Evidence for the Upward Spiral Stands Steady: A Response to Nickerson (2017)

Barbara L. Fredrickson[1]

& Bethany E. Kok[2]

[1]University of North Carolina at Chapel Hill, Department of Psychology and Neuroscience

[2]Max Planck Institute for Human Cognition and Brain Sciences, Department of Social Neuroscience

Author Note: Correspondence concerning this article should be addressed to Barbara L. Fredrickson, University of North Carolina at Chapel Hill, Department of Psychology and Neuroscience, CB 3270, Chapel Hill, NC 27516. E-mail: blf@unc.edu

Evidence for the Upward Spiral Stands Steady: A Response to Nickerson (2017)

In 2013, we reported in *Psychological Science* on a longitudinal field experiment in which we and our coauthors randomized participants to receive positive-emotions training over six weeks (or not), in order to test theory-driven hypotheses regarding the pathways by which positive emotions might build physical health (Kok et al., 2013; data publicly available on the Open Science Framework: https://osf.io/jazfy/). Our results revealed that, amidst the unpredictability of field settings, positive-emotions training produced statistically significant improvements in a marker of physical health, improvements that were mediated by psychological processes. Nickerson (in press), in her Commentary and based on her reanalyses of our data, claims that we have overstated our findings. As detailed below, we find the empirical basis for Nickerson's claim to be questionable.

Considering the Influence of Statistical Outliers

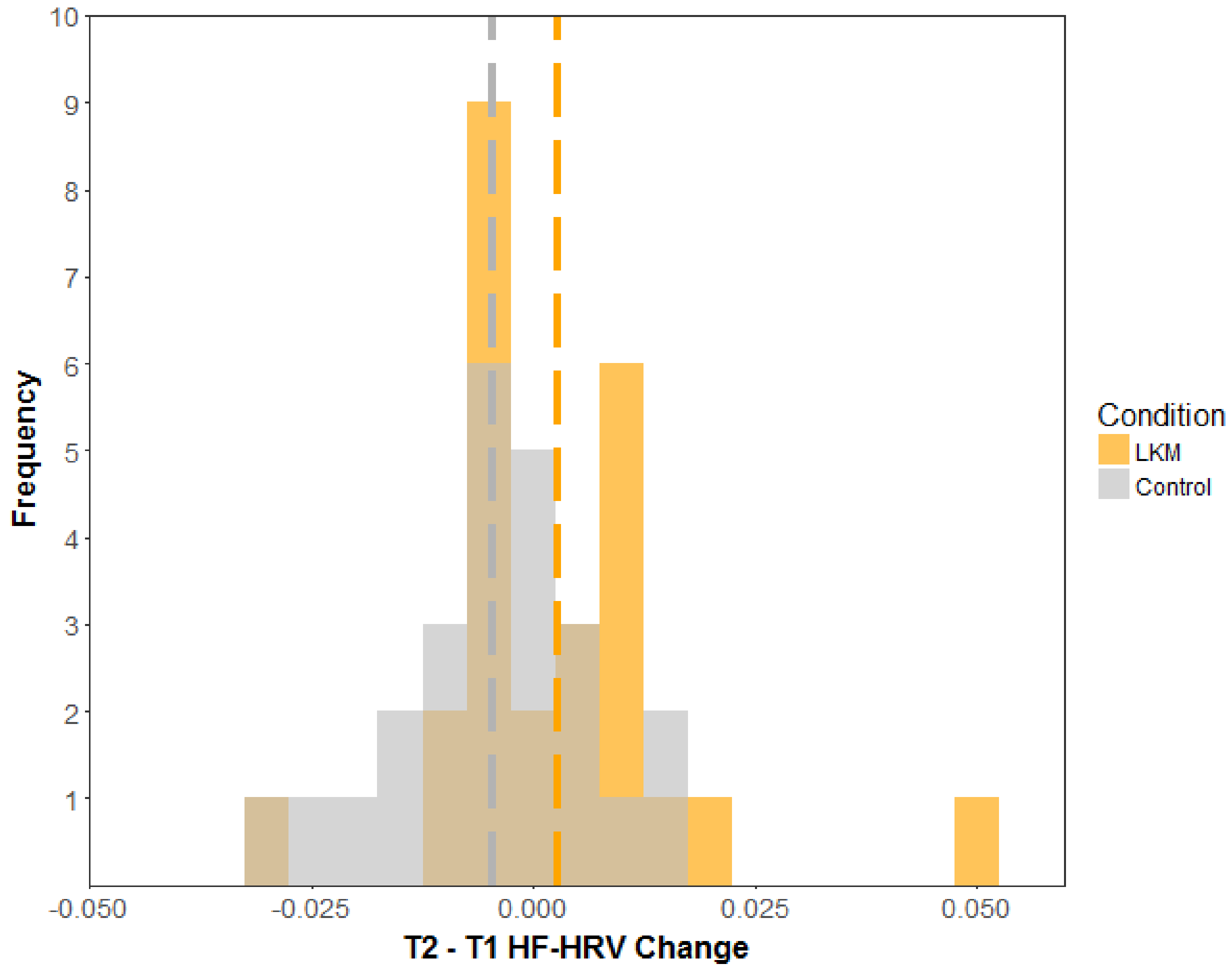
In discussing "unusual participants," Nickerson argues that extreme and implausible values for high-frequency heart rate variability (HF-HRV) for seven participants led us to overstate the effects of the positive-emotions training deployed, known as loving-kindness meditation (LKM). Nickerson, however, uses a method of identifying influential observations that does not fit the analysis we conducted. Specifically, although Nickerson characterizes seven datapoints as "unusual," she spotlights five participants with the highest scores for endpoint vagal tone (Nickerson, in press, Figure 1). However, the ultimate variable of interest in the analysis we reported in our 2013 research article represents *change in vagal tone* from baseline to end-of-study (see Kok et al., 2013, Figure 2). As such, her focus on endpoint vagal tone is inappropriate.

Closer inspection of Nickerson's Figure 1 reveals that of the five datapoints that she characterizes as extreme, the three rightmost values in her Figure 1 show end-of-study values that are consistent with their corresponding baseline values. Even though these three values are high relative to the other HF-HRV scores, the change in vagal tone that they represent is not itself disproportionately high. Thus, in our statistical model, which rests on change in vagal tone, these three datapoints do not represent statistical outliers.

Nickerson also wonders how our team could disagree with the authors of a previously published Commentary on our 2013 paper (Heathers, Brown, Coyne, and Friedman, 2015) regarding the biological validity of vagal tone values. It is because – as stated in our published response to Heathers et al. (Kok & Fredrickson, 2015) – our determination of biological validity was based on having an outside expert (blind to experimental condition) inspect the raw electrocardiogram (ECG) data for each participant in question – an inspection Heathers et al. did not do. The expert we consulted was James Long, of the James Long Company, who has provided software and hardware solutions for psychophysiological research since 1979. Long determined that all values in our sample, excepting the two identified and excluded in Kok & Fredrickson (2015; Participants #557004 and #557027), were the result of biologically plausible ECG recordings and not of error.[1] As such, it is inappropriate to exclude them from analysis (Raw R-wave times along with James Long's assessments of the disputed cases are publicly

[1] We thank Nickerson for bringing to our attention one additional participant (#557055) whose data we had inadvertently failed to send to James Long for evaluation in 2015. Upon inspecting ECG data for this participant in 2017, Long noted a single ectopic beat near the beginning of the 2-min resting phase used to determine end-of-study vagal tone. This single ectopic beat had artifactually produced an atypically high HF-HRV value (i.e., 0.0106). Long recommended that we re-compute this participant's HF-HRV value by excluding the first 16 seconds of data, which contained the ectopic beat. Doing so resulted in a corrected value more typical for our sample (i.e., 0.000167). All data representations and analyses reported here use this corrected value.

available on the Open Science Framework: https://osf.io/jazfy/). The Figure below presents the histograms of change in vagal tone by experimental condition, as illustrated by the difference scores created by subtracting baseline HF-HRV from end-of-study HF-HRV.[2]



[2] We chose to illustrate change in vagal tone using difference scores in this Figure to retain the interpretability of x-axis values above and below zero as positive and negative change over time, respectively. We note that statistical models reported here and in Kok et al. (2013) predict change in vagal tone, as indexed by residualized change scores. We provide a supplementary Figure that presents corresponding condition-specific histograms based on residualized change scores at https://osf.io/jazfy/. It conveys comparable distributions, outliers, and condition-specific means.

Figure caption: Condition-specific histograms of change in vagal tone, the outcome variable in Kok et al. (2013). Mean change for each experimental condition is represented by the corresponding dashed vertical line.

Note. Change in vagal tone computed as end-of-study HF-HRV minus baseline HF-HRV, using the square-root transformed values used in Kok et al. (2013). HF-HRV = high-frequency heart rate variability.

Nickerson's approach to the seven "unusual participants" that she designates was to exclude them from her reanalysis of our data. A truism in statistics is that statistically significant findings can be rendered nonsignificant with sufficient reductions in sample size and, correspondingly, in statistical power. In this light, it is noteworthy that Nickerson's approach reduced the available sample by more than 13%.

We agree with Nickerson that the rightmost value in the LKM histogram of our outcome variable, change in vagal tone (see Figure), may plausibly function as an influential statistical outlier. We do not agree, however, with her strategy of addressing outliers by expunging data. An alternative analytic method for removing the disproportionate influence of outliers, one that does not reduce statistical power, is to Winsorize any extreme values for change in vagal tone by replacing them with values from the 95th percentile (Rivest, 1994). This approach simultaneously reduces the distortion that can be caused by outlying values while retaining most of the information provided by the datapoint (i.e., that the individual had a high value).

After Winsorizing the data (and again excluding the two cases identified as biologically implausible in Kok & Fredrickson, 2015), empirical support for the three hypotheses of our original report stands steady. Specifically, we reran the model reported in Kok et al. (2013, see

Figure 2) using the 95th percentile method of Winsorizing extreme datapoints for change in vagal tone. This method adjusted six outlying values,[3] including the rightmost value of concern to Nickerson (see Figure). The model fit continued to be acceptable (RMSEA = 0.084, 90% CI = 0.063/0.103; CFI = 0.945; TLI = 0.943), and the results continue to support our three hypotheses: (1) that the practice of LKM increased positive emotions ($b = 0.055$, $z = 3.770$, $p < .001$), with baseline vagal tone moderating this effect such that participants who began the study with higher values of HF-HRV showed greater increases in positive emotions over nine weeks' time ($b = 0.047$, $z = 3.003$, $p = .003$); (2) that increased positive emotions predicted increased perceived social connections ($b = 1.065$, $z = 4.184$, $p < .001$); and (3) that increased perceived social connections predicted larger changes in vagal tone over the course of the study ($b = 3.795$, $z = 2.008$, $p = .045$).

Within-person Analyses?

Nickerson argues that evidence for upward spirals needs to rest on within-person analyses rather than between-person analyses. We agree with Nickerson that within-person tests are of considerable interest. We find fault, however, with her claim that she provides a within-person test, and show that her between-person test rests on data stripped of much of its information value.

A key strength of Kok et al. (2013) is its experimental design, which randomized study participants either to a Control Condition or Intervention Condition (LKM). Importantly, *both* our original analysis *and* the reanalysis Nickerson offers in the latter section of her Commentary are between-person tests that compare these two groups. Even though both analytic approaches

[3] Values for change in vagal tone adjusted by Winsorizing included those for participants #557012, #557016, #557037, #557059, #557065, and #557070.

– ours and hers – include derived variables to represent within-person changes (i.e., latent scores for our analysis; computed indices of change in her reanalysis), both statistical approaches ultimately compare these within-person variables using between-person tests. Statistical experts argue that any effects assessed using single derived values per person to represent within-person change on a repeated measure – whether as independent or dependent variables – are unambiguously between-person effects (Curran, Howard, Bainter, Lane & McGinley, 2014). So, although Nickerson claims that her reanalysis provides a test of within-person effects, it does not.

The primary difference between our original analysis and Nickerson's reanalysis is that she chose to drastically reduce the information value of our data before conducting her reanalysis. She did this by dichotomizing the data. (We note that only one of Nickerson's 195 computed indices of change – across 3 variables for each of 65 participants – resulted in a value of "no change." As such, Nickerson's trichotomization strategy in effect reduces to dichotomization.) Dichotomizing continuous data is strongly discouraged by statistical experts because it has been decisively linked to a wide variety of adverse statistical consequences, including (1) loss of information about individual differences, (2) loss of effect size and power, and (3) loss of measurement reliability (MacCallum, Zhang, Preacher & Rucker, 2002). As such, Nickerson's tallies (in the rightmost column of her Table 1) drastically diminish the information value of the original data. With her extreme approach to data reduction, it is foreseeable that statistical significance would be lost in most any analysis of these tallies. To sum, Nickerson's approach of conducting a between-person test on dichotomized data is far less sensitive than our original analysis that used the full information within our continuous data.

Although the trio of hypotheses tested in Kok et al. (2013) continue to be supported by the initial data we collected in 2007, we acknowledge that the scientific study of upward spiral processes remains in its infancy (for a review, see Fredrickson & Joiner, in press). Only a handful of studies, for instance, have incorporated biological variables into tests of the prospective and reciprocal relations that form upward spiral dynamics (i.e., Burns et al., 2008; Kok & Fredrickson, 2010).[4] Whereas multiple laboratories report cross-sectional positive associations between cardiac vagal tone and well-being (both physical and mental; e.g., Bhattacharyya, Whitehead, Rakhit, & Steptoe, 2008; Geisler, Vennewald, Kubiak, & Weber, 2010; Marsland, Gianaros, Prather, Jennings, Neumann, & Manuck, 2007; Oveis, Cohen, Gruber, Shiota, Haidt, & Keltner, 2009; Wang, Lu, & Qin, 2013; yet for null results see Silvia, Jackson, & Sopko, 2014 and Sloan et al., 2017; for nonlinear associations see Kogan, Gruber, Shallcross, Ford & Mauss, 2013), to date, no direct or conceptual replications of the longitudinal, field experiment reported in Kok et al. (2013) have been published. Of relevance for further longitudinal investigations of upward spiral dynamics, new statistical tools are now available to provide rigorous and simultaneous tests of between-person and within-person effects over time (e.g., group iterative multiple model estimation, see Beltz, Wright, Sprague, & Molenaar, 2016; latent curve models with structured residuals, see Curran, Howard, Bainter, Lane, & McGinley, 2014). These advanced statistical tools, however, require larger samples (e.g., $N = 250$) and more frequent repeated assessments (e.g., $T = 5$) than are available in the dataset on which Kok et al. (2013) was based. Rigorous and well-powered tests are thus still needed to further examine

---

[4] We note that Kok & Fredrickson (2010) examined individual differences in upward spiral dynamics using the same dataset as used in Kok et al. (2013), with statistical controls for group assignment. Across these two publications, we conclude that upward spirals between positive emotions and cardiac vagal tone occur naturally to greater or lesser degrees (Kok & Fredrickson, 2010) and that individuals' upward spirals can be nudged to advance more swiftly with random assignment to positive emotions training (Kok et al., 2013).

whether and how self-generated positive emotions improve vagal tone and other objective markers of physical health, and whether they do so in an upward spiral dynamic.

References

Beltz, A. M., Wright, A, G. C., Sprague, B. N., & Molenaar (2016). Bridging the nomothetic and ideographic approaches to the analysis of clinical data. *Assessment, 23(4),* 447-458.

Bhattacharyya, M. R., Whitehead, D. L., Rakhit, R., & Steptoe, A. (2008). Depressed mood, positive affect, and heart rate variability in patients with suspected coronary artery disease. *Psychosomatic Medicine, 70*, 1020-1027.

Burns, A. B., Brown, J. S., Sachs-Ericsson, N., Plant, E. A., Curtis, J. T., Fredrickson, B. L., & Joiner, T. E. (2008). Upward spirals of positive emotion and coping: Replication, extension, and initial exploration of neurochemical substrates. *Personality and Individual Differences, 44*(2), 360-370. doi:10.1016/j.paid.2007.08.015

Curran, P. J., Howard, A. L., Bainter, S. A., Lane, S. T., & McGinley, J. S. (2014). The separation of between-person and within-person components of individual change over time: A latent curve model with structured residuals. *Journal of Consulting and Clinical Psychology, 82(5)*, 879-894.

Fredrickson, B. L. & Joiner, T. (in press). Reflections on positive emotions and upward spirals. *Perspectives on Psychological Science*.

Geisler, F. C. M., Vennewald, N., Kubiak, T., & Weber, H. (2010). The impact of heart rate variability on subjective well-being is mediated by emotion regulation. *Personality and Individual Differences, 49*, 723-728.

Heathers, J. A. J., Brown, N. J. L., Coyne, J. C. & Friedman, H. L. (2015). The elusory upward spiral: A reanalysis of Kok et al., (2013). *Psychological Science, 26*, 1140-1143.

Kogan, A., Gruber, J., Shallcross, A. J., Ford, B. Q., & Mauss, I. B. (2013). Too much of a good thing? Cardiac vagal tone's nonlinear relationship with well-being. *Emotion, 13,* 599-604.

Kok, B. E., Coffey, K. A., Cohn, M. A., Catalino, L. I., Vacharkulksemsuk, T., Algoe, S. B., ... & Fredrickson, B. L. (2013). How positive emotions build physical health: Perceived positive social connections account for the upward spiral between positive emotions and vagal tone. *Psychological Science, 24*, 1123–1132.

Kok, B. E., & Fredrickson, B. L. (2010). Upward spirals of the heart: Autonomic flexibility, as indexed by vagal tone, reciprocally and prospectively predicts positive emotions and social connectedness. *Biological Psychology, 85*(3), 432-436. doi:10.1016/j.biopsycho.2010.09.005

Kok, B. E., & Fredrickson, B. L. (2015). Evidence for the upward spiral stands steady: A response to Heathers, Brown, Coyne, and Friedman (2015). *Psychological Science, 26*, 1144–1146.

MacCallum, R.C., Zhang, S., Preacher, K.J. & Rucker, D.D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7(1)*, 19-40.Sloan

Marsland, A. L., Gianaros, P. J., Prather, A. A., Jennings, R., Neumann, S. A., Manuck, S. B. (2007). Stimulated production of proinflammatory cytokines covaries inversely with heart rate variability. *Psychosomatic Medicine, 69*, 709-716.

Nickerson, C. A. (in press). Overstating the effects of loving-kindness meditation: Comment on Kok et al. (2013). *Psychological Science.*

Oveis, C., Cohen, A. B., Gruber, J., Shiota, M. N., Haidt, J., & Keltner, D. (2009). Resting respiratory sinus arrhythmia is associated with tonic positive emotionality. *Emotion, 9,* 265-270.

Rivest, L. (1994). Statistical properties of Winsorized means for skewed distributions. *Biometrika, 81*, 373-383.

Silvia, P. J., Jackson, B. A., & Sopko, R. S. (2014). Does baseline heart rate variability reflect stable positivity emotionality? *Personality and Individual Differences, 70*, 183-187.

Sloan, R. P., Schwarz, E., McKinley, P. S., Weinstein, M., Love, G., Ryff, C. et al. (2017). Vagallly-mediated heart rate variability and indices of well-being: Results of a nationally representative study. *Health Psychology, 36*, 73-81.

Wang, Z., Lu, W., & Qin, R. (2013). Respiratory sinus arrhythmia is associated with trait positive affect and positive emotional expressivity. *Biological Psychology, 93*, 190-196.

Author Contributions

B. L. Fredrickson and B. E. Kok each drafted sections of the manuscript and each provided critical revisions. Both authors approved the final version of the manuscript for submission.

Acknowledgements

The authors wish to thank Aaron Boulton, Sara Algoe, Ann Firestine, and James Long for their assistance with preparing this response.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

The data reanalyzed in this article were collected with support from a research grant awarded to Principal Investigator Barbara L. Fredrickson by the National Institute of Mental Health (MH59615) of U.S. National Institutes of Health (NIH). Support for the lead author's time came from a Research and Study Assignment from the College of Arts and Sciences at the University of North Carolina at Chapel Hill, a Cattell Sabbatical Award from the Association for Psychological Science, as well as from three research grants awarded to the lead author by the NIH. These were a National Institute for Nursing Research Grant (NR012899), an award supported by the NIH Common Fund, which is managed by the NIH Office of the Director/Office of Strategic Coordination; a National Cancer Institute Research Grant (CA170128); and a National Center for Complementary and Integrative Health Research Grant (AT007884). These funding agencies played no role in the decision to publish or the preparation of the manuscript.