. 1993;28(1–2):76–81.

46. Schmid Mast M, Gatica-Perez D, Frauendorfer D, Nguyen L, Choudhury T. Social Sensing for Psychology: Automated Interpersonal Behavior Assessment. Curr Dir Psychol Sci. 2015 Apr 1;24(2):154–60.

Table 1. Sample demographics

	Targets		Raters	
Demographic Variable	LTM (n = 16)	MNP (n = 83)	Undergraduates (n = 86)	Meditation Teachers (n = 25)
Age, mean(SD)	50.62 (9.56)	48.79 (11.13)	19.07 (3.37)	52.80 (12.25)
Female, n(%)	8 (50)	53 (63.9)	54 (62.8)	9 (36.0)
Race/ethnicity				
non-Hispanic white, n(%)	14 (87.5)	76 (91.6)	55 (64.0)	20 (80.0)
non-Hispanic black, n(%)	0 (0.0)	0 (0.0)	5 (5.8)	1 (4.0)
Asian, n(%)	2 (12.5)	2 (2.4)	15 (17.4)	0 (0.0)
Native American, n(%)	0 (0.0)	0 (0.0)	1 (1.2)	0 (0.0)
Hispanic, any race, n(%)	0 (0.0)	4 (4.8)	1 (1.2)	1 (4.0)
More than one race/ethnicity, n(%)	0 (0.0)	1 (1.2)	8 (9.3)	1 (4.0)
Did not want to respond, n(%)	0 (0.0)	0 (0.0)	1 (1.2)	2 (8.0)

Note: LTM = long-term meditator; MNP = meditation naïve participant.

Table 2. Descriptions of six surveys

Survey	Sample	Traits assessed
1	21 undergraduates	[a]Calm/emotionally stable
2	21 undergraduates	Attractive
3	22 undergraduates	[a]Anxious/easily upset, [a]Sympathetic/warm, [a]Dependable/self-disciplined
4	22 undergraduates	[a]Critical/quarrelsome, [a]Disorganized/careless, Comfortable in their own skin
5	14 meditation teachers	[a]Anxious/easily upset
6	11 meditation teachers	[a]Sympathetic/warm, Comfortable in their own skin, Mindful

Note: [a]Six items drawn from Ten Item Personality Inventory (Gosling et al., 2003) assessing conscientiousness, agreeableness, and neuroticism.

Table 3. LTMs versus MNPs at baseline

Rating Domain	LTM	MNP	*d*	*p*
Attractive	2.81 (0.72)	2.67 (0.58)	0.23	0.406
Agreeable	4.26 (0.85)	3.93 (0.78)	0.42	0.162
Conscientious	4.54 (0.64)	4.18 (0.49)	0.70*	0.039
Neurotic	3.24 (0.65)	3.64 (0.66)	-0.61*	0.045
Comfortable	4.62 (0.74)	4.18 (0.70)	0.62*	0.045
Mindful	4.75 (0.79)	4.31 (0.60)	0.69*	0.039

Note: Sample sizes: LTM = 16, MNP = 83. LTM = Long-term meditation practitioners; MNP = meditation naïve participants at pre-test; *d* = Cohen's *d*; Comfortable = "comfortable in their own skin"; Big-Five personality dimensions assessed using two-item scales drawn from the Ten Item Personality Inventory;[36] *p* = *p*-values from independent samples t-test, adjusted for false discovery rate;[39] **p* < .050.

Table 4. Pre-post comparison for MNPs randomized to three conditions

	MBSR			HEP			Waitlist			
Rating Domain	Pre	Post	*d*	Pre	Post	*d*	Pre	Post	*d*	*p*
Attractive	2.66 (0.64)	2.62 (0.65)	-0.06	2.56 (0.52)	2.55 (0.54)	-0.02	2.81 (0.58)	2.77 (0.57)	-0.07	0.991
Agreeable	3.66 (0.72)	3.69 (0.74)	0.04	3.93 (0.81)	4.17 (0.79)	0.3	4.20 (0.74)	4.25 (0.82)	0.06	0.978
Conscientious	4.07 (0.40)	4.08 (0.50)	0.02	4.08 (0.58)	4.18 (0.57)	0.17	4.39 (0.42)	4.36 (0.54)	-0.06	0.978
Neurotic	3.87 (0.64)	3.85 (0.61)	-0.03	3.65 (0.69)	3.39 (0.62)	-0.4	3.40 (0.58)	3.36 (0.62)	-0.07	0.978
Comfortable	3.94 (0.71)	4.01 (0.71)	0.1	4.18 (0.71)	4.41 (0.70)	0.33	4.41 (0.64)	4.51 (0.72)	0.15	0.978
Mindful	4.33 (0.69)	4.28 (0.57)	-0.08	4.30 (0.56)	4.40 (0.61)	0.17	4.31 (0.55)	4.30 (0.60)	-0.02	0.978

Note: Sample sizes: MBSR = 27, HEP = 29, waitlist = 27. MNP = meditation naïve participants; MBSR = Mindfulness-Based Stress Reduction; HEP = Health Enhancement Program; Pre = pre-test; Post = post-test; *d* = within-group (i.e., pre-post) Cohen's *d*; Comfortable = "comfortable in their own skin"; Big-Five personality dimensions assessed using two-item scales drawn from the Ten Item Personality Inventory;[36] *p* = *p*-values from time X group ANOVA models, adjusted for false discovery rate;[39] **p* < .050.

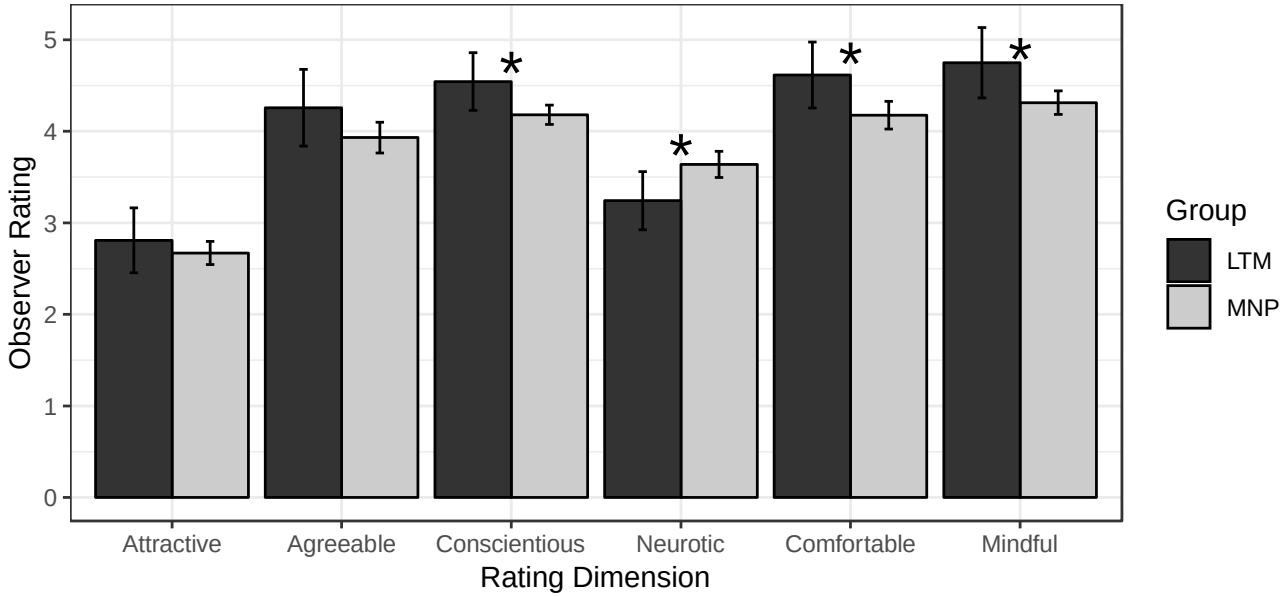

Figure 1. Long-term meditation (LTM) practitioners are rated as less neurotic and more conscientious, "comfortable in their own skin," and mindful compared to meditation naïve participants (MNP). Target sample includes 16 LTMs and 83 MNPs. Rater sample includes 25 meditation teachers and 86 undergraduates.

Supplemental Materials Table 1. Sample R code for computing intraclass correlation coefficient (ICC3)

```
vp <- function(data.g, formula.g) {   # variance partitioning
  lmer.out <- lmer(data = data.g, formula = formula.g)
  var.residual<-attr(VarCorr(lmer.out), "sc")^2
  var.comp <- ldply(VarCorr(lmer.out))
  names(var.comp) <- c("Factor", "Variance")
  var.comp <- rbind(var.comp, data.frame("Factor" = "Residual", "Variance" =
var.residual))
  var.comp$Percent <- round(var.comp$Variance / sum(var.comp$Variance) * 100, 1)
  attr(var.comp, "mer") <- lmer.out
  class(var.comp) <- c("data.frame", "G")
  var.comp
}

#T1_calm
oneFac <- reshape(data=df[,c("SID",paste("WB",sids,"_T1_calm", sep=""))],
                  varying=paste("WB",sids,"_T1_calm", sep=""),
v.names=c("rating"),timevar="Target",sep="_", direction="long")
oneFac$id <- NULL; oneFac$Rater <- factor(oneFac$SID); oneFac$Target <-
factor(oneFac$Target)
lme1 <- lmer(data=oneFac, formula = rating ~ 1 + (1|Rater) + (1|Target))
vp1 <- vp(data=oneFac, formula=rating ~ 1 + (1|Rater) + (1|Target)); varT <-
vp1$Variance[1]; varR <- vp1$Variance[2]; varRTe <- vp1$Variance[3]
varT/(varT + varRTe/table(table(df[!is.na(df$WB101_T1_calm),"SID"]))) #.84
```

Still facial photographs of long-term meditators are perceived by naïve observers as less neurotic, more conscientious and more mindful than non-meditating controls

Simon B. Goldberg[1,2]*, Matthew Hirshberg[2], Lawrence Y. Tello[2,3], Helen Y. Weng[4], Lisa Flook[2,5], & Richard J. Davidson[2,6,7]

[1]Department of Counseling Psychology, University of Wisconsin – Madison, Madison, WI, USA

[2]Center for Healthy Minds, University of Wisconsin – Madison, Madison, WI, USA

[3]National Eye Institute, National Institutes of Health, Bethesda, MD, USA

[4]Osher Center for Integrative Medicine, University of California, San Francisco, San Francisco, CA, USA

[5]Learning Policy Institute, Palo Alto, CA, USA

[6]Department of Psychology, University of Wisconsin – Madison, Madison, USA

[7]Department of Psychiatry, University of Wisconsin – Madison, Madison, USA

*Correspondence to: Simon B. Goldberg, Department of Counseling Psychology, University of Wisconsin – Madison, 335 Education Building, 1000 Bascom Mall, Madison, WI, 53706

sbgoldberg@wisc.edu

Abstract

The impact of meditation training on self-report psychological variables is well-established. Although meditation training is purported to have interpersonal impacts, whether naïve observers perceive differences associated with long- and short-term meditation training is largely unknown. The current study provided a stringent test of this possibility through observer ratings of a very thin slice of expressive behavior: still photographs. Photographs were drawn from a larger study investigating differences between long-term meditators (LTM) and meditation naïve participants (MNP) who were exposed to one of three experimental conditions. Photographs of ninety-nine targets (16 LTMs, 83 MNPs) were taken at baseline, prior to the randomization of MNPs to an eight-week mindfulness meditation course (mindfulness-based stress reduction; n=27), an active control comparison condition (health enhancement program; n=29), or a waitlist control group (n=27) and again after the training period. Pre- and post-intervention photographs were then rated by 25 meditation teachers and 86 undergraduate raters on five domains theoretically linked to meditation training. Results indicated that relative to MNPs, LTMs were rated as less neurotic and more conscientious, mindful, and "comfortable in their own skin" at baseline (ds=0.61 to 0.70, ps<.050), although not more agreeable or attractive. Results were largely unchanged when controlling for five observable confounds (age, gender, race/ethnicity, body mass index, attractiveness). No evidence was found supporting experimental effects of short-term meditation training on observer ratings. Thus, it seems that if meditation is associated with observable differences in facial behavior, effects may be limited to long-term training.

Keywords: mindfulness; meditation; observer ratings; personality; thin slices

There has been a dramatic increase in interest in meditation over the past several decades. Recent data from the National Health Interview Survey showed a three-fold increase in past year meditation use between 2012 and 2017 in the United States (4.1% to 14.3%).[1] Scientific interest in the topic has also increased exponentially. Experimental data from randomized controlled trials (RCTs) have shown reductions in psychiatric symptoms and improvements in well-being in the context of relatively brief (e.g., eight-week) meditation interventions, in both clinical[2,3] and non-clinical populations.[4] A related body of research has examined differences between individuals with extensive meditation experience (i.e., long-term meditators [LTM]) and meditation naïve participants (MNP). Relative to MNPs, LTMs have often shown lower psychiatric symptoms and higher well-being[5] along with a host of biological differences indicating less reactive stress physiology and improved emotion and attention regulation,[6–9] although not always.[10] Meditation training has been linked via meta-analysis to lower neuroticism and higher attention and empathy.[11] Further, dispositional mindfulness, a psychological construct purported to be cultivated in the context of various forms of meditation training (e.g., mindfulness meditation),[12,13] is correlated with several aspects of personality including neuroticism (r = -.45), conscientiousness (r = .32), and agreeableness (r = .22).[14]

Theories on the effects of meditation practice claim that they will be embodied and impact both psychological and physiological variables.[15] Although the effects of meditation practice impacting psychological variables as reported by the practitioner have been well-established,[2,5] little research has examined how meditation practice may impact interpersonal perceptions by others, including close relationships and strangers. Theories suggest that meditation practice enhances a sense of interconnectedness, and may also impact others through greater interpersonal connection and prosocial behavior.[16–20] Most research has been conducted

on first-person reports of effects, and rarely report on second-person observer effects, to examine whether the effects of meditation practice can be detectible by others.[21] However, interpersonal perception through second-person observers is important, both as a potential indicator of internal states (e.g., mood, well-being) as well as a proxy for how an individual may be perceived in daily life. Moreover, being perceived more favorably by others (e.g., as happier) may lead to more positive expectations of social interactions,[22] which could in turn produce beneficial inter- and intrapersonal effects. Thus, the domain of interpersonal perception may be one *in* which and *through* which beneficial effects of meditation practice appear. We tested the possibility that meditation training may be associated with differences in second-person observer effects with a stringent criterion, using a thin slice of behavior from still photographs and observers with no familiarity with the participants.

A wide variety of observer rating paradigms have been employed in social psychology. In their review of 263 independent samples, Connelly and Ones[23] demonstrated that observer ratings show predictive validity (e.g., of academic and job performance), incremental to and at times better than self-ratings. Although, the accuracy of observer ratings (i.e., inter-rater agreement, self-other agreement) varied considerably depending on a variety of factors including who was providing the rating (e.g., family member vs. stranger) and the trait being rated (e.g., extraversion vs. neuroticism).

One widely used observer rating paradigm involves ratings made by strangers (i.e., at zero acquaintance), often through brief excerpts or "thin slices" of expressive behavior.[24] Ratings made by individuals who have more information about targets (e.g., family members, coworkers) tend to outperform ratings at zero acquaintance (e.g., when predicting self-ratings of personality[23]). However, zero-acquaintance ratings have high ecological validity: in theory, they

reflect how an individual is perceived by strangers, a key aspect of interpersonal perception in daily life.[25,26] Further, from the perspective of evolutionary psychology, the ability to detect personality traits at zero acquaintance has clear adaptive utility (e.g., detection of emotional instability).[27] Within zero-acquaintance ratings, a variety of factors appear to influence the accuracy of ratings, including the type of information available (e.g., audio plus visual vs. text/electronic communication) and the trait being rated (e.g., extraversion vs. neuroticism).[23] The Realistic Accuracy Model (RAM) provides a framework for understanding this variability, noting that accurate judgments of personality require relevant information to be available and both detected and utilized by raters in making judgments.[26,28] This model helps understand why personality traits with higher visibility (e.g., extraversion) are more accurately rated at zero acquaintance than less visible traits (e.g., neuroticism), as well as why accuracy generally increases with the quantity and quality of available information (e.g., audio plus visual information yields more reliable ratings than text/electronic communication).[23] Similarly, the Self-Other Knowledge Asymmetry (SOKA) model highlights how the accuracy of ratings varies across domains, with self-ratings showing higher accuracy than observer ratings for traits low in visibility (e.g., neuroticism).[29]

It is important to note that while observer rating paradigms involving thin slices of behavior at zero acquaintance are intriguing and potentially informative, they should not be interpreted as necessarily measuring reality. Such ratings show only modest correlations with self-ratings (especially for less observable traits like neuroticism, $r = .08$)[23] and can be heavily influenced by stereotypes and other perceiver effects (i.e., rating biases within the individual making the rating).[30] Further, these judgments are not necessarily stable over time, tending to become more accurate as acquaintanceship increases.[31] Nonetheless, ratings at zero

acquaintance do provide information about how an individual is perceived by strangers. The incremental validity of such ratings over and above more conventional self-report measures is an issue that still warrants further study.

To our knowledge, only one prior study examined the effect of meditation training on thin slices of behavior using ratings at zero acquaintance. In two small independent samples (*n*s=26 and 20), Choi, Karremans, and Barendregt[32] found that novice meditators were rated as looking happier after a meditation retreat and experienced meditators were rated as looking happier relative to control participants based on brief video snippets. While intriguing, this work included only one rating dimension, a modest sample size of targets (i.e., those being rated), and a lack of random assignment, limiting conclusions that can be drawn about the experimental effects of meditation training. Further, it is unclear whether an even thinner slice of behavior (e.g., a still photograph, which has been used in other thin slices research),[22] may allow detection of training-related correlates.

The current study sought to address these limitations and explore the boundaries of information required to detect meditation-related differences through second-person observation. We employed one of the thinnest slices of expressive behavior – a still photograph – which provided a more stringent test of the lower limits of information for detecting correlates of meditation training. Relative to videos, photographs are less unlikely to provide potentially confounding contextual cues as to whether an individual engages in meditation practice (e.g., based on the language they use or the topics they discuss). Photographs were obtained in the context of a larger study examining differences between LTMs and MNPs as well as the experimental effects of meditation training through an RCT. Observers included undergraduate raters as well as meditation teachers to represent a range of perspectives from naïve to expert.

Assessments were made across several dimensions theoretically linked to meditation practice and embodiment,[33] including three of the Big Five personality traits.[14] Of note, rating dimensions were selected based on their theoretical relevance to meditation training and included several constructs reflecting internal states despite their low visibility (e.g., neuroticism).

Materials and Method

Participants

This study was approved by the University of Wisconsin-Madison Institutional Review Board. Written consent was obtained from participants.

Targets. Targets were recruited as part of a larger study investigating the effects of long- and short-term meditation training.[9,34] Participants included as targets in the current study consented to have their photographs used in future research. A sample of 16 long-term meditators (LTM; age=50.62, *SD*=9.56 years, 8 female, 14 non-Hispanic white, 10 attended graduate school, see Table 1 for full demographics) were recruited at meditation centers and through mailing lists. To be included, LTMs had to have practiced Vipassana and compassion/loving-kindness meditation for at least three years, have a daily practice of 30 minutes or more, and have attended at least three residential meditation retreats lasting five days or more (see Rosenkranz et al., 2016). LTMs had an average of 8,774 lifetime hours of meditation practice (range=1,439 to 32,612; *SD*=7,041). A sample of 83 meditation naïve participants (MNP; age=48.79, *SD*=11.13 years, 53 female, 76 non-Hispanic white, 42 attended graduate school) matched on age and gender were recruited in the Madison, WI area using online and print media. Demographic matching within the larger trial was focused on age and gender due to the association between these variables with a variety of neurobiological and psychological variables. Only MNPs who provided images at both pre- and post-test were rated.

The LTM and MNP groups did not differ by age, gender, race/ethnicity, or education (*p*s>.050). Participants in both groups were excluded if they had used psychotropic medications, had a psychiatric diagnosis in the past year, or had a history of bipolar or schizophrenic disorders, brain damage, or seizures.

Raters. Two samples of raters were recruited. A sample of 25 Buddhist meditation teachers (age=52.80, *SD*=12.25 years, 9 female, 20 non-Hispanic white) were recruited through meditation centers. Inclusion criteria were self-identification as a meditation teacher and experience leading residential meditation retreats. A sample of 86 undergraduate raters (age=19.07, *SD*=3.37 years, 54 female, 55 non-Hispanic white) were recruited through psychology courses at the University of Wisconsin–Madison.

Procedure

Intervention. Following baseline assessment, which included collection of photographs of spontaneous emotion (see below), MNPs were randomly assigned to one of three conditions: mindfulness-based stress reduction (MBSR),[15] health enhancement program (HEP),[34] or a waitlist (WL) control condition. MBSR is a standardized, eight-week mindfulness intervention involving instruction in formal (e.g., sitting meditation) and informal (e.g., attentiveness during daily life) mindfulness practice. MBSR was delivered in the typical group format by experienced MBSR instructors. HEP is an active control condition designed specifically to match MBSR as closely as possible, while not including mindfulness content. HEP includes mild physical activity, functional movement, nutrition education, and music and imagery designed to enhance psychological health.[34] HEP was delivered in an eight-week, group format by instructors with expertise in HEP content but no background in mindfulness. Participants in the WL condition received no intervention. The current MNPs were a subset of a sample of 130

participants enrolled in the larger trial.[34] Among the 86 MNP who provided consent to have their photographs used in future research, 27 were assigned to MBSR, 29 to HEP, and 27 to WL.

Photographs. Prior to randomization, LTMs and MNPs were photographed by a research assistant blind to study condition (i.e., LTM vs. MNP). MNPs were photographed again post-intervention. Photographs were taken as part of study visits that were equivalent for LTMs and MNPs. Participants were told "we're just going to take your photo," allowing a spontaneous facial expression (i.e., not restricted to a neutral expression). Spontaneous facial expressions have been shown to yield more accurate observer ratings of personality.[35] Color photographs were subsequently cropped at the neck and resized so that participants' heads were approximately the same size.

Rating paradigm. Raters, blind to study condition, provided ratings of still photographs. Both undergraduate and meditation teacher raters completed ratings through online surveys. Undergraduate raters completed surveys in the laboratory and meditation teachers completed surveys remotely. Each item was phrased as follows, with the respective trait varied across items: "How [trait] is this person? Please respond from 1 (not at all) to 7 (very)." We chose traits to be rated due to their relationship to well-being and embodiment, both of which are theoretically cultivated through meditation practice.[4,14,33] Of note, these dimensions were selected based on their potential association with meditation training, rather than their visibility (e.g., extraversion was not rated). Six items were drawn from the Ten Item Personality Inventory[36] assessing the Big-Five personality dimensions of conscientiousness, agreeableness, and neuroticism (two items for each). Two novel items assessed mindfulness ("mindful") and embodiment ("comfortable in their own skin"). One item assessed attractiveness as a potential confound.[22]

Due to the large number of photographs and dimensions being assessed, six separate surveys were created (see Table 2). Across four separate samples, undergraduate raters rated two items for each of the three Big Five personality traits (conscientiousness, agreeableness, neuroticism) as well as attractiveness and "comfortable in their own skin." We anticipated greater difficulty recruiting meditation teachers, so planned to have them rate a subset of items in order to increase reliability of available ratings. Across two separate samples, meditation teacher raters rated a single item for neuroticism and agreeableness, along with "comfortable in their own skin" and "mindful."

Data Analysis

Inter-rater reliability of observer ratings was determined using Shrout and Fleiss's[37] intra-class correlation coefficient (ICC) for fixed judges (i.e., ICC3, in which each target is rated by all judges; see Supplemental Materials Table 1 for sample R code). Observer ratings were then aggregated across raters and within target, yielding a single rating per target for each item. For items assessing personality dimensions, a composite score was computed by averaging across the two items (reverse scored as appropriate). To compare LTMs and MNPs at baseline, regression models were constructed predicting observer ratings from LTM status. Subsequent models controlled for observable confounds (age, gender, race/ethnicity, observer-rated attractiveness, body mass index). Analysis of variance (ANOVA) models examined intervention-related changes in observer ratings across the three randomized conditions (MBSR, HEP, WL). Our study was powered to detect large to moderate differences between LTM and MNP groups and large differences between the three randomized conditions.[38] However, *a priori* power calculations were for the larger trial from which these data are drawn. Specifically, the original study was powered to detect large to moderate (d = 0.74) between group effects based on prior

fMRI studies. The larger trial proposed samples of 36 participants per group (total n = 144 across four groups) to allow for attrition. To control for potential Type I error due to conducting tests across six target dimensions in our primary analyses (i.e., LTM vs. MNP at baseline, intervention-related time X group effect for MNPs), we controlled for false discovery rate using Benjamini and Hochberg's[39] method implemented using the 'p.adjust' function in R.[40] This method provides a more powerful alternative to Bonferroni-type adjustments and has been recommended for theory-driven contexts specifically.[41]

The larger trial from which these data were drawn was not pre-registered. At the time the larger trial was funded (2008), pre-registration was not widespread. A variety of self-report, behavioral, and neuroimaging data were collected as part of the larger trial and are not reported here. Similarly, the data collected via observer ratings reported here were not pre-registered. However, between-group comparisons (i.e., LTMs vs. MNPs, MBSR vs. HEP vs. waitlist) on personality dimensions and embodiment items were planned *a priori*. Along with rating attractiveness, raters also assessed "old" and "healthy" as two additional potential confounds, but those data are not reported here. An additional rating task was completed by a subsample of raters in which observers were shown a pre- and post-test photograph and were asked to choose which occurred following a well-being intervention. There was no indication that observers were able to predict post-intervention photographs above chance (overall, or for the active groups [MBSR, HEP] specifically). These data are not reported here.

Results

All items showed adequate inter-rater reliability (ICC≥.75).[42] The two item personality scales showed adequate internal consistency reliability ($\alpha \geq .80$). Ratings from meditation teachers and undergraduates were highly correlated ($rs \geq .66$, $ps < .001$) and rater type (i.e.,

meditation teacher vs. undergraduate) did not moderate the association between group (LTM vs. MNP at baseline, MBSR vs. HEP vs. WL in ANOVAs) and ratings (*p*s>.050). Ratings were therefore combined to increase reliability.[37]

Means and standard deviations for the six constructs assessed are included in Tables 3 and 4. At baseline, LTMs were rated more highly than MNPs on conscientiousness, comfortable in their own skin, and mindful and were rated lower than MNPs on neuroticism (absolute value of *d*s=0.61 to 0.70, *p*s<.050; Figure 1, Table 3). No differences were observed for ratings of agreeableness or attractiveness (*p*s>.050). Results were unchanged controlling for observable confounds, with the exception of two models (LTM status predicting comfortable and agreeableness) in which LTM status became a marginally significant predictor (*p*s=.085, .089, respectively) when confounds were statistically controlled. A sensitivity analysis separated ratings from undergraduates and meditation teachers. Using undergraduate ratings only, LTMs were rated as more conscientious and comfortable than MNPs, but no longer less neurotic (*p*=.052). Using meditation teacher ratings only, LTMs were rated as less anxious and more mindful, but no longer more comfortable (*p*=.161).

Time by group interaction terms were used to model changes in observer ratings for MNPs over the course of intervention (MBSR, HEP, WL; Table 4). No time by group effects were observed on any of the six dimensions assessed (*p*s>.050). Results remained unchanged controlling for observable confounds. Exploratory *post hoc* analyses examined time by group interactions for each pairing (i.e., MBSR vs. HEP, MBSR vs. WL, etc.) and pre-post within-group changes. No time by group effects or pre-post within-group changes were observed (*p*s>.050).

Discussion

The current study examined the association of long- and short-term meditation training with observer ratings of personality and related dimensions. Photographs were taken of LTMs and MNPs at baseline, and again following MNPs' completion of training based on randomization to a mindfulness meditation training (MBSR), an active control condition (HEP), or a waitlist. High inter-rater reliability was obtained across all rating dimensions. Results supported the possibility that long-term meditation training may be linked with perceived differences in neuroticism and conscientiousness by strangers despite minimal information. Significant differences in the moderate range were observed on ratings of these personality traits, as well as on novel items intended to assess constructs specific to meditation (comfortable in their own skin, mindful). While results do not imply these reflect actual differences between LTMs and MNPs on these traits (and the broader interpersonal perception literature suggests ratings of internal states by strangers may not reflect actual differences),[23,27] they also do not appear to reflect merely globally positive ratings of LTMs. In particular, LTMs and MNPs did not differ on ratings of agreeableness and attractiveness. Further, results did not appear to be primarily driven by differences in observable confounds (e.g., age, attractiveness). Results were most robust when ratings from undergraduates and meditation teachers were combined.

In contrast, no evidence was found suggesting effects of short-term meditation training on observer ratings. In fact, neither MBSR nor the active control condition (HEP) differed from the waitlist group in observer-rated changes over the course of the eight-week interventions. Thus, if meditation is associated with changes in observers' perception of personality (which our design cannot demonstrate for LTMs due to non-random assignment), these effects may be restricted to long- rather than short-term training.

The possibility that long-term meditation practice may be associated with observer's perception of neuroticism and conscientiousness is intriguing, albeit qualified by a lack of randomization of participants to the LTM and MNP conditions. Interpersonal perceptions matter; they show incremental validity beyond self-report for predicting a variety of outcomes (e.g., teacher effectiveness, academic performance)[23,24] and are not confounded with known biases in self-report (e.g., social desirability) that may be particularly pernicious when assessing internal processes impacted by meditation training (e.g., mindfulness).[13,43] Our results support the notion that long-term meditation practice is associated with more favorable perceptions by others in still photographs, on dimensions with potentially important intra- and interpersonal consequences.[44] The ways in which these perceptions made from still photographs may relate to perception in daily life is not clear from the current study, of course. Nonetheless, it is possible that LTMs may experience a more welcoming interpersonal environment, vis-à-vis others' initial impressions.[22] Should that be the case, improved interpersonal interactions could be both an outcome of long-term practice as well as a mechanism through which long-term practice yields benefits in other domains (e.g., well-being, quality of life).

It is important to highlight that perceived differences between LTMs and MNPs were evident using a very small sample of expressive behavior (i.e., still photographs). While this may reflect the potency of potential differences (i.e., they could be detected from minimal information), the observed effect should be contextualized within the broader thin slices and zero acquaintance ratings literature. Meta-analytic evidence has highlighted the modest reliability for ratings by strangers of internal states based on still visual cues (e.g., r_{rr} = .25 for emotional stability) as well as the low self-other agreement for ratings by strangers of internal states (e.g., r = .08 for emotional stability, r = .12 for openness to experience).[23] Therefore, perceived

differences must be interpreted as simply that – differences in perception – and may or may not reflect differences in LTMs' and MNPs' actual internal states. This is particularly so given the small amount of information available to observers (e.g., still photographs). Again, still photographs are clearly not analogous to how interpersonal perception typically occurs within daily life – we almost always have access to considerably more information when interacting with others. However, our results nonetheless suggest that observers somehow perceived LTMs more favorably across several domains highly relevant to meditation training, despite the limited information provided.

It would be worthwhile continuing to unpack these initial findings in future studies. One potentially fruitful future direction implied by the Realistic Accuracy Model would be obtaining samples of behavior in situations with higher relevance to the traits being assessed which may be more likely to provide available relevant information that could be detected and utilized by raters to make valid judgments.[26] Recent work has shown that ratings of neuroticism are more accurate in trait-relevant situations (e.g., socially stressful situations).[27] A potential future direction could be obtaining samples of behavior in such situations (e.g., Trier Social Stress Task),[9,45] ideally within the context of random assignment to short-term meditation training. Perhaps tasks could be used and/or developed that provide information about other aspects impacted by meditation training assessed here (e.g., mindfulness, embodiment, conscientiousness). It could also be illuminating to determine which aspects of behavior raters are using as cues to differentiate LTMs and MNPs or the short-term effects of training. This could be done using machine learning and social sensing technologies.[46] Future studies could also continue to explore potential signals detectable through still photographs, perhaps opening the door to measurement

strategies that have not been widely implemented but could hold promise for the detection of emotional signals in daily life (e.g., examining facial expressions obtained through smartphones). A future study with a larger sample of LTMs could also explore potential dose effects of meditation on observer ratings of personality across LTMs (i.e., is more training associated with more positive perceptions). The small sample of LTMs in the current study (n = 16) prohibited a proper assessment of this possibility.

The lack of short-term training effects is harder to interpret and could be due to a genuine lack of impact or an insensitivity of our observer-rated measures to these potentially more subtle changes. Pre-post analyses also relied on a subset of participants, which reduced our statistical power to detect effects (although very small pre-post effects suggest power is not the sole issue). It seems the most prudent conclusion is that short-term meditation training does not impact interpersonal perceptions of personality observable through photographs. As noted above, it could be valuable to assess the impact of short-term training in future studies through samples of behavior drawn from contexts more likely to provide cues reflecting target constructs of interest (e.g., socially stressful situations, tasks requiring conscientiousness).

Key limitations of the current study include a lack of random assignment to LTM and MNP conditions (introducing a risk of selection bias) and a relatively modest number of LTMs. The available sample was below that for which the power analysis was conducted, which had assumed a moderate-to-large effect size (d = 0.74). Thus, it is very likely that the current study was underpowered to detect more modest effects. Our use of photographs may have obscured changes detectable through a thicker slice of behavior (e.g., video recordings) or behaviors with higher relevance to the constructs of interest. We also relied primarily on personality dimensions which may be less sensitive to short-term training, given their trait-like nature. While we

focused on dimensions theoretically linked to meditation training and embodiment, we did not assess two of the five Big Five dimensions (i.e., extraversion, openness). Neither group of raters were experts in personality, so our results merely indicate a lay interpretation of the items that were rated, rather than a scientifically-based understanding of the specific personality dimensions assessed (although the Big Five items were drawn from self-report measures used to assess personality in the general population).[36] Further, the majority of our sample of raters were young, undergraduate students (mean age = 19.07). This group may be prone to a variety of perceiver effects that introduced bias into the ratings (e.g., associated with rating targets who are on average 30 years older), raising questions about the degree to which the ratings of photographs in the current study may generalize to interpersonal perceptions made by individuals of a wider variety of ages in the context of daily life.

These limitations notwithstanding, we believe this is the first study to demonstrate that naïve observers perceive LTMs differently than MNPs on several domains theoretically linked with mindfulness training, even through very minimal samples of behavior (i.e., still photographs). This work highlights the possibility that LTMs are perceived differently in daily life, which in theory could both represent an outcome of long-term training as well as a potential mechanism through which long-term training confers psychological and interpersonal benefits. Future work investigating the interpersonal effects of mindfulness training is warranted.

Acknowledgments

The authors would like to thank William Hoyt for his valuable input on design of the observer rating paradigm and comments on a previous draft of this manuscript and Donal MacCoon, David Bachhuber, Tammi Kral, Amy Cayo, and Melissa Rosenkranz who contributed to planning and execution of the larger study from which targets were drawn. This work was supported by a grant from the National Center for Complementary and Integrative Health (NCCIH) P01AT004952 to Richard J. Davidson and a grant from the Mind & Life Institute (Varela Award) to Simon B. Goldberg. Richard J. Davidson is the founder, president, and serves on the board of directors for the non-profit organization, Healthy Minds Innovations, Inc.

References

1. Clarke TC, Barnes PM, Black LI, Stussman BJ, Nahin RL. Use of yoga, meditation, and chiropractors among U.S. adults aged 18 and over. Hyattsville, MD: National Center for Health Statistics; 2018.

2. Goldberg SB, Tucker RP, Greene PA, Davidson RJ, Wampold BE, Kearney DJ, et al. Mindfulness-based interventions for psychiatric disorders: A systematic review and meta-analysis. Clin Psychol Rev. 2018 Feb;59:52–60.

3. Goyal M, Singh S, Sibinga EMS, Gould NF, Rowland-Seymour A, Sharma R, et al. Meditation Programs for Psychological Stress and Well-being: A Systematic Review and Meta-analysis. JAMA Intern Med. 2014 Mar 1;174(3):357–68.

4. Khoury B, Sharma M, Rush SE, Fournier C. Mindfulness-based stress reduction for healthy individuals: A meta-analysis. J Psychosom Res. 2015 Jun 1;78(6):519–28.

5. Lykins E, Baer RA. Psychological functioning in a sample of long-term practitioners of mindfulness meditation. J Cogn Psychother. 2009;23(3):226–41.

6. Fox KCR, Dixon ML, Nijeboer S, Girn M, Floman JL, Lifshitz M, et al. Functional neuroanatomy of meditation: A review and meta-analysis of 78 functional neuroimaging investigations. Neurosci Biobehav Rev. 2016 Jun 1;65:208–28.

7. Kaliman P, Álvarez-López MJ, Cosín-Tomás M, Rosenkranz MA, Lutz A, Davidson RJ. Rapid changes in histone deacetylases and inflammatory gene expression in expert meditators. Psychoneuroendocrinology. 2014 Feb 1;40:96–107.

8. Lutz A, Greischar LL, Rawlings NB, Ricard M, Davidson RJ. Long-term meditators self-induce high-amplitude gamma synchrony during mental practice. Proc Natl Acad Sci. 2004 Nov 16;101(46):16369–73.

9. Rosenkranz MA, Lutz A, Perlman DM, Bachhuber DRW, Schuyler BS, MacCoon DG, et al. Reduced stress and inflammatory responsiveness in experienced meditators compared to a matched healthy control group. Psychoneuroendocrinology. 2016 Jun 1;68:117–25.

10. Lykins ELB, Baer RA, Gottlob LR. Performance-Based Tests of Attention and Memory in Long-Term Mindfulness Meditators and Demographically Matched Nonmeditators. Cogn Ther Res. 2012 Feb 1;36(1):103–14.

11. Sedlmeier P, Eberth J, Schwarz M, Zimmermann D, Haarig F, Jaeger S, et al. The psychological effects of meditation: A meta-analysis. Psychol Bull. 2012;138(6):1139–71.

12. Brown KW, Ryan RM. The benefits of being present: Mindfulness and its role in psychological well-being. J Pers Soc Psychol. 2003 Apr;84(4):822–48.

13. Goldberg SB, Tucker RP, Greene PA, Simpson TL, Hoyt WT, Kearney DJ, et al. What can we learn from randomized clinical trials about the construct validity of self- report measures of mindfulness? A meta-analysis. Mindfulness. 2019;

14. Giluk TL. Mindfulness, Big Five personality, and affect: A meta-analysis. Personal Individ Differ. 2009 Dec 1;47(8):805–11.

15. Kabat-Zinn J. Full catastrophe living: Using the wisdom of your body and mind to face stress, pain, and illness (Revised ed.). New York: Random House; 2013.

16. Leiberg S, Klimecki O, Singer T. Short-Term Compassion Training Increases Prosocial Behavior in a Newly Developed Prosocial Game. PLOS ONE. 2011 Mar 9;6(3):e17798.

17. Weng HY, Fox AS, Shackman AJ, Stodola DE, Caldwell JZK, Olson MC, et al. Compassion Training Alters Altruism and Neural Responses to Suffering. Psychol Sci. 2013 Jul 1;24(7):1171–80.

18. Fredrickson BL, Cohn MA, Coffey KA, Pek J, Finkel SM. Open hearts build lives: Positive emotions, induced through loving-kindness meditation, build consequential personal resources. J Pers Soc Psychol. 2008 Nov;95(5):1045–62.

19. Hutcherson CA, Seppala EM, Gross JJ. Loving-kindness meditation increases social connectedness. Emotion. 2008;8(5):720–4.

20. Lindsay EK, Young S, Brown KW, Smyth JM, Creswell JD. Mindfulness training reduces loneliness and increases social contact in a randomized controlled trial. Proc Natl Acad Sci. 2019 Feb 26;116(9):3488–93.

21. Davidson RJ, Kaszniak AW. Conceptual and Methodological Issues in Research on Mindfulness and Meditation. Am Psychol. 2015 Oct;70(7):581–92.

22. Harker L, Keltner D. Expressions of positive emotion in women's college yearbook pictures and their relationship to personality and life outcomes across adulthood. J Pers Soc Psychol. 2001 Jan;80(1):112–24.

23. Connelly BS, Ones DS. An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. Psychol Bull. 2010;136(6):1092–122.

24. Ambady N, Bernieri FJ, Richeson JA. Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. In: Advances in Experimental Social Psychology [Internet]. Academic Press; 2000 [cited 2019 Mar 28]. p. 201–71. Available from: http://www.sciencedirect.com/science/article/pii/S0065260100800064

25. Hirschmüller S, Link to external site this link will open in a new window, Egloff B, Link to external site this link will open in a new window, Nestler S, Back MD. The dual lens model: A comprehensive framework for understanding self–other agreement of personality judgments at zero acquaintance. J Pers Soc Psychol. 2013 Feb;104(2):335–53.

26. Funder DC. Accurate Personality Judgment. Curr Dir Psychol Sci. 2012 Jun 1;21(3):177–82.

27. Hirschmüller S, Egloff B, Schmukle SC, Nestler S, Back MD. Accurate Judgments of Neuroticism at Zero Acquaintance: A Question of Relevance. J Pers. 2015;83(2):221–8.

28. Funder DC. On the accuracy of personality judgment: A realistic approach. Psychol Rev. 1995 Oct;102(4):652–70.

29. Vazire S. Who knows what about a person? The self–other knowledge asymmetry (SOKA) model. J Pers Soc Psychol. 2010 Feb;98(2):281–300.

30. Kenny DA. Interpersonal perception: A social relations analysis. New York: Guilford Press; 1994.

31. Lee K, Ashton MC. Acquaintanceship and self/observer agreement in personality judgment. J Res Personal. 2017 Oct 1;70:1–5.

32. Choi Y, Karremans JC, Barendregt H. The happy face of mindfulness: Mindfulness meditation is associated with perceptions of happiness as rated by outside observers. J Posit Psychol. 2012 Jan 1;7(1):30–5.

33. Farb N, Daubenmier JJ, Price CJ, Gard T, Kerr C, Dunn B, et al. Interoception, Contemplative Practice, and Health. Front Psychol [Internet]. 2015 Jun 1 [cited 2019 Mar 28];6. Available from: https://doaj.org

34. MacCoon DG, Imel ZE, Rosenkranz MA, Sheftel JG, Weng HY, Sullivan JC, et al. The validation of an active control intervention for Mindfulness Based Stress Reduction (MBSR). Behav Res Ther. 2012 Jan 1;50(1):3–12.

35. Naumann LP, Vazire S, Rentfrow PJ, Gosling SD. Personality Judgments Based on Physical Appearance. Pers Soc Psychol Bull. 2009 Dec 1;35(12):1661–71.

36. Gosling SD, Rentfrow PJ, Swann WB. A very brief measure of the Big-Five personality domains. J Res Personal. 2003 Dec 1;37(6):504–28.

37. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. Psychol Bull. 1979;86(2):420–8.

38. Cohen J. A power primer. Psychol Bull. 1992;112(1):155.

39. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B Methodol. 1995;57(1):289–300.

40. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2018.

41. Glickman ME, Rao SR, Schultz MR. False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies. J Clin Epidemiol. 2014 Aug 1;67(8):850–7.

42. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016 Jun 1;15(2):155–63.

43. Grossman P. Defining mindfulness by how poorly I think I pay attention during everyday awareness and other intractable problems for psychology's (re)invention of mindfulness: Comment on Brown et al. (2011). Psychol Assess. 2011 Dec;23(4):1034–40.

44. Friedman HS, Kern ML. Personality, Well-Being, and Health. Annu Rev Psychol. 2014 Jan 3;65(1):719–42.

45. Kirschbaum C, Pirke K-M, Hellhammer DH. The 'Trier Social Stress Test' – A Tool for Investigating Psychobiological Stress Responses in a Laboratory Setting. Neuropsychobiology. 1993;28(1–2):76–81.

46. Schmid Mast M, Gatica-Perez D, Frauendorfer D, Nguyen L, Choudhury T. Social Sensing for Psychology: Automated Interpersonal Behavior Assessment. Curr Dir Psychol Sci. 2015 Apr 1;24(2):154–60.

Table 1. Sample demographics

	Targets		Raters	
Demographic Variable	LTM (n = 16)	MNP (n = 83)	Undergraduates (n = 86)	Meditation Teachers (n = 25)
Age, mean(SD)	50.62 (9.56)	48.79 (11.13)	19.07 (3.37)	52.80 (12.25)
Female, n(%)	8 (50)	53 (63.9)	54 (62.8)	9 (36.0)
Race/ethnicity				
non-Hispanic white, n(%)	14 (87.5)	76 (91.6)	55 (64.0)	20 (80.0)
non-Hispanic black, n(%)	0 (0.0)	0 (0.0)	5 (5.8)	1 (4.0)
Asian, n(%)	2 (12.5)	2 (2.4)	15 (17.4)	0 (0.0)
Native American, n(%)	0 (0.0)	0 (0.0)	1 (1.2)	0 (0.0)
Hispanic, any race, n(%)	0 (0.0)	4 (4.8)	1 (1.2)	1 (4.0)
More than one race/ethnicity, n(%)	0 (0.0)	1 (1.2)	8 (9.3)	1 (4.0)
Did not want to respond, n(%)	0 (0.0)	0 (0.0)	1 (1.2)	2 (8.0)

Note: LTM = long-term meditator; MNP = meditation naïve participant.

Table 2. Descriptions of six surveys

Survey	Sample	Traits assessed
1	21 undergraduates	[a]Calm/emotionally stable
2	21 undergraduates	Attractive
3	22 undergraduates	[a]Anxious/easily upset, [a]Sympathetic/warm, [a]Dependable/self-disciplined
4	22 undergraduates	[a]Critical/quarrelsome, [a]Disorganized/careless, Comfortable in their own skin
5	14 meditation teachers	[a]Anxious/easily upset
6	11 meditation teachers	[a]Sympathetic/warm, Comfortable in their own skin, Mindful

Note: [a]Six items drawn from Ten Item Personality Inventory (Gosling et al., 2003) assessing conscientiousness, agreeableness, and neuroticism.

Table 3. LTMs versus MNPs at baseline

Rating Domain	LTM	MNP	*d*	*p*
Attractive	2.81 (0.72)	2.67 (0.58)	0.23	0.406
Agreeable	4.26 (0.85)	3.93 (0.78)	0.42	0.162
Conscientious	4.54 (0.64)	4.18 (0.49)	0.70*	0.039
Neurotic	3.24 (0.65)	3.64 (0.66)	-0.61*	0.045
Comfortable	4.62 (0.74)	4.18 (0.70)	0.62*	0.045
Mindful	4.75 (0.79)	4.31 (0.60)	0.69*	0.039

Note: Sample sizes: LTM = 16, MNP = 83. LTM = Long-term meditation practitioners; MNP = meditation naïve participants at pre-test; *d* = Cohen's *d*; Comfortable = "comfortable in their own skin"; Big-Five personality dimensions assessed using two-item scales drawn from the Ten Item Personality Inventory;[36] *p* = *p*-values from independent samples t-test, adjusted for false discovery rate;[39] *$p < .050$.

Table 4. Pre-post comparison for MNPs randomized to three conditions

	MBSR			HEP			Waitlist			
Rating Domain	Pre	Post	*d*	Pre	Post	*d*	Pre	Post	*d*	*p*
Attractive	2.66 (0.64)	2.62 (0.65)	-0.06	2.56 (0.52)	2.55 (0.54)	-0.02	2.81 (0.58)	2.77 (0.57)	-0.07	0.991
Agreeable	3.66 (0.72)	3.69 (0.74)	0.04	3.93 (0.81)	4.17 (0.79)	0.3	4.20 (0.74)	4.25 (0.82)	0.06	0.978
Conscientious	4.07 (0.40)	4.08 (0.50)	0.02	4.08 (0.58)	4.18 (0.57)	0.17	4.39 (0.42)	4.36 (0.54)	-0.06	0.978
Neurotic	3.87 (0.64)	3.85 (0.61)	-0.03	3.65 (0.69)	3.39 (0.62)	-0.4	3.40 (0.58)	3.36 (0.62)	-0.07	0.978
Comfortable	3.94 (0.71)	4.01 (0.71)	0.1	4.18 (0.71)	4.41 (0.70)	0.33	4.41 (0.64)	4.51 (0.72)	0.15	0.978
Mindful	4.33 (0.69)	4.28 (0.57)	-0.08	4.30 (0.56)	4.40 (0.61)	0.17	4.31 (0.55)	4.30 (0.60)	-0.02	0.978

Note: Sample sizes: MBSR = 27, HEP = 29, waitlist = 27. MNP = meditation naïve participants; MBSR = Mindfulness-Based Stress Reduction; HEP = Health Enhancement Program; Pre = pre-test; Post = post-test; *d* = within-group (i.e., pre-post) Cohen's *d*; Comfortable = "comfortable in their own skin"; Big-Five personality dimensions assessed using two-item scales drawn from the Ten Item Personality Inventory;[36] *p* = *p*-values from time X group ANOVA models, adjusted for false discovery rate;[39] $*p < .050$.

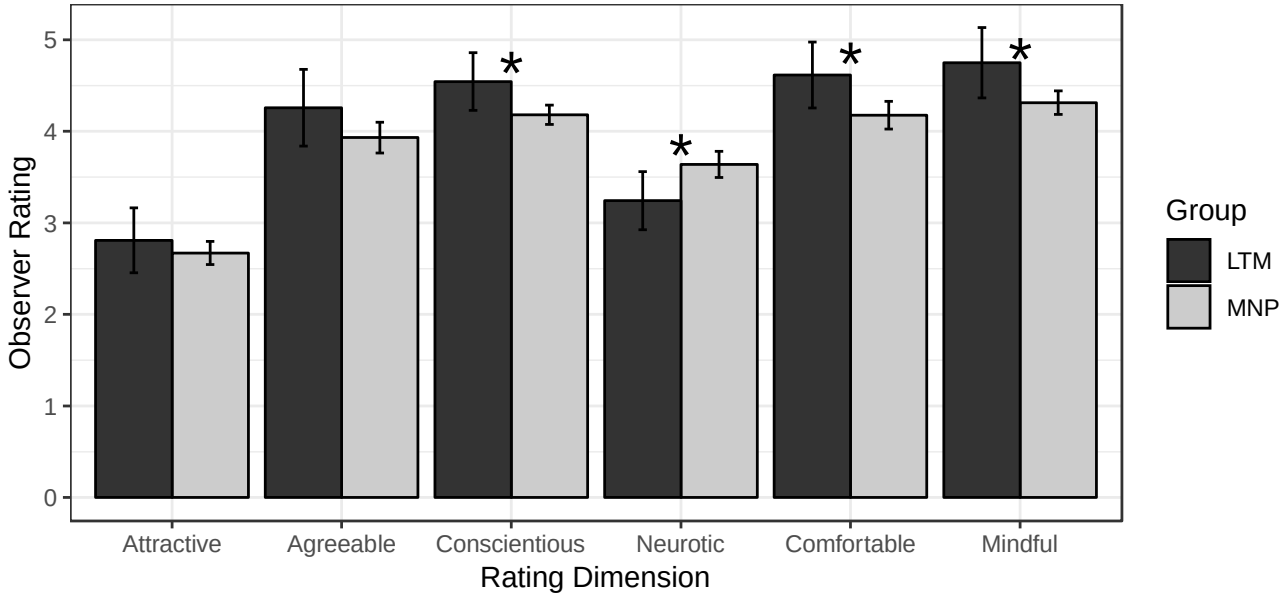

Figure 1. Long-term meditation (LTM) practitioners are rated as less neurotic and more conscientious, "comfortable in their own skin," and mindful compared to meditation naïve participants (MNP). Target sample includes 16 LTMs and 83 MNPs. Rater sample includes 25 meditation teachers and 86 undergraduates.

Supplemental Materials Table 1. Sample R code for computing intraclass correlation coefficient (ICC3)

```
vp <- function(data.g, formula.g) {   # variance partitioning
  lmer.out <- lmer(data = data.g, formula = formula.g)
  var.residual<-attr(VarCorr(lmer.out), "sc")^2
  var.comp <- ldply(VarCorr(lmer.out))
  names(var.comp) <- c("Factor", "Variance")
  var.comp <- rbind(var.comp, data.frame("Factor" = "Residual", "Variance" =
var.residual))
  var.comp$Percent <- round(var.comp$Variance / sum(var.comp$Variance) * 100, 1)
  attr(var.comp, "mer") <- lmer.out
  class(var.comp) <- c("data.frame", "G")
  var.comp
}

#T1_calm
oneFac <- reshape(data=df[,c("SID",paste("WB",sids,"_T1_calm", sep=""))],
                  varying=paste("WB",sids,"_T1_calm", sep=""),
v.names=c("rating"),timevar="Target",sep="_", direction="long")
oneFac$id <- NULL; oneFac$Rater <- factor(oneFac$SID); oneFac$Target <-
factor(oneFac$Target)
lme1 <- lmer(data=oneFac, formula = rating ~ 1 + (1|Rater) + (1|Target))
vp1 <- vp(data=oneFac, formula=rating ~ 1 + (1|Rater) + (1|Target)); varT <-
vp1$Variance[1]; varR <- vp1$Variance[2]; varRTe <- vp1$Variance[3]
varT/(varT + varRTe/table(table(df[!is.na(df$WB101_T1_calm),"SID"]))) #.84
```