BIOBEHAVIORAL THREAT SENSITIVITY AND AMYGDALA VOLUME: A TWIN NEUROIMAGING STUDY

Supplement

Contents

1. Recruitment and Pre-Selection

The MRI study reported in this article was part of a larger grant-funded research project focusing on quantification of variations in threat sensitivity through use of self-report and physiological response measures. The first phase of the larger project entailed collection of questionnaire data, via mail, from adult twins identified through the Minnesota Twin Registry, a birth-based record of all twins born in Minnesota during two major spans of time (1935-1955 and 1971-1989; Iacono et al., 1999; Lykken, Bouchard, McGue, & Tellegen, 1990); participants in this questionnaire sample consisted of 2,511 same-sex monozygotic (MZ) and dizygotic (DZ) twins, born during the years 1971-1985, who had not previously been tested in other research studies. Data for 11 different fear/fearlessness scales collected from this sample were used to formulate a report-based measurement model for the construct of dispositional fear; this model included a general fear/fearlessness factor on which all scales loaded, along with residual factors accounting for additional covariance among certain scales. Details of the participant sample for this work (N = 2,511), the fear-fearlessness model, and twin-biometric analyses of its constituent factors, are reported by Kramer, Patrick, Krueger, and Gasperi (2012).

Scores on the general factor of this fear/fearlessness model – quantified using the 55-item Trait Fear (TF-55) scale described in the main article – were used to select and recruit a subsample of twin pairs for lab testing; the purpose of the lab testing phase of the project was to identify physiological indicators of general fear/fearlessness to allow for this construct to be quantified as biobehavioral threat sensitivity (i.e., by combining physiological indices of threat reactivity with report-based fear [TF-55] scores) and to examine diagnostic correlates of threat sensitivity quantified in this manner. Data for a total of 508 twin participants were collected in this lab testing phase; as described in the paper by Yancey, Venables, and Patrick (2016)

referenced in the main article, physiological and self-report measures needed to quantify biobehavioral threat sensitivity were available for 454 of these 508 participants.

The sample for the current MRI study (n = 44) comprised a subset of the 454 lab-test participants for whom biobehavioral threat scores were computed, as described by Yancey et al. (2016). MZ twin pairs were targeted for testing because available resources and time precluded testing of both MZ and DZ twin pairs in sufficient numbers for biometric analyses to be performed. As noted in the main article, we selected one member of each twin pair (denoted 'twin A') for testing on the basis of higher or lower levels of biobehavioral threat sensitivity (without other restrictions), and we tested the co-twin of each pair (designated 'twin B') without any restrictions; the sample of 44 utilized in our reported analyses comprises all twin pairs for which structural MRI data were collected from both pair members ('A' and 'B'). Regarding the criteria for selecting twin A's, we sought to represent each quartile of the distribution of scores on the biobehavioral threat (THT) index about equally, but our ability to do so was constrained by the availability of previously-tested twin pairs for rescheduling and also by resources and time for collection of MRI test data as a whole. Notwithstanding these constraints, the scatterplots for amygdala volumes (left, right) as a function of threat sensitivity scores shown in Figure 1 of the main article indicate that we were effective in representing varying levels of biobehavioral THT in the MRI test sample.

2. Analyses Using FreeSurfer Software to Quantify Amygdala Gray Matter Volume

There are different methods for performing a volumetric analysis that could have been applied to the neuroimaging data of the current study. The two most prominent of these are voxel-based morphometry (VBM) and algorithmic segmentation using the FreeSurfer software package (Fischl, 2012). We decided *a priori* to use VBM based on findings from a recent study demonstrating greater accuracy for VBM in quantifying amygdala volume relative to

quantification based on manual segmentation and visual inspection, given the shape and location of this structure within the brain (Grimm et al., 2015). However, to directly address the issue of comparability of findings for VBM and FreeSurfer in the current dataset, we report here results from analyses using the FreeSurfer 5.3 image analysis suite (Fischl, 2012) to quantify volumes of the left and right amygdala.

Procedure for Quantifying Amygdala Gray Matter Volume. FreeSurfer morphometric procedures have demonstrated good test-retest reliability for data derived from different scanner systems with varying field strengths (Han et al., 2006; Reuter et al., 2012). The standard FreeSurfer pipeline (Fischl et al., 2004) was utilized to process MRI data from the current study and derive intracranial volumes (ICVs) for use in analyses. Specifically, amygdala subfield segmentation was performed using the new automated algorithm available in FreeSurfer 6.0 (Saygin et al., 2017). This method employs a refined probabilistic atlas constructed by three independent neuroanatomists from 10 ex vivo brains and 39 in vivo brains using both manual and automated methods. Applying Bayesian inference, the atlas is used to automatically segment the amygdala into 9 bilateral nuclei (i.e., 18 total segments). This procedure was validated on data from 374 individuals with and without Alzheimer's disease and 262 others with and without Autism Spectrum Disorder, and demonstrated greater accuracy in identifying individuals from these special populations than prior versions of FreeSurfer.

Results using FreeSurfer-Defined Amygdala Volume. The volumetric analysis using FreeSurfer yielded similar, but not identical, results to those using VBM. When derived using Freesurfer, gray matter volumes for the left and right amygdala correlated quite strongly (r = .70, $p < .001$), but to a somewhat lesser degree than when derived using VBM (r = .80). Similarly, twin concordances for gray matter volumes of the left and right amygdala were moderate and significant when derived using FreeSurfer (ICCs = .42 and .49, respectively, Fs(21, 21) = 2.42

and 2.94, *p*s = .01 and .02), but lower than when derived using VBM (ICCs = .78 and .66, *F*s(21, 21) = 8.00 and 4.84, *p*s < .001.

In addition, for FreeSurfer, as for VBM, right amygdala gray matter volume was significantly and negatively related to both THT (*r* = –.33, *p* = .03) and Social Phobia symptoms (*r* = –.30, *p* = .049), but showed a non-significant negative association with the fear disorder symptom composite (*r* = –.04, *p* = .79). For left amygdala volume, these correlations were also negative, but each fell below the .05 threshold for statistical significance (THT: *r* = –.19, *p* = .24; Social Phobia symptoms: *r* = –.16, *p* = .29; fear disorder composite: *r* = –.05, *p* = .74). Correlations for amygdala volume scores with Depression symptoms and dysphoric disorder composite scores were nonsignificantly positive for the FreeSurfer analyses (*r*s for right amygdala with Depression and dysphoric disorder composite scores = .05 and .07, respectively, *p*s = .76 and .67; *r*s for left amygdala = .16 and .22, respectively, *p*s = .29 and .16) as compared to nonsignificantly negative for the VBM analyses.

This comparative analysis appears to support our *a priori* decision to utilize the VBM quantification method, based on findings reported by Grimm et al. (2015). Within the current study, interhemispheric correlations and twin concordances were somewhat higher for amygdala volumes when quantified using VBM as compared to FreeSurfer. In addition, predicted correlations for amygdala volume scores with THT scores, and also with Social Phobia and fear disorder symptom scores, were higher in magnitude when quantified using VBM as compared to FreeSurfer.

3. Results for Control Region (Hippocampus)

Gray matter volumes (GMVs) of the left and right hippocampus, quantified using VBM, were highly correlated with corresponding GMVs for the amygdala, but not significantly

correlated with biobehavioral threat sensitivity (THT), Social Phobia or Depression symptom scores, or symptom composite scores for either fear disorders or dysphoric disorders (see Supplementary Table A). Further, regression analyses revealed that the modest, non-significant relationships evident for hippocampus GMV with THT scores, Social Phobia symptoms, and fear disorder composite scores were fully accounted for by covariance between amygdala and hippocampus GMVs. For left amygdala and left hippocampus as co-predictors of THT, the overall regression model was significant (F[2, 40] = 4.01, p = .03), with only the amygdala showing a unique predictive contribution (B = –.47, p = .02, Cohen's f^2 = .28 for amygdala, versus B = .11, p = .55, f^2 = .01 for hippocampus). For right amygdala and right hippocampus as co-predictors, the overall model R fell below significance (F[2, 40] = 2.36, p = .11), with the amygdala showing unique prediction at a trend level (B = –.34, p = .069, f^2 = .13) and the hippocampus exhibiting no unique contribution (B = .02, p = .92, f^2 = .00). Results were similar for regression models using hippocampus and amygdala GMVs (either left or right) as co-predictors of Social Phobia symptoms and fear disorder composite score. The overall model Rs were weaker, as were the regression coefficients (Bs) for the amygdala as co-predictor, but in all cases the association for the hippocampus was reduced within the model relative to its bivariate correlation, and greatly exceeded by the association for the amygdala.

4. Correlations of Scale and Physiological Components of Biobehavioral THT Index with Amygdala Volume: Findings and Interpretation

Examination of correlations with left and right amygdala volume for individual indicators of threat sensitivity (THT) within the current MRI study sample (N = 43) revealed robust negative rs for the TF-55 scale indicator (–.38 & –.36, respectively, ps = .01 and .02), commensurate with those for biobehavioral THT scores (rs = –.40 & –.32, ps = .01 and .03).

Unexpectedly, however, a composite of the three physiological indicators showed weaker, nonsignificant negative *r*s with left and right amygdala volume (*r*s = –.15 & –.06, respectively).

Our interpretation of this pattern of results is that levels of biobehavioral THT among participants tested in the current MRI study, relative to all participants within the larger (*N* = 454) sample from which they were drawn, were reflected more in TF-55 scores than in physiological indicator scores. This interpretation is based in part on the well-documented point (Dawes, 1975; see also, e.g., Elwert & Winship, 2014) that constituent indicators of a composite (e.g., factor-score) predictor variable can exhibit very different relations with one another, and with criterion measures, in subsamples drawn from the larger test sample used to formulate a prediction model. It is based also on corollary evidence from the current study sample.

More specifically: Dispositional threat sensitivity (THT) – conceptualized as an underlying (latent) trait that affects responses in different measurement modalities – was operationalized in the current study by scores on a factor reflecting variance in common among the TF-55 scale and three physiological indicators of aversive-stimulus reactivity, extracted from data for a much larger project sample (*N* = 454; Yancey et al., 2016). Within this larger sample, higher factor scores were indicative, on average, of higher levels of each of the four indicators of THT; however, for individual participants at any given level of the trait dimension as quantified by factor scores, particular indicators may determine factor scores more so than others (cf. Dawes, 1975; Ewert & Winship, 2014). As indicated in the main article, individuals were selected for MRI testing based on THT factor scores – but without specific consideration of their positions on particular THT indicators (i.e., without requiring that positions along the trait continuum be determined as much by the physiological indicators as by the TF-55 scale indicator).

As evidence for the position that the three physiological indicators were less indicative of biobehavioral THT in the MRI subsample than in the full project sample, a regression analysis using TF-55 scores and physiology-only composite scores to predict THT factor scores in the MRI

test sample indicated markedly higher representation for TF-55 scores (*B*s = .80 and .61, respectively), whereas a counterpart regression analysis for the twin sample as a whole (*N* = 454) revealed somewhat higher representation for physiology-only composite scores (*B*s = .66 versus .60). Even more direct evidence is provided by a comparison of correlations with a separate physiological composite – consisting of 2 brain and 2 facial indices of aversive response separate from those represented in the biobehavioral factor – that was used as a criterion measure in the Yancey et al. (2016) study (see Figure 2, p. 400, and accompanying text). Whereas a composite of the three physiological indicators included in the biobehavioral factor showed a robust correlation with this neurophysiological criterion in the full *N* = 454 sample ($r = .27$, $p < .001$), it showed only a weak positive correlation with this criterion measure in the current MRI sample ($r = .08$) – similar to the mean of its *r*s (in reverse) with right and left amygdalae (= –.11). The implication is that correlations of the three physiological indicators with amygdala volume may well have been higher (i.e., more robustly negative) in the full participant sample – where scores on these indicators contributed more on average to estimated levels of biobehavioral THT.

Although compelling, this interpretation is *post hoc* in nature, and therefore speculative. Further research is needed to corroborate our finding of a negative relationship between dispositional threat sensitivity and amygdala volume, and to further establish the value of a multi-method, biobehavioral approach to quantifying this trait dimension. In particular, work with larger participant samples is needed to demonstrate that a biobehavioral THT factor defined more strongly by physiological indicators than psychological-scale indicators effectively predicts variations in amygdala volume.

Supplementary References

Dawes, R. M. (1975). Graduate admission variables and future success. *Science*, 187(4178), 721-723.

Elwert, F., & Winship, C. (2014). Endogenous selection bias: The problem of conditioning on a collider variable. *Annual Review of Sociology*, *40*, 31-53.

Fischl, B. (2012). FreeSurfer. Neuroimage, 62(2), 774-781.

Fischl, B., Van Der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D. H., ... & Caviness, V. (2004). Automatically parcellating the human cerebral cortex. Cerebral cortex, 14(1), 11-22.

Grimm, O., Pohlack, S., Cacciaglia, R., Winkelmann, T., Plichta, M. M., Demirakca, T., & Flor, H. (2015). Amygdalar and hippocampal volume: a comparison between manual segmentation, Freesurfer and VBM. Journal of Neuroscience Methods, 253, 254-261.

Han, X., Jovicich, J., Salat, D., van der Kouwe, A., Quinn, B., Czanner, S., ... & Maguire, P. (2006). Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. Neuroimage, 32(1), 180-194.

Iacono, W. G., Carlson, S. R., Taylor, J., Elkins, I. J., & McGue, M. (1999). Behavioral disinhibition and the development of substance-use disorders: Findings from the Minnesota Twin Family Study. Development and Psychopathology, 11(4), 869-900.

Kramer, M. D., Patrick, C. J., Krueger, R. F., & Gasperi, M. (2012). Delineating physiologic defensive reactivity in the domain of self-report: Phenotypic and etiologic structure of dispositional fear. Psychological Medicine, 42(06), 1305-1320.

Lykken, D. T., Bouchard, T. J., McGue, M., & Tellegen, A. (1990). The Minnesota Twin Family Registry: Some Initial Findings. Acta Geneticae Medicae et Gemellologiae: Twin Research, 39(1), 35-70.

Reuter, M., Schmansky, N. J., Rosas, H. D., & Fischl, B. (2012). Within-subject template estimation for unbiased longitudinal image analysis. Neuroimage, 61(4), 1402-1418.

Saygin, Z. M., Kliemann, D., Iglesias, J. E., van der Kouwe, A. J., Boyd, E., Reuter, M., ... & Fischl, B. (2017). High-resolution magnetic resonance imaging reveals nuclei of the human amygdala: manual segmentation to automatic atlas. Neuroimage, 155, 370-382.

Yancey, J. R., Venables, N. C., & Patrick, C. J. (2016). Psychoneurometric operationalization of threat sensitivity: Relations with clinical symptom and physiological response criteria. Psychophysiology, 53(3), 393-405.

Supplementary Table A. *Correlations of gray matter volume (GMV) scores for hippocampus with amygdala GMV scores, threat sensitivity scores, Social Phobia and Depression symptom scores, and fear disorder and dysphoric disorder composite scores*

	Hippocampus GMV	
	Left	Right
Amygdala GMV		
Left	.63**	.60**
Right	.58**	.55**
Threat Sensitivity	–.18	–.17
Social Phobia symptoms	–.19	–.23
Depression symptoms	–.10	–.07
Fear Disorder composite	–.23	–.27
Dysphoric Disorder composite	–.10	–.07