SUPPLEMENTARY MATERIAL:

Joint factorial structure of psychopathology and personality

Tom Rosenström[1,2,*], Line Cecilie Gjerde[1,3], Robert F. Krueger[4], Steven H. Aggen[5], Nikolai Olavi Czajkowski[1,3], Nathan A. Gillespie[5], Kenneth S. Kendler[5,6,7], Ted Reichborn-Kjennerud[1,8], Fartein Ask Torvik[1,3,†], Eivind Ystrom[1,3,8,†]

[1]Department of Mental Disorders, Norwegian Institute of Public Health, Oslo, Norway;

[2]Department of Psychology and Logopedics, University of Helsinki, Finland;

[3]Department of Psychology, University of Oslo, Norway;

[4]Department of Psychology, University of Minnesota, USA;

[5]Department of Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA;

[6]Deparment of Human and Molecular Genetics, Virginia Commonwealth University, Richmond, VA, USA;

[7]Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, USA;

[8]Institute of Clinical Medicine, University of Oslo, Norway;

[8]PharmacoEpidemiology and Drug Safety Research Group, School of Pharmacy, University of Oslo, Norway;

[*]Correspondence to tom.rosenstrom@helsinki.fi;

[†]Joint senior authors

Supplementary contents

Analysis and sensitivity analysis of response styles

When we included raw personality-trait sum scores and psychiatric diagnoses to the same exploratory factor analysis with a bi-factor rotation, we observed similar loading patterns as in the main text (Supplementary Table S1) but with one more specific factor (Figure S1). From this, we also noted that inclusion of PID-5-NBF compulsivity did not alter the overall structure either. We then examined orthogonal-, correlated-, and bi-factor solutions with response styles variables included, and always found one personality factor out of the four factors that overlapped with response style on BFI questionnaire (Supplementary Table S2). Moreover, the response style variables represented the strongest loadings on the factor, whereas psychiatric disorders did not load on it. This suggested that response style in BFI inventory represented a source of confounding for a psychopathology model. In contrast, response style in PID-5-NBF was related to general psychopathology. However, the removal of the BFI response styles did not have major effects on the bi-factor rotated structure of the three first factors.

How response style variables were computed?

Each of the 44 Big Five Inventory items is a statement directed at aspects of the test taker's normal personality, which he or she responds to using an ordinal-valued, 5-point Likert scale from "strongly disagree" (score 1) to "strongly agree" (score 5). The statements are balanced so that for some statements a high score is associated with a high trait value and for other statements the high score is associated with a low trait value. Then, Extreme Response Style (ERS) is defined as the number of items where the subject endorsed the extreme alternatives (scores 1 and 5), whereas Acquiescent Response Style (ARS) is a sum of weights 1 for "agree", 2 for "strongly agree", and 0 for other

endorsements. Analogous variables were computed for the pathological personality traits, assessed with four-level ordinal response format. All continuous-valued variables (ERS, ARS, and personality traits) were standardized to mean of zero and variance of one.

How the raw averages and correlations looked like?

For comprehensiveness, Figure S2 shows all correlations in the data investigated in the main text. The Table 1 in the main text showed prevalence of the disorders, and although we used standardized personality trait scores, the Supplementary Table S3 complements these information with unstandardized score averages.

Shortly on parallel analysis method for determining factor number

In this section, we briefly explain our use of Parallel Analysis (PA) method for convenience, although all the content is also available in pertinent literature (Horn 1965; Garrido *et al.* 2013; Rosenström *et al.* 2017). The logic of the PA method, as applied here, is the following. Geometrically, the eigenvalues $\lambda_1, ..., \lambda_d$, of a correlation matrix correspond to dilations or contractions of the underlying d-dimensional data cloud to the directions of their corresponding eigenvectors. Because correlation is computed as covariance for standardized variables having variance 1, any dilation ($\lambda_i > 1$ for some i) must imply correlations and be balanced by contractions ($\lambda_j < 1$ for some $j \neq i$). If $\lambda_i = 1$ for all i, the data is uncorrelated. For each $\lambda_i > 1$, there must be a corresponding dimension of linear dependence (correlations). Thus, in an infinitely large sample, modeling *fewer* underlying factors for the data than the number of eigenvalues exceeding 1, would directly imply unmodeled correlations. In a finite sample, however, sampling variance induces some chance correlations even when uncorrelated processes generate the data, and the PA simulation characterizes how much on average for a given

sample size. Thus, in our case, modeling fewer factors than there are (ordered) data-derived eigenvalues exceeding the corresponding PA values, would imply that the associated 'comorbidity model' misses some comorbidity. Modeling more factors would instead imply that sampling variance ('noise') is being interpreted. Thus, only one sensible number of factors exists.

While this is the essential logic, please see Garrido *et al.* (2013) for treatment of ordinal-valued data and Rosenström *et al.* (2017) for assessing sensitivity to within-twin-pair correlations with upper and lower bounds. Briefly, within-twin-pair correlations reduce the amount of independent information in the available observations, thus reducing the effective sample size. But whatever the correlation, the effective sample size will be between the number of unique twin pairs and the number of unique twins in the data. If both the sample sizes deliver the same conclusion in PA, then it holds for intermediate correlations as well. This frequently occurs in samples as large as herein. While it is not immediately obvious how the ordinal-valued data should be handled, extensive simulations of Garrido *et al.* (2013) suggested that one should compute the real-data eigenvalues from a polychoric correlation matrix.

Technical supplement to behavior genetic analysis

The twin ACE model of behavior genetics was used to partition the variance-covariance matrix of estimated factor scores into distinct contributions from additive genetic (A) sources of variance, common/shared environmental (C) sources that tend to make twins similar, and non-shared environmental (E) sources of variance that tend to make twins dissimilar, using *a priori* knowledge that monozygotic twins share 100% of their segregating genes and dizygotic twins on average 50% (Neale & Cardon 1992). In ordinary cross-sectional data one has degrees of freedom for distinguishing only single covariance matrix, but in twin analyses, there are extra degrees of freedom from both dizygotic and monozygotic cross-twin covariances; altogether, a three-fold number of degrees of freedom. A multigroup structural equation model for monozygotic (one group) and

dizygotic (the other group) twins can be identified and estimated, such that respective within- and cross-twin covariance structures are

$$\begin{pmatrix} A+C+E & A+C \\ A+C & A+C+E \end{pmatrix} \text{ and } \begin{pmatrix} A+C+E & \frac{1}{2}A+C \\ \frac{1}{2}A+C & A+C+E \end{pmatrix},$$

where A, C, and E are freely estimable within- and cross-trait variance-covariance matrices for genetic influences, shared environmental influences, and non-shared environmental influences, respectively.

Rotations under multivariate normal and non-normal distributions

From a narrow technical viewpoint, the correlated-factor and the bi-factor rotation are just different 'faces' of the same symmetric three-dimensional object, a latent population distribution. However, the latent distribution of factor models (a multivariate normal distribution) is typically chosen to facilitate computation, not because it necessarily best reflects 'the nature' (Lei & Lomax 2005). It can be shown that all rotations produce equivalent fits for multivariate normal distribution *and only for* that distribution (Hyvärinen *et al.* 2001). If all but one of the latent dimensions have non-normal distributions, a uniquely interpretable rotation exists. For an analogy, if three microphones are recording simultaneous speech of three individuals in a room, only one 'rotation' of the received speech signals retrieves the three original non-mixed speeches (up to permutation of individuals) (Hyvärinen *et al.* 2001). While present *computation* (or those generally seen in psychology and psychiatry) cannot identify such a unique rotation, one that best advances science may nevertheless exist. We and others have put forth practical and logical arguments in favor of the bi-factor rotation, and perhaps future works will increasingly develop technical arguments to support or refute those.

Supplementary Table S1. Exploratory bi-factor model without regressing out the response styles

		Bi-factor rotation			
Data type	**Variable**	***General factor: "Psychopathology"***	***Specific factor #1: "Internalizing"***	***Specific factor #2: "Externalizing"***	***Specific factor #3: "Personality"***
Interview	Alcohol use disorder	**0.506**	-0.052	**0.397**	-0.043
	Substance use disorders	**0.539**	0.082	**0.519**	-0.025
	Major depressive episode	**0.502**	**0.451**	0.041	0.031
	Panick attack	**0.576**	**0.54**	0.087	-0.005
	Agoraphobia	**0.612**	**0.668**	0.026	0.006
	Social phobia	**0.655**	**0.495**	-0.04	-0.101
	Specific phobias	**0.378**	**0.436**	-0.031	0.08
	Generalized anxiety disorder	**0.629**	**0.446**	-0.009	-0.038
	Antisocial personality traits	**0.539**	-0.194	**0.485**	0.002
	Psychotic-like experiences	**0.459**	0.262	0.232	0.012
	Manic experiences	**0.44**	0.228	0.273	0.025
BFI	Extraversion	**-0.412**	-0.012	**0.517**	**0.34**
	Agreeableness	**-0.427**	0.295	0.095	0.124
	Conscientiousness	**-0.573**	0.307	-0.031	**0.337**
	Neuroticism	**0.64**	0.172	-0.306	-0.119
	Openness	0.067	0.054	0.299	0.198
PID-5-NBF	Negative emotionality	**0.72**	0.000	**-0.416**	0.221
	Detachment	**0.563**	-0.136	**-0.406**	0.052
	Antagonism	**0.441**	**-0.359**	0.088	**0.38**
	Disinhibition	**0.609**	-0.214	-0.007	0.142
	Compulsivity	**0.424**	0.036	-0.253	**0.449**
	Psychoticism	**0.637**	-0.079	-0.144	**0.372**

Note: Loadings above √0.1 are highlighted; "BFI" = Big Five Inventory; "PID-5-NBF" = Personality Inventory for DSM-5, Norwegian Brief Form

Supplementary Table S2. Various exploratory factor analysis rotations with response style (RS) variables included (a large table on the next page).

		Varimax				Promax							Bi-factor
Data type	**Variable**	**F1**	**F2/RS**	**F3**	**F4**	**F1**	**F2/RS**	**F3**	**F4**	**int**	**p**	**ext**	**RS**
Interview	Alcohol use disorder	**0.572**	-0.05	0.247	0.13	**0.601**	-0.093	0.121	-0.058	0.061	**0.471**	**0.426**	-0.074
	Substance use disorders	**0.636**	0.057	**0.369**	0.064	**0.667**	0.005	0.263	-0.154	0.179	**0.49**	**0.515**	0.022
	Major depressive episode	0.14	0.031	**0.637**	0.151	0.071	0.077	**0.636**	0.031	**0.505**	**0.432**	0.032	0.053
	Panick attack	0.182	0.02	**0.759**	0.128	0.11	0.061	**0.766**	-0.032	**0.615**	**0.482**	0.077	0.039
	Agoraphobia	0.116	0.036	**0.88**	0.136	0.022	0.095	**0.911**	-0.023	**0.739**	**0.504**	0.014	0.064
	Social phobia	0.125	-0.079	**0.793**	0.201	0.035	-0.019	**0.794**	0.044	**0.638**	**0.533**	-0.012	-0.046
	Specific phobias	0.044	0.014	**0.535**	0.126	-0.022	0.062	**0.549**	0.043	**0.441**	**0.328**	-0.037	0.038
	Generalized anxiety disorder	0.162	-0.027	**0.73**	0.209	0.08	0.029	**0.717**	0.065	**0.567**	**0.529**	0.019	0.002
	Antisocial personality traits	**0.681**	0.001	0.147	0.195	**0.723**	-0.046	-0.025	0.017	-0.077	**0.531**	**0.491**	-0.026
	Psychotic-like experiences	**0.327**	0.009	**0.459**	0.092	0.308	0.007	**0.416**	-0.062	**0.32**	**0.405**	0.229	0.005
	Manic experiences	**0.365**	-0.001	**0.415**	0.071	**0.359**	-0.015	**0.365**	-0.091	0.278	**0.39**	0.273	-0.012
BFI	Extraversion	0.23	**0.514**	**-0.322**	-0.3	**0.32**	**0.435**	-0.307	-0.258	-0.262	-0.29	**0.37**	**0.449**
	Agreeableness	-0.271	**0.483**	-0.008	-0.302	-0.274	**0.487**	0.139	-0.186	0.138	**-0.402**	-0.073	**0.465**
	Conscientiousness	**-0.348**	**0.447**	-0.157	-0.278	**-0.348**	**0.457**	-0.016	-0.117	0.01	**-0.475**	-0.15	**0.435**
	Neuroticism	0.027	-0.294	**0.527**	0.4	-0.071	-0.207	**0.464**	0.312	**0.355**	**0.555**	-0.196	-0.233
	Openness	0.303	**0.359**	0.05	0.001	**0.326**	**0.341**	0.007	-0.02	-0.03	0.137	0.269	**0.334**
	Extreme response style in BFI	-0.126	**0.887**	0.024	0.008	-0.167	**0.959**	0.084	0.214	0.022	-0.105	-0.096	**0.894**
	Acquiescent response style in BFI	0.14	**0.802**	0.122	0.079	0.114	**0.852**	0.117	0.191	0.033	0.12	0.095	**0.8**
PID-5-NBF	Negative emotionality	-0.023	-0.124	**0.363**	**0.795**	-0.17	0.051	0.203	**0.855**	0.08	**0.766**	**-0.432**	-0.015
	Detachment	-0.049	-0.254	0.199	**0.649**	-0.159	-0.121	0.059	**0.694**	-0.013	**0.584**	**-0.38**	-0.164
	Antagonism	**0.4**	0.016	-0.148	**0.564**	**0.385**	0.068	**-0.382**	**0.586**	**-0.41**	**0.573**	0.067	0.05
	Disinhibition	**0.319**	-0.048	0.137	**0.598**	0.263	0.033	-0.064	**0.584**	-0.139	**0.678**	-0.028	0.004
	Compulsivity	0.014	0.039	0.148	**0.606**	-0.085	0.169	0.012	**0.692**	-0.069	**0.54**	-0.296	0.117
	Psychoticism	0.207	0.024	0.206	**0.711**	0.112	0.151	0.01	**0.75**	-0.091	**0.737**	-0.183	0.101
	Extreme response style in PID5	-0.237	0.195	-0.186	**-0.749**	-0.148	0.078	0.036	**-0.747**	0.122	**-0.787**	0.18	0.116
	Acquiescent response style in PID5	0.067	0.008	0.223	**0.906**	-0.074	0.193	0.008	**1.012**	-0.108	**0.833**	**-0.404**	0.12

Supplementary Table S3. Unstandarized mean scores for the personality traits, plus their standard errors (SD), and numbers of observations they were based on

Variable	$n_{available}$	Mean	SD
Extraversion (range 1–5)	2295	3.50	0.64
Agreeableness (range 1–5)	2293	3.94	0.42
Conscientiousness (range 1–5)	2295	3.87	0.47
Neuroticism (range 1–5)	2295	2.50	0.67
Openness (range 1–5)	2291	3.32	0.55
Negative emotionality (range 1–4)	2294	1.32	0.42
Detachment (range 1–4)	2295	1.54	0.54
Antagonism (range 1–4)	2296	1.23	0.31
Disinhibition (range 1–4)	2296	1.57	0.49
Compulsivity (range 1–4)	2295	1.40	0.52
Psychoticism (range 1–4)	2293	1.27	0.43

Supplementary Figures

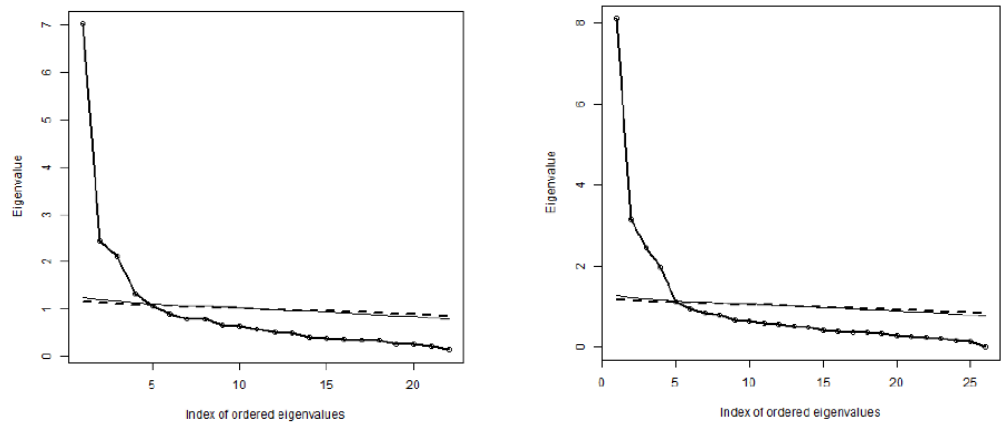

Figure S1. Parallel analysis results without (left) and with (right) response style variables. Number of observed eigenvalues (circles) above parallel-analysis ('zero-correlation') lines indicates the correct number of factors (i.e., 4 in both cases). Note response styles were not regressed out, but left out, in the analysis of left panel. Thus, they were implicitly present.

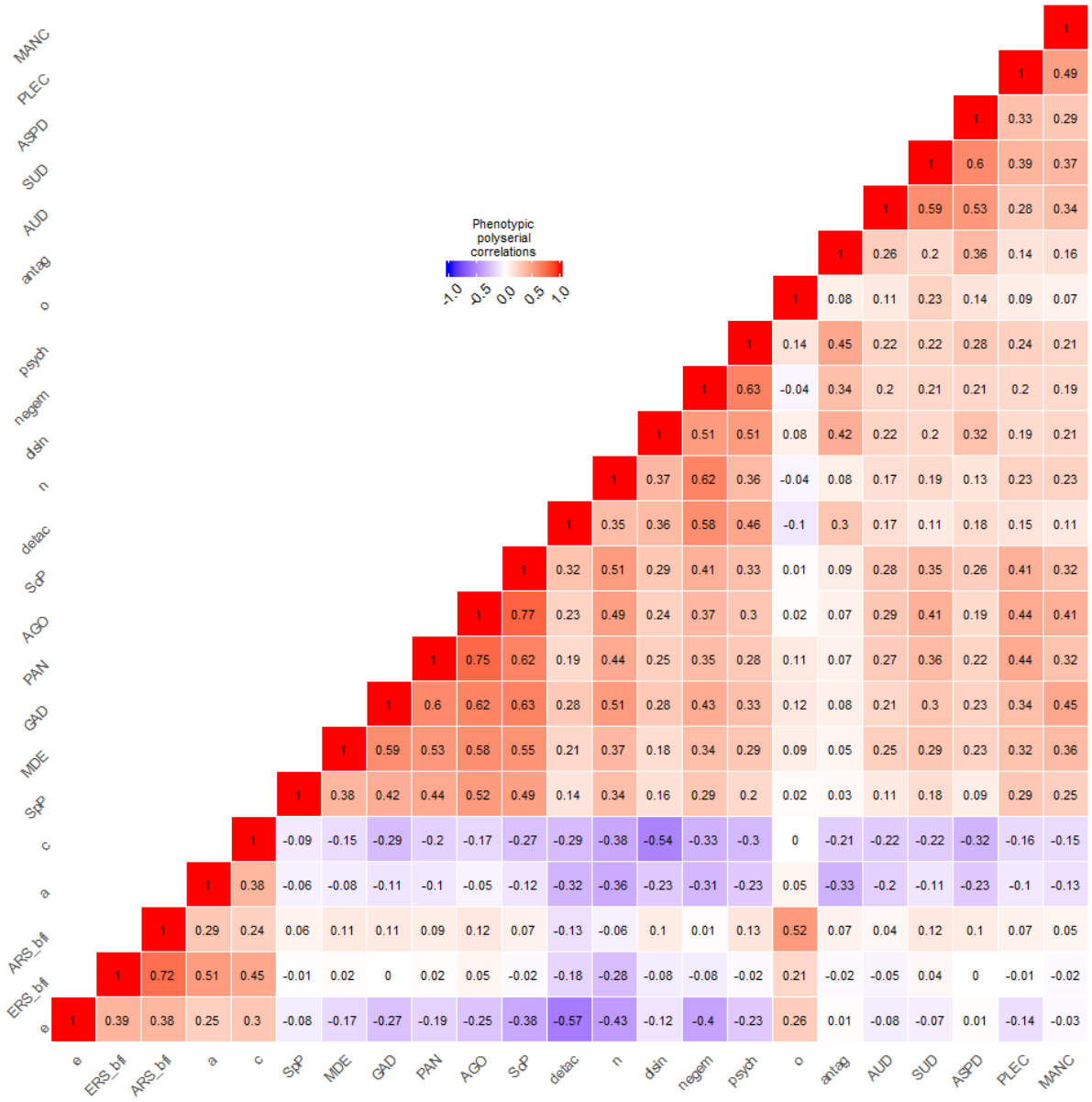

Figure S2. Phenotypic correlations under liability-threshold modelling for ordinal data. Abbreviations: “e” = Extraversion in Big Five Inventory (BFI); “ERS_bfi” = Extreme Response Style in BFI; “ARS_bfi” = Acquiescent Response Style in BFI; “a” = Agreeableness in BFI; “c” = Conscientiousness in BFI; “SpP” = Specific Phobias; “MDE” = Major Depressive Episode; “GAD” = Generalized Anxiety Disorder; “PAN” = Panick attack; “AGO” = Agoraphobia; “SoP” = Social Phobia; “detach” = Detachment in Personality Inventory for DSM-5, Norwegian Brief Form (PID-5-NBF); “n” = Neuroticism in BFI; “disin” = Disinhibition in PID-5-NBF; “negem” = Negative emotionality in PID-5-NBF; “psych” = Psychoticism in PID-5-NBF; “o” = Openness to experience in BFI; “antag” = Antagonism in PID-5-NBF; “AUD” = Alcohol Use Disorder or Dependency; “SUD” = Substance Use Disorder; “ASPD” = Antisocial personality disorder traits; “PLEC” = Psychotic-like experiences, categorical score; “MANC” = Manic experiences, categorical score.