To know personality is to measure it

Introducing a Dutch brief form of the Multidimensional Personality Questionnaire

Eigenhuis, A.

Chapter 6

Structural brain associations of personality do not replicate

Annemarie Eigenhuis

Jan H. Kamphuis

Andries R. van der Leij

Arjen Noordhof

Ruud Wetzels

H. Steven Scholte

Abstract

Numerous studies have shown structural brain-personality associations. However, results are heterogeneous and inconsistent. We hypothesized that many of the reported associations actually are chance findings and put this to the test by using a design that featured (a) a large representative young adult sample (N = 922); (b) an exploratory and confirmatory subsample to search and cross-validate associations. We used the Multidimensional Personality Questionnaire (MPQ) to operationalize personality. Both Gray Matter (GM) and White Matter (WM) were assessed employing Voxel-Based Morphometry (VBM) and Tract-Based Spatial Statistics (TBSS) respectively. Although a multitude of associations appeared in the exploratory subsample, none of these associations survived cross-validation in the confirmatory sample. In contrast, gender did provide replicable associations with brain structure. These results suggest that there are no reliable direct associations between personality trait scores and brain structure.

Introduction

Personality can be defined as *"an individual's characteristic patterns of thought, emotion, and behavior, together with the psychological mechanisms – hidden or not – behind those patterns"* (Funder, 2016, p. 5). In this definition a distinction is made between characteristic patterns and the psychological mechanisms behind these patterns. It is generally accepted that individual differences in the characteristic patterns can be summarized by a concise set of personality traits (Funder, 2016). The psychological mechanisms underlying these traits are currently typically sought in the brain. The quest for brain determinants of personality requires the assumption of trait-realism, which entails that traits *'exist'* independently of their measurement. Within trait-realism traits are thought to reflect psychobiological structures underlying an extended family of behavioral dispositions (Tellegen, 1991a).

An abundant number of studies relating personality trait scores to brain structure have been published (limiting ourselves to studies that employed broadband personality instruments, see e.g. Bjørnebekk et al., 2013; DeYoung et al., 2010; Kapogiannis, Sutin, Davatzikos, Costa, & Resnick, 2013; Liu et al., 2013; Lu et al., 2014; Riccelli, Toschi, Nigro, Terracciano, & Passamonti, 2017; Xu & Potenza, 2012). Many of these studies reported significant associations between personality trait scores and brain structure, suggesting a direct basis of personality in brain structure. Yet, results are heterogeneous and inconsistent. Results are heterogeneous in the sense that single traits have been reported to be associated with many different and large brain structures bearing many different functions. For example Neuroticism has been related to brain structure in many frontal areas, in the temporal lobes, the basal ganglia, in numerous gyri, the anterior cingulate, the caudate, the cerebellum etcetera. Inconsistency is shown by the observation that very seldom a specific brain structure is reported by more than one study. Of the above-mentioned associations only the associations between Neuroticism and the orbitofrontal cortex and the precentral gyrus were reported twice across the above referenced studies. Moreover, some studies find (heterogeneous) associations for all domains of personality (DeYoung et al., 2010; Kapogiannis et al., 2013; Riccelli et al., 2017), while others report on null findings for certain traits (e.g. Bjørnebekk et al., 2013; Liu et al., 2013; Xu & Potenza, 2012), or ignore some domains (e.g. Lu et al., 2014). The heterogeneity and inconsistency of the reported associations pose problems for the credibility of this line of research and make the formulation of expectations on converging and diverging patterns a challenge.

We hypothesized that the heterogeneity and inconsistency in the reported structural personality – brain associations are due to many of the associations being chance findings. Two lines of evidence support this hypothesis. First, most studies employed relatively small samples (typically $N \approx$ 50-100 with the exceptions of Bjørnebekk et al., 2013, N = 265 and Riccelli et al., 2017, N = 507). The smaller the sample size of a study is, the smaller its power which makes it hard to identify small effects. Simultaneously, the reliability of the estimation of the effect is weak (all other factors being equal such as the representativeness of the sample). The account of many relatively large effects in the literature therefore suggest that to a high degree publication bias and file drawer problems are at play (Yarkoni, 2015). Second, it was recently shown that the type I error rate (i.e. observing an association that actually does not exist) is underestimated in functional imaging studies, resulting in more chance findings than anticipated (Eklund, Nichols, & Knutsson, 2016). This observation most probably generalizes to structural imaging to the degree that similar techniques for multiple comparison correction are used.

Problems associated with publication bias, file drawer mechanisms and underestimated type I error rates are not restricted to personality neuroscience. It has recently been shown that only around one-third of effects reported in psychology journals can be replicated in independent data sets (Open Science Collaboration, 2015). This finding highlights the need for study designs that employ large representative samples and that cross-validate their results. Chances of observing the same chance finding in two (semi-)independent data sets is negligible. We do not know of any studies into the neuroanatomical correlates of personality that have cross-validated their results.

Here we present a study examining structural personality – brain associations. Both Gray Matter (GM) and White Matter (WM) were assessed by employing Voxel-Based Morphometry (VBM) and Tract-Based Spatial Statistics (TBSS) respectively. We used a large young adult sample (N = 922), stratified on gender and educational level (Centraal Bureau voor de Statistiek, 2009) to be representative for the Dutch general population. The sample size enabled us to divide the data into exploratory and confirmatory subsamples in order to be able to cross-validate our results. We expected, based on previous studies, to find a multitude of personality – brain associations in the exploratory sample, of which many would not survive cross-validation. To ascertain that any lack of replication could not be attributed to type II errors (i.e. not finding an association while there actually is one), we also searched for brain associations with gender, for which more consistent results are

reported in the literature (Chou, Cheng, Chen, Lin, & Chu, 2011; Menzler et al., 2011; Ruigrok et al., 2014).

For measurement of the full domain of personality we used the Multidimensional Personality Questionnaire (MPQ; Tellegen & Waller, 2008), a broadband personality instrument that has been shown to have good reliability, validity and utility across a diversity of samples (see Method section). We selected the MPQ over one of the more commonly used Five-Factor Model (FFM) instruments for two reasons. First, the MPQ departs from the above described realist trait perspective which makes it principally suited for an exploration of the neuroanatomical correlates of personality. The current MPQ is the result of an ongoing inductive-deductive test construction process: the development and refinement of the model and the instrument are guided by both personality theory and empirical evidence (Tellegen & Waller, 2008). Because of its partly theoretical basis, the model underlying the MPQ converges with the influential neurobiological model of personality developed by Depue and colleagues (see Depue, 1996). In contrast, the origin of the FFM lays in the psycholexical tradition, which has been agnostic about the ontological status of personality traits and is therefore largely based on a non-theoretical inductive approach to test construction. Second, the MPQ covers two levels of the personality hierarchy on which considerable consensus exists in the literature (see Figure 6.1; DeYoung, 2006; Saucier, 2008; Tellegen & Waller, 2008). The FFM comprises only one level of the hierarchy. It is noteworthy that, as can also be inferred from Figure 6.1, FFM traits can be described in terms of combinations of MPQ traits (chapter 2; Tellegen & Waller, 2008). Consequently, the results from the current study also have implications for FFM findings.

Method

Participants

The sample employed here consisted of 965 Dutch young adults for whom MPQ and imaging data were available. Data were gathered as part of a larger project that had as overarching objective to relate a broad range of individual difference variables to imaging data. The original goal of the project was to get 1,000 participants, but due to monetary constraints data acquisition was stopped before this goal was attained. The sample was stratified by gender and educational level (Centraal Bureau voor de Statistiek, 2009) to be representative for the Dutch general population. Data of 38 persons were removed due to

irregularities in the imaging data (i.e., incidental findings, technical errors during scanning, incomplete recordings and/or excessive subject movement), and data of another 5 participants were disregarded because their MPQ-BF-NL was invalid (see chapter 2 for criteria). The final sample (N = 922) consisted of 440 (48%) men, and age ranged from 20 to 26 with a mean of 22.9 years (SD = 1.70). 87 participants (9%) were currently enrolled in or did finish primary or basic secondary education (VMBO or MBO level 1/2), 397 participants (43%) were currently enrolled in or did finish more advanced secondary education (HAVO/VWO, MBO level 3/4), and 438 participants (48%) were currently enrolled in or did finish college or university.

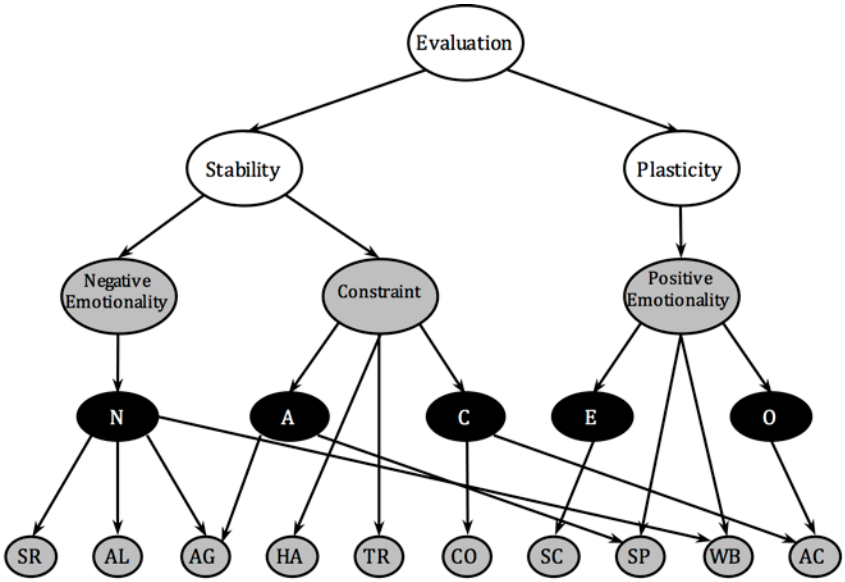

Figure 6.1. The MPQ covers two levels in the hierarchical structure of personality. Gray shapes represent MPQ constructs; black shapes represent FFM constructs. N = FFM Neuroticism; A = FFM Agreeableness; C = FFM Conscientiousness; E = FFM Extraversion; O = FFM Openness; SR = MPQ Stress Reaction; AL = MPQ Alienation; AG = MPQ Aggression; HA = MPQ Harm Avoidance; TR = MPQ Traditionalism; CO = MPQ Control; SC = MPQ Social Closeness; SP = MPQ Social Potency; WB = MPQ Wellbeing; AC = MPQ Achievement.

The sample was randomly split into two equal halves (each n = 461), an exploratory and a confirmatory sample. The two subsamples were matched on age and gender. The exploratory subsample consisted of 221 (48%) men and the confirmatory sample consisted of 219 (48%) men. Descriptives for age were the same in both subsamples as in the full sample. There were no significant differences in mean personality trait scores between the

two subsamples as shown by t-tests, and a Chi-square test showed that educational level also did not differ between the two sample halves.

Personality measured by the Multidimensional Personality Questionnaire

The Dutch brief form of the Multidimensional Personality Questionnaire (MPQ-BF-NL) was used (chapter 2). The 11 MPQ-BF-NL primary trait scales (see lower-order constructs in Figure 6.1) are measured by 12 binary items each. Including three extra items needed for determining scores on validity scales, the full measure consists of 135 items. The primary trait scales coalesce into three higher-order factors: Positive Emotionality (PEM), Negative Emotionality (NEM), and Constraint (CON). PEM is comprised of primary trait scales Wellbeing (WB), Social Potency (SP), Achievement (AC), and Social Closeness (SC). NEM includes Stress Reaction (SR), Aggression (AG) and Alienation (AL), and CON includes Control (CO), Harm Avoidance (HA) and Traditionalism (TR). Absorption (AB) is not allocated to any of the three higher-order factors. Additionally, there are two validity scales, Variable Response Inconsistency (VRIN) and True Response Inconsistency (TRIN) that are comprised of item pairs that are either content matched and inversely keyed (VRIN), or opposite in content but keyed the same direction (TRIN). Consequently, VRIN is a measure of inconsistent responding, while TRIN is a measure of acquiescence or systematic disagreement.

The original U.S. MPQ scales have demonstrated good reliability in a variety of samples and theoretically predicted correlations with other instruments (Tellegen & Waller, 2008). Moreover, the scale scores have been shown to predict behavior (Kamp, 1986), to distinguish between different forms of psychopathology (Miller et al., 2003), and to predict clinical variables better than most other personality scales (Grucza & Goldberg, 2007). Compared to its full length U.S. counterpart, the MPQ-BF-NL showed similar internal consistencies in the general population, with Cronbach's alpha in the range of .75 to .84 for most scales, and between .70 and .73 for Traditionalism, Harm Avoidance, and Aggression. Moreover, higher-order structure and correlational patterns of the MPQ-BF-NL are quite similar to the U.S. (chapter 2).

Table 6.1 provides reliabilities and further descriptions of the measurement domain for each personality trait scale in the sample employed in the current study. Normalized T

Table 6.1. Descriptions of high scorers on scales and Cronbach's alphas as well as means and standard deviations

			M (SD)		
Scale	Description of a high scorer	α^a	Total	Males	Females
WB	Has a happy, cheerful disposition; feels good about self and sees a bright future	.82	50.6 (10.3)	50.9 (10.5)	50.3 (10.1)
SP	Is forceful and decisive; fond of influencing others; fond of leadership roles	.85	51.5 (9.9)*	53.1 (10.0)	50.0 (9.6)
AC	Works hard; enjoys demanding projects and working long hours	.78	50.5 (10.2)	50.9 (9.8)	50.2 (10.5)
SC	Is sociable, likes people, and turns to others for comfort	.81	55.7 (9.6)*	55.2 (9.4)	56.1 (9.8)
SR	Is nervous, vulnerable, sensitive, prone to worry	.84	50.5 (10.0)	48.1 (9.5)	52.7 (9.9)
AG	Hurts others for own advantage; will frighten and cause discomfort for others	.76	54.5 (10.0)*	57.0 (9.9)	52.2 (9.5)
AL	Feels mistreated, victimized, betrayed, and the target of false rumors	.83	51.7 (9.3)	52.0 (9.2)	51.4 (9.4)
CO	Is reflective, cautious, careful, rational, planful	.83	45.0 (11.1)*	43.6 (10.6)	46.2 (11.5)
HA	Avoids excitement and danger; prefers safe activities even if they are tedious	.75	40.4 (7.6)*	37.9 (7.1)	42.7 (7.4)
TR	Desires a conservative social environment; endorses high moral standards	.67	42.9 (9.1)*	42.4 (9.1)	43.4 (9.1)
AB	Is responsive to evocative sights and sounds; readily captured by entrancing stimuli	.79	51.3 (10.2)*	52.2 (10.2)	50.5 (10.2)
PEM	Efficacious; ready to experience positive emotions; actively involved in social/work environments	-	52.7 (10.3)*	53.8 (10.7)	51.8 (9.8)
NEM	Experiences elevated levels of negative emotions such as fear, anxiety, and anger; antagonistic	-	52.1 (10.0)*	51.8 (10.0)	52.4 (10.1)
CON	Endorses social norms; acts in a cautious and restrained manner; avoids thrills	-	40.7 (9.6)*	38.2 (8.6)	43.0 (9.9)

Note. Underlined values indicate that males and females differ significantly at $\alpha = .05$, false discovery rate (fdr) corrected; WB = Wellbeing; SP = Social Potency; AC = Achievement; SC = Social Closeness; SR = Stress Reaction; AG = Aggression; AL = Alienation; CO = Control; HA = Harm Avoidance; TR = Traditionalism; AB = Absorption; PEM = Positive Emotionality; NEM = Negative Emotionality; CON = Constraint.

[a] Cronbach's alpha is not available for higher-order scales since these are calculated using a regression strategy instead of simple summation.

* fdr corrected $p < .001$ (different from general population mean).

scores (*M* = 50; *SD* = 10) were used, benchmarked on the distribution of scores in the Dutch general population (see chapter 2). Reliabilities of the MPQ-BF-NL were similar to the ones observed in the general population (chapter 2). All Cronbach's alphas exceeded .75, except for the one for Traditionalism, which was .67. In comparison to the general population mean scale scores did differ notably for some of the scales (especially mean scores for Social Closeness were higher and mean scores for the Constraint scales were lower). These differences may be explained by the relatively young age of our sample. In the employed sample males scored on average significantly higher on Social Potency, Aggression, Absorption and Positive Emotionality than females. Females showed greater mean levels of Stress Reaction, Control, Harm Avoidance and Constraint. These patterns in gender differences are largely in line with what is consistently found by others (Roberts, Caspi, & Moffitt, 2001), although one would also expect females to generally score higher on Social Closeness, and not to find a difference on Absorption.

Analytic strategy

Gray Matter (GM) was assessed through Voxel-Based Morphometry (VBM) using T1 scans. White Matter (WM) was assessed through Tract-Based Spatial Statistics (TBSS) using Diffusion-Weighted Imaging (DWI) scans. For each of the primary trait scales of the MPQ-BF-NL and for gender ROIs were selected that were associated with the variable (i.e. scale or gender) of interest. Associations were then tested for robustness in the confirmatory sample.

Image acquisition

Magnetic resonance images were acquired using a 3-T scanner (Philips). The participant's head was immobilized using foam pads to reduce motion artifacts and earplugs were used to moderate scanner noise. Three T1 weighted anatomical recordings were acquired. [3DTI turbo field echo; Echo Time (TE), 4.6 ms; Repetition Time (TR), 8.3 ms; Flip Angle (FA) 8°; 160 transversal slices with a thickness of 1 mm, an in plane resolution of $1mm^2$ and a Field Of View (FOV) of 256^2 mm]. In the same session three diffusion-weighted recordings were acquired [Spin Echo-Echo Planer Imaging; TE 73.3 ms; TR 6313 ms; FA 90°, 60 transversal slices with a thickness of 2 mm, an in plane resolution of $2mm^2$ and a FOV of 224^2 mm; b0 = 1000 s/mm^2]. Also in the same session a BOLD-MRI recording was made, which was not used in the current analyses.

Preprocessing

VBM (GM). Data were preprocessed with FSL-VBM (Smith et al., 2004). First, structural images were brain extracted (Smith, 2002). Next, tissue type segmentation was performed using FAST4 (Zhang, Brady, & Smith, 2001). The resulting GM partial volume images were then aligned to MNI 152 standard space, using the affine registration. The resulting images were averaged to create a study-specific template, to which the native GM images were then nonlinearly reregistered with a method that uses a b-spline representation of the registration warp field (Andersson, Jenkinson, & Smith, 2007b; Rueckert et al., 1999). In order to correct for local expansion or contraction, the registered partial volume images were then modulated. The modulated segmented images were smoothed with an isotropic Gaussian kernel with a σ of 4 mm.

TBSS (WM). Data were preprocessed with FSL-TBSS (Smith et al., 2006). First eddy current correction was applied to the raw diffusion data. Next Fractional Anisotropy (FA) images were created by fitting a tensor model to the motion corrected diffusion data using FMRIB's Diffusion Toolbox (FDT), and then brain extracted using Brain Extraction Tool (BET; Smith, 2002). All participants' FA data were then aligned into a common space using FMRIB's nonlinear image registration tool (FNIRT; Andersson et al., 2007b; Andersson, Jenkinson, & Smith, 2007a), which uses a b-spline representation of the registration warp field (Rueckert et al., 1999). Next, the mean FA image was created and thinned to create a mean FA skeleton that represents the centers of all tracts common to the group. Each participant's aligned FA data was then projected onto this skeleton.

ROIs. For the selection of ROIs that were associated in the exploratory sample with each of the MPQ-BF-NL scales (primary trait and higher-order), and with gender, we employed massive voxelwise permutation-based nonparametric testing on both GM and WM data. Brain size, age and gender were included as covariates. This yielded, for each variable, a brain-map with significance values for the association per voxel for both GM and WM data. To determine how many voxels needed to be connected, corrected for multiple comparisons, given a pre-specified type I error rate, we used a Monte-Carlo simulation procedure (AlphaSim; Ward, 2000). Although we are aware that procedures like these likely suffer from inflated type I error rates (Eklund et al., 2016), this generally is the approach taken in similar studies. Note that in the present study potential type I errors are less problematic because the associations that were observed in the exploratory sample were tested in the confirmatory sample.

For the associations between the personality scales and GM a threshold of $p<0.01$ was used. Emerging clusters were selected if they consisted of more than 712 voxels using AlphaSim. For the associations between the personality scales and WM a threshold of $p<0.05$ was used and the subsequently emerging clusters where selected if they consisted of more than 92 voxels using AlphaSim. Thresholds were stricter for GM than for WM because VBM data is smoother than TBSS data, resulting in more and larger - and less informative - clusters for a given threshold. For the associations between gender and brain measures we chose the same minimum number of voxels, but more conservative thresholds (GM: $p<0.001$; WM: $p<0.01$), because otherwise the resulting clusters would have been too large to interpret. For the determination of correlations in the confirmatory sample, residual trait scores from a regression model with brain size, age, and gender as predictors were used in order to correct for influence of theses variables.

Results

Gray matter volume correlates

In the exploratory sample 21 ROIs were identified in which GM volume correlated with specific MPQ-BF-NL trait scales. Table 6.2 describes the ROIs and lists the correlations between the relevant trait-scale and mean GM volume in this area for both the exploratory and the confirmatory samples. Within the exploratory sample, all correlations ranged from |.12| to |.26| and were significant. However, none of these correlations replicated: in the confirmatory sample all correlations reduced to around zero and lost their significance.

In contrast, replicable GM volume correlations were found for gender in ten ROIs. One of these ROIs encapsulated the region between the two hemispheres and the cerebral spinal fluid. This ROI clearly represented systematic registration and brain extraction differences between large and small brains (males on average have larger brains than females), and we decided to disregard it. The remaining nine ROIs are described in Table 6.2. Correlations ranged from |.15| to |.27| in the exploratory sample and from |.09| to |.20| in the confirmatory sample. Males showed on average relatively (i.e. corrected for total brain size) more GM within the limbic system, including the basal ganglia and in the anterior part of the temporal pole. Females showed on average relatively more GM in the ventral medial prefrontal cortex, the anterior and posterior cingulate gyrus and around the central gyrus. These differences are in concordance with the literature (Ruigrok et al., 2014), except that we did not observe a clear pattern of lateralization.

Table 6.2. Associations between mean gray matter volume in ROIs and personality scale scores as well as gender observed in the exploratory sample (n = 461) and their replicability in the confirmatory sample (n =461)

					Gender/age/brain size corrected r[a]	
Scale	Lat.	ROI	MNI coordinates	mm^3 (# voxels)	Exp.	Conf.
WB						
	left	middle and inferior temporal gyrus, posterior division	-61, -20, -18	6288 (786)	-.20***	.02
	right	lateral occipital cortex, superior division; precuneous cortex	26, -79, 33	9400 (1175)	-.23***	.04
	right	frontal pole; superior frontal gyrus	19, 27, 54	9440 (1180)	-.20***	.00
SP						
	left	insular cortex; caudate; putamen; pallidum	-24, 2, 3	7400 (925)	.20***	.03
AC						
	left	cerebellum	-37, -67, -50	9856 (1232)	.17***	-.03
	right	lateral occipital cortex, superior division	32, -74, 30	6888 (861)	-.18***	.00
	left	postcentral gyrus; superior parietal lobule; supramarginal gyrus, anterior division; lateral occipital cortex, superior division	-33, -45, 45	7800 (975)	-.20***	-.01
SC						
	left	frontal pole	-40, 47, 10	6432 (804)	-.19***	-.02
AL						
	left/right	subcallosal cortex	0, 6, -16	8632 (1079)	.20***	.06
	left/right	right cingulate gyrus, posterior division; brain-stem; cerebellum	-2, -40, -13	7008 (876)	.20***	.07
	right	thalamus; caudate	14, -5, 15	5992 (749)	.19***	.12
TR						
	left	superior temporal gyrus, posterior division; supramarginal gyrus, posterior division; angular gyrus; lateral occipital cortex, superior division; planum temporale	-56, -45, 21	14704 (1838)	.21***	.04
AB						
	right	frontal pole, paracingulate gyrus	14, 54, 16	10656 (1332)	-.22***	.00
	left/right	occipital pole	-7, -96, 12	10368 (1296)	-.19***	.05
PEM						
	left	superior, middle and inferior temporal gyrus, posterior division	-65, -31, -9	6176 (772)	-.20***	.03
	left	frontal pole	-35, 46, 20	11880 (1485)	-.26***	.08
	right	lateral occipital cortex, superior division	33, -78, 28	7440 (930)	-.19***	.02
	right	frontal pole; superior frontal gyrus	18, 34, 50	12088 (1511)	-.21***	.04

Table 6.2. (continued)

Scale	Lat.	ROI	MNI coordinates	mm^3 (# voxels)	Gender/age/brain size corrected r[a] Exp.	Conf.
CON						
	left	supramarginal gyrus, posterior division; angular gyrus	-58, -53, 29	7240 (905)	.09	.04
	left	lateral occipital cortex, superior division	-25, -72, 47	6304 (788)	.12*	.04
	right	lateral occipital cortex, superior division	34, -68, 45	6136 (767)	.12*	.08
Gender[b]						
	left/right	frontal pole; left middle frontal gyrus; left paracingulate gyrus; cingulate gyrus, anterior division	-2, 36, 9	54328 (6791)	.27***	.20***
	left	precentral gyrus; postcentral gyrus	-36, -27, 48	6656 (832)	.16***	.14***
	right	middle frontal gyrus; postcentral gyrus; central opercular cortex	42, -10, 33	16832 (2104)	.23***	.16***
	left/right	cingulate gyrus, posterior division	-1, -44, 30	9536 (1192)	.15***	.09**
	left	postcentral gyrus	-36, -27, 48	6656 (832)	.16***	.14***
	right	temporal pole; parahippocampal gyrus, anterior division; temporal fusiform cortex, anterior division	24, 0, -40	10016 (1252)	-.18***	-.20***
	left	temporal pole; frontal orbital cortex; parahippocampal gyrus, anterior division;	-29, 10, -33	15256 (1907)	-.21***	-.18***
	right	putamen; pallidum	24, 0, 0	8168 (1021)	-.19***	-.11**
	left	caudate; pallidum	-21, 3, 6	7848 (981)	-.19***	-.08*

Note. Lat. = Lateralization; ROI = Region of Interest; Exp. Exploratory sample; Conf. = Confirmatory sample. WB = Wellbeing; SP = Social Potency; AC = Achievement; SC = Social Closeness; SR = Stress Reaction; AG = Aggression; AL = Alienation; CO = Control; HA = Harm Avoidance; TR = Traditionalism; AB = Absorption; PEM = Positive Emotionality; NEM = Negative Emotionality; CON = Constraint.

[a]Pearson correlations are reported for all associations except for gender for which Kendall's taus are given. Also, for gender data were not corrected for gender. [b]Positive associations indicate relatively larger (corrected for total brain size) GM volume in females.

*false discovery rate (fdr) corrected $p < .05$. **fdr corrected $p < .01$. ***fdr corrected $p < .001$.

White matter volume correlates

In the exploratory sample 41 ROIs were identified in which WM correlated with specific MPQ-BF-NL trait scales. Table 6.3 describes the ROIs and lists the correlations between the relevant trait scale and mean WM in this area for both the exploratory and the confirmatory samples. Correlations ranged from |.08| to |.19| in the exploratory sample, but did not replicate as they reduced to non-significant values around zero in the confirmatory sample.

For gender in contrast, replicable correlations were observed. Six of the seven observed correlations (ranging from |.14| to | .19|) in the exploratory sample were significant in the confirmatory sample as well (ranging from |.11| to |.15|). Males showed bilaterally more coherent WM tracts in the superior longitudinal fasciculus and in the anterior thalamic radiation. Females showed bilaterally more coherent WM tracts in the inferior fronto-occipital fasciculus. These results are partly inconsistent with previous research that reported more coherent WM tracts throughout the brain in males than in females (Chou et al., 2011; Menzler et al., 2011).

Discussion

In the present study we did not find any robust associations between personality trait scores and brain structure, which strongly suggests that the inconsistency and heterogeneity across previous studies is attributable to chance findings. We deem our results reliable because we used a very large representative (young adult) sample and we searched for associations in an exploratory subsample and tested the results in a confirmatory subsample. As expected, a multitude of structural personality-brain associations appeared in the exploratory subsample of which none replicated in the confirmatory sample. The absence of robust associations in the present study is unlikely to be due to an undesirable high type II error rate because nearly all associations that were found for gender in the exploratory sample were confirmed in our confirmatory sample.

Our results contradict the notion that individual differences in specific brain structures are a direct cause of individual differences in personality trait standings. Direct causation would imply the existence of an identity relation between brain and personality (Lewis, 1966). The notion of an identity relation between brain structure and personality now seems untenable, as robust direct associations are a necessary requirement for such a relation. Moreover, in the field of neuroscience it is becoming more and more accepted that brain

Table 6.3. Associations between mean white matter volume in ROIs and personality scale scores as well as gender observed in the exploratory sample (*n* = 461) and their replicability in the confirmatory sample (*n* = 461)

Scale	ROI	MNI coordinates	mm[3b]	Gender/age/brain size corrected *r*[a] Exp.	Conf.
Wellbeing					
	Anterior thalamic radiation	-3, -9, 4	598	.17**	-.02
	Cortico spinal tract	-6, -22, -16	301	.13**	-.07
	Inferior fronto-occipital fasciculus	-9, -11, -6	623	.17**	-.05
	Inferior longitudinal fasciculus	-28, -17, -15	446	.18**	-.02
	Uncinate fasciculus	-8, 4, -10	182	.13**	-.02
Social Potency					
	Anterior thalamic radiation	-19, 15, 7	231	-.14**	.08
Social Closeness					
	Cortico spinal tract	25, -20, 18	182	-.16**	.05
	Inferior fronto-occipital fasciculus	31,-22,-3	252	-.16**	.01
	Inferior longitudinal fasciculus	43, -19, -15	142	-.14**	-.09
Stress Reaction					
	Anterior thalamic radiation	-1, -9, 3	476	-.15**	-.03
	Cingulum hippocampus	13, -26, -19	101	-.11*	.00
	Cortico spinal tract	-1, -21, -19	284	-.13**	.07
	Inferior fronto-occipital fasciculus	-26, -9, -5	149	-.17**	.06
	Inferior longitudinal fasciculus	-22, -16, -16	358	-.18**	.04
Aggression					
	Anterior thalamic radiation	-7, 14, 6	608	-.15**	.01
	Cingulum cingulate gyrus	-4, -1, 32	321	-.15**	.06
	Cortico spinal tract	-1, -22, 7	1015	-.16**	-.02
	Forceps major	7, -68, 10	195	-.13**	-.09
	Inferior fronto-occipital fasciculus	0, -12, -1	1484	-.16**	-.07
	Inferior longitudinal fasciculus	-16, -37, -9	728	-.14**	-.07
	Superior longitudinal fasciculus	-7, -27, 30	1044	-.12*	-.08
	Superior temporal longitudinal fasciculus	-22, -26, 29	241	-.08	-.05
	Uncinate fasciculus	-9, 4, -10	193	-.09	-.08
Alienation					
	Cortico spinal tract	14, -20, -15	164	-.17**	.07
	Inferior fronto-occipital fasciculus	4, -69, 1	112	-.15**	-.08
	Inferior longitudinal fasciculus	-36, -26, -13	192	-.19**	-.03
Control					
	Anterior thalamic radiation	7, 7, 7	147	.14**	.01
	Cortico spinal tract	9, -22, -15	306	.14**	-.01
	Inferior fronto-occipital fasciculus	28, -22, -4	454	.16**	-.01
	Inferior longitudinal fasciculus	40, -35, -9	229	.15**	.04
	Superior longitudinal fasciculus	36, -26, 31	378	.15**	-.04
	Uncinate fasciculus	-2, 5, -9	117	.09	.05
Negative Emotionality					
	Anterior thalamic radiation	-7, 0, 5	762	-.16**	-.01
	Cingulum cingulate gyrus	-4, -7, 34	268	-.11*	.07
	Cortico spinal tract	-3, -21, -3	928	-.15**	.06
	Forceps major	6. -61, 11	280	-.13**	-.03
	Inferior fronto-occipital fasciculus	-1, -26, -2	1471	-.17**	-.06
	Inferior longitudinal fasciculus	-12, -33, -10	1066	-.16**	-.05
	Superior longitudinal fasciculus	-6, -21, 29	1190	-.11*	-.05
	Superior temporal longitudinal fasciculus	-22, -23, 29	278	-.09	-.06
	Uncinate fasciculus	-18, 4, -12	160	-.11*	-.01

Table 6.3. (continued)

Scale	ROI	MNI coordinates	mm^{3b}	Gender/age/brain size corrected r^a Exp.	Conf.
Gender					
	Anterior thalamic radiation	-12, 17, 9	110	.14***	.05
	Anterior thalamic radiation	0, 2, 6	317	-.19***	-.12***
	Cingulum cingulate gyrus	-5, 2, 32	192	-.16***	-.15***
	Cortico spinal tract	3, -23, 27	330	.19***	.11***
	Inferior fronto-occipital fasciculus	-14, 0, -3	268	.17***	.15***
	Superior longitudinal fasciculus	-14, -23, 27	521	-.17***	-.15***
	Superior temporal longitudinal fasciculus	-32, -27, 28	199	-.15***	-.14***

Note. ROI = Region of Interest; Exp. = Exploratory sample; Conf. = Confirmatory sample.
[a]Pearson correlations are reported for all associations except for gender for which Kendall's taus are given. Also, for gender data were not corrected for gender. [b]mm^3 = # of voxels. [c]Positive associations indicate relatively more WM (corrected for total brain size) in females.
*false discovery rate (fdr) corrected $p < .05$. **fdr corrected $p < .01$. ***fdr corrected $p < .001$.

areas do not operate in isolation, but are organized in complex interacting networks (Smith et al., 2009).

Rather than a direct consequence of brain structure, we deem it more plausible that personality is an emergent phenomenon. Emergence entails the view that basic phenomena interact in such a way as to result in complex phenomena that exhibit properties the more basic phenomena do not possess (Kim, 1999). Applied to personality this would mean that the different determinants of personality, such as neurotransmitter systems, brain structure, brain function and different experiences, engage in complex interactions that combine and transform the properties into specific responses, that result in scale scores on a personality questionnaire. The scores in some sense do comprise their determinants, but it would be a major oversimplification to expect direct links between constituents of one of these systems with the personality trait scores.

More likely, the measurement of the complex systems that lie at the basis of personality should be sought in brain function, instead of brain structure. Much in line with Tellegen's theorizing about the 'if S, then R' propositional nature of personality traits (Tellegen, 1985), the expression of the system may be observable in functional responses to relevant stimuli (e.g. fear responses to threatening stimuli) rather than in structural variation. Also, the measurement of these systems should incorporate the complex and interactional nature of the networks. Dynamic causal modeling makes such investigations now feasible (Friston, Harrison, & Penny, 2003). An example of a model that may operationalize a system that lies at the basis of MPQ Negative Emotionality (or its FFM counterpart Neuroticism) is the

psychobiological system proposed by Depue and colleagues (see Depue, 1996) that comprises different transmitter systems influencing the modulatory function of the Locus Ceruleus (LC) in directing orientation and attention, cascading down to the functioning of many different brain areas. This system is thought to serve a warning-alarm function and to ultimately cause individual differences in the sensitivity to experience fear and anxiety and the like.

Conclusion

Not a single structural personality-brain association replicated, while several robust gender-brain associations emerged. We argue that personality is better conceptualized as an emergent construct resulting from complex interactions between more basic processes. The realist trait perspective may well hold, but the search for the psychobiological substrates of personality should be redirected to functional dynamic causal brain models.