

*This is a pre-print version for scholarly re-distribution. Please, cite the original published paper: Rosenström, T., and Jokela, M. (2017) A Parsimonious Explanation of the Resilient, Undercontrolled, and Overcontrolled Personality Types. Eur. J. Pers., doi: 10.1002/per.2117 (a direct link: <http://dx.doi.org/10.1002/per.2117>)*

## **A parsimonious explanation of the resilient, undercontrolled, and overcontrolled personality types**

Tom Rosenström<sup>1,2\*</sup>

Markus Jokela<sup>1</sup>

<sup>1</sup>Department of Psychology and Logopedics,  
Faculty of Medicine, University of Helsinki, Finland

<sup>2</sup>Department of Mental Disorders,  
Norwegian Institute of Public Health, Oslo, Norway

\*Correspondence:

Tom Rosenström, University of Helsinki,  
Siltavuorenpenger 1A, 00014, Helsinki, Finland.

E-mail: [tom.rosenstrom@helsinki.fi](mailto:tom.rosenstrom@helsinki.fi)

Tel: +358 44 5397665

The work was sponsored by the Emil Aaltonen Foundation.

## Abstract

**Objective:** Researchers applying clustering methods have found that the five most commonly studied personality traits (the “Big Five”) appear to form three prototypes, known as Resilient, Undercontrolled, and Overcontrolled (RUO) personality types. The analysis has been replicated cross-nationally and the results have been reasonably robust. However, these findings do not necessarily imply discontinuities or nonlinearities in the Big Five data. We study whether the RUO types can arise from typical Big Five intercorrelations alone.

**Method:** We used data from a previous meta-analysis of inter-trait correlations ( $N = 144,117$  participants) and simulated Normally distributed observations with this correlation structure. Applying three different clustering algorithms ( $k$ -means, hierarchical agglomerative, and model-based) with three-cluster solutions to the simulated data, we examined whether the known correlations alone can give rise to the RUO typology. The simulated results were compared with previous empirical findings.

**Results:** A simple multivariate normal distribution with the Big Five correlation structure was sufficient to generate the RUO typology in 3 cluster solutions for all the three clustering methods.

**Conclusions:** Contrary to the RUO typology “carving personality description at its joints”, linear correlations typical for correlations among Big Five traits can create RUO types even in the absence of any points of discontinuity.

Keywords: Personality types; Statistical methods; Five factor inventory; Clustering methods

## Introduction

The factorial structure of personality traits has been studied extensively based on continuously distributed traits (Allport, 1937; John, Naumann, & Soto, 2008). The most widely used Five Factor Model consists of higher-order traits of Openness (O), Conscientiousness (C), Extraversion (E), Agreeableness (A), and Neuroticism (N) (Digman, 1990; John et al., 2008; McCrae & Costa, 2003). Another line of research has examined whether people can be clustered empirically into distinct prototypical personality types. Evidence from multiple studies shows that the five personality dimensions appear to form three prototypes, known as Resilient, Undercontrolled, and Overcontrolled (RUO) personality types (Asendorpf, Borkenau, Ostendorf, & Van Aken, 2001; Asendorpf & van Aken, 1999; Robins, John, Caspi, Moffitt, & Stouthamer-Loeber, 1996). A Resilient type person is one with a lower than average Neuroticism, and higher than average value in the four other traits, whereas the Overcontrolled type has the opposite pattern. The Undercontrolled personality type is most often observed to have higher than average Openness, Extraversion, and Neuroticism, and lower than average Conscientiousness and Agreeableness. The prototypes have been replicated in several different linguistic and cultural traditions (e.g., Alessandri et al., 2014; Klimstra, Hale, Raaijmakers, Branje, & Meeus, 2010; Meeus, Van de Schoot, Klimstra, & Branje, 2011; Pesonen et al., 2008; Robins et al., 1996; Specht, Luhmann, & Geiser, 2014) and thereby appear reasonably robust (but see Asendorpf, 2002, for a discussion on failed replications).

Despite all efforts towards a personality system of uncorrelated traits, the actual measures do show nontrivial correlations between traits (DeYoung, Quilty, & Peterson, 2007; van der Linden, te Nijenhuis, & Bakker, 2010). At first sight, moderate trait inter-correlations seem unrelated to the question of prototype clusters, and with model-based (finite mixture model) clustering methods and correct number of estimated cluster this is indeed the case (Fraley & Raftery, 2002; MacKay, 2003; McLahlan & Peel, 2000). However, the clustering methods used in research on personality

typologies have often relied on older clustering algorithms (e.g., Alessandri et al., 2014; Asendorpf et al., 2001). A known problem with methods such as *k*-means clustering is that they lack representation for the size and shape of clusters, and mere correlations between the measures can therefore give rise to spurious clusters (see e.g. chapters 20 and 22 in MacKay, 2003). Furthermore, even model-based clustering methods can give rise to replicable spurious clusters when a wrong number of clusters is specified, and researchers have argued for the use of models with cluster numbers that do not provide strictly the best fit to data in terms of the applied information criteria (e.g., Alessandri et al., 2014).

Figures 1 and 2 demonstrate the issue in a two-dimensional case. The upper-left panels give an example of simulated uncorrelated data (spherical; Figure 1) and correlated data (Figure 2). The other panels show fifty 2-cluster solutions from three different widely-used clustering methods. When one is estimating two cluster centers and the ‘true’ underlying data has only one spherical cluster, “replicable and robust” results are unattainable (see Figure 1). However, when the underlying data has only one *correlated* cluster of observations, the two cluster centers arising from the algorithms are highly “replicable and robust” (Figure 2). A crucial difference between the newer model-based and the older *k*-means and Ward’s algorithms is that the older methods are based on spherical cluster representations and favor 2 clusters in this example when comparing models with different numbers of clusters; the newer method “models” elliptic clusters and therefore correctly favors a single elliptic cluster when comparing models with different numbers of cluster (Figure 3; see Appendix for explanation of model-selection criterion).

In exploratory cluster analysis, choosing an appropriate number of clusters is a very difficult task that depends on specifying the correct model (Alessandri et al., 2014; Asendorpf et al., 2001; Fraley & Raftery, 2002; MacKay, 2003). We seek to answer a more restricted question: do the observed inter-correlations of the Big Five traits suffice to give rise to RUO topology in three cluster solutions? Given the properties of the clustering methods used to infer the RUO personality

types, the mere inter-correlations of the Big Five traits might underlie the personality prototypes, but this is not self-evident and will be tested here. If the RUO typology does not arise from known inter-correlations of traits, then the typology may reflect some non-linear processes that give rise to distinct personality types. If, however, the RUO typology can arise from simple linear correlations, the value of the typology would be undermined, giving further support for continuous models of personality and psychopathology (e.g., Haslam, Holland, & Kuppens, 2012; Kotov et al., 2017; Krueger & Markon, 2014; Markon, Chmielewski, & Miller, 2011; Tyrer, Reed, & Crawford, 2015; Widiger & Mullins-Sweatt, 2009).

The aim of this contribution is not to argue that personality clusters, taxa, or other nonlinearities would not exist, or that clustering methods should never be used. The aim is to argue that clusters can arise from linear correlations in the absence of any discontinuity in the data, unless these correlations are taken in account by use of proper method or removed prior to the analysis (e.g., by ‘sphering’ of data as in Hyvärinen, Karhunen, & Oja, 2001). In many ways, Normal distribution is the simplest, or most parsimonious, explanation of data; for example, it is the maximum entropy distribution (assumes the least structure for a given covariance; Cover & Thomas, 2006, p. 254), result from summed independent influences, and is the standard model for errors experiment (Klenke, 2008). Relying on categorical data is not simply an alternative for linear correlations with equal but different strengths and weaknesses. When underlying taxon does not exist, use of categorical formulations has been criticized on many grounds, time and again (Kotov et al., 2017; Maxwell & Delaney, 1993; Royston, Altman, & Sauerbrei, 2006; van der Sluis, Posthuma, Verhage, & Dolan, 2013).

## **Method**

### *Data*

Correlation estimates between the Big Five dimensions used in this simulation study were drawn

from published data. First, we made use of the estimates from a recent meta-analysis of 212 studies and 144,117 individuals (van der Linden et al., 2010). These represent a comprehensive literature search on all correlations for measures of “Big Five” or “Five Factor Model” personality traits. Second, as a validation and sensitivity analysis, we used correlation matrix from Eugene-Springfield community sample of 481 individuals (DeYoung et al., 2007). DeYoung et al. (2007) recruited the sample by mail from a list of homeowners, and they fulfilled NEO-PI-R questionnaire. Most participants in the sample identified as White, and they spanned age range from 20 to 85 years (mean was 52.5 and s.d. 12.6) and all educational levels. Table 1 summarizes these two correlation matrices.

The previously observed correlation matrices were used to generate 10 000 five-dimensional multivariate normally distributed observations with the given correlation structure and mean vector of zeros. These simulated individuals were studied by *k*-means cluster analysis. Because we simulated the data from a single multivariate Normal distribution, we knew that no true clusters existed in the data.

As an empirical point of comparison, we used average cluster centers from a four-nation study by Alessandri et al. (2014). They used samples of college students from Italy ( $n = 322$ ), the United States ( $n = 499$ ), Spain ( $n = 420$ ), and Poland ( $n = 235$ ), striving to replicate the findings of Asendorpf et al. (2001) cross-nationally.

In the very end, we use a full dataset from the International Personality Item Pool (IPIP) project to compare clustering solutions in correlated and correlation-removed data (Goldberg et al., 2006). The data is freely available ([http://personality-testing.info/\\_rawdata/](http://personality-testing.info/_rawdata/)) and contains information from 19 719 independent individuals’ responses to the 50-item IPIP Big-Five-Factor Markers personality measure (<http://personality-testing.info/tests/IPIP-BFFM/>). The data has been gathered through an online form and then distributed for research purposes, and at the time of writing, was last updated 18<sup>th</sup> of May 2014.

### *Statistical procedure*

All computations were performed in R software, Linux 64-bit version 3.2.2, and a full R script for the analyses is provided as supplementary online material. Multivariate random vectors were generated using the “mvtnorm” package. Having generated the simulated participants' Big Five profiles (five-dimensional random vectors;  $N = 10\,000$ ), we fitted the  $k$ -means clustering algorithm (“kmeans” R-function) with three clusters (i.e.,  $k = 3$ ) and default options. This entire procedure of simulating the data and running the  $k$ -means clustering algorithm was repeated 100 times, and the average cluster centers/means and their 95% confidence intervals were plotted, yielding our main results. In addition to the  $k$ -means algorithm, we repeated the entire procedure using a version of Ward’s hierarchical, or agglomerative, clustering as implemented in the “hclust” function of the “stats” R package (with method “ward.D”). Two closely related but slightly different versions of Ward’s algorithm exist in the literature, and we used the one based on Euclidian distances rather than their squares (Murtagh & Legendre, 2014). We also replicated our analysis using model-based clustering as implemented in the “mclust” R package, version 5.2 (Fraley & Raftery, 2002; McLahlan & Peel, 2000). This method seeks initial values by agglomerative clustering and then fits a mixture of multivariate Gaussian distributions using expectation-maximization algorithm (Dempster, Laird, & Rubin, 1977).

The patterns of cluster centers were compared to the average RUO personality prototypes found by Alessandri et al. (2014). In addition to visual comparison, we used pattern distance based on Euclidian/Frobenius norm; the Euclidian distance between the cluster centers. Differences for distances across conditions were compared using (exact enumeration) permutation tests.

As a minor inconvenience, the clustering algorithm may label the clusters differently for different instances of the 100 repetitions. For consistent results, we re-labeled the clusters in each run so that the first cluster (related to “Resilient” type) had the lowest Neuroticism mean among the

three clusters, the second cluster (related to “Overcontrolled” type) had the lowest Extraversion mean, and the third was the remaining cluster, or a random remaining cluster if both Neuroticism and Extraversion were lowest in the same cluster (a very rare occurrence in samples with Eugene-Springfield correlation structure; never occurred for data simulated with the meta-analytic correlations). Notice that clusters are renamed also in empirical research, often based on their similarity with the RUO types. Thus, our procedure can be thought of as simulating results from 100 identical replication studies.

Usually researchers make several different kinds of sensitivity analyses for determining the optimal number of clusters in cluster analysis, and their replicability across different runs of the algorithm (Alessandri et al., 2014; Asendorpf et al., 2001; Fraley & Raftery, 2002; MacKay, 2003). However, different indicators have often been inconsistent in past research, and yet researchers have tended to argue for three clusters (Alessandri et al., 2014). To keep things simple, we just wanted to test if three-cluster solutions generally replicate the RUO personality structure when applied to a dataset with the Big Five correlation structure but no real clusters. Finally, we also studied the relative importance of trait correlations vs cluster re-labeling in the ensuing  $k$ -means solution by gradually decreasing the trait inter-correlations—at zero correlations, any systematic clustering must be due to the re-labeling.

The simulation studies establish that correlations are *sufficient* to create RUO clusters under certain methods, but not that they *necessarily* create them in raw, real data. Points of discontinuity (or rarity) could exist alongside the correlations. To further assess whether RUO types necessarily arise from trait correlations, we estimated 3-cluster  $k$ -means clustering solutions in the original IPIP data and in whitened, or sphered, IPIP data; 95% bootstrap percentile confidence intervals of cluster means were derived from 2000 bootstrap resamples (Efron & Tibshirani, 1993). Whitening is an invertible linear transformation that removes correlations and standardizes the variables, but does not render multiple clusters into a one cluster. Mathematically, if  $S$  is observed covariance,  $S =$

$UDU'$  its eigendecomposition, and  $X$  the original  $n$ -by-5 data matrix ( $n$  is sample size), then the matrix product  $XUD^{-1}$  yields a whitened, uncorrelated dataset (e.g., Hyvärinen, Karhunen, & Oja, 2001). Clusters that are produced by correlations only should vanish after whitening.

## Results

Figure 4 presents the average  $k$ -means cluster centers for the simulated data with the correlation structure derived from the previous meta-analysis, and compares those with the average results from the cross-national replication of RUO types. It is fair to say that this replicates the RUO personality types quite well. The same held for hierarchical agglomerative clustering (Figure 5) and for model-based clustering using a mixture of three multivariate Gaussian distributions (Figure 6). Also the Eugene-Springfield correlation structure yielded the RUO types (Figure 7). In summary, three-cluster solutions resulted in RUO personality types when the clustering methods were applied to strictly continuously (multivariate normally) distributed simulated data having experimentally observed Big Five inter-correlations.

When the meta-analytic trait correlations (Table 1) were diluted by multiplying them with a coefficient less than one, the cluster means found from simulated data using  $k$ -means algorithm were gradually drawn towards zero, except for the traits that were used to designate the cluster label (Figure 8). This latter effect is a sampling bias caused by the procedure of cluster re-labeling. The RUO pattern is strikingly persistent at the face of diluting trait correlations, which illustrates the importance of the statistical effects of cluster re-labeling. Nevertheless, the distance between simulated and observed cluster means was higher for zero-correlation simulation (resilient = 1.056, undercontrolled = 0.739, and overcontrolled = 1.494) than for the observed-correlation simulation (resilient = 0.483, undercontrolled = 0.830, and overcontrolled = 0.558), suggesting that also the trait correlations are an important determinant of the cluster solutions. Correlations had a statistically significant contribution for Resilient and Overcontrolled clusters ( $p = 0.016$  and  $p =$

0.028, resp.), but not for Undercontrolled cluster ( $p = 0.578$ ).

In the “3-means” analysis of the full IPIP data, Resilient and Overcontrolled types replicated reasonably well over the bootstrap resamples, whereas the Undercontrolled type did not (Figure 9, top row of panels). None of the types were distinguishable after whitening of the data, and only the effects of selection based on Neuroticism and Extraversion (i.e., cluster re-labeling) consistently replicated across the bootstrap resamples (Figure 9, bottom row). This strongly suggests that all clustering in the IPIP data is due to either trait correlations or cluster re-labeling.

### Discussion

The results of this simulation study show that the RUO personality types can result merely from the inter-correlations of the Big Five traits. The findings raise a concern that RUO personality types could be an epiphenomenon of the Big Five correlation structure instead of genuine data-point clustering. Similar concerns may apply to other clusters of personality types reported in the literature (e.g., Hertzberg & Roth, 2006). Even if there were no other structure in the data than a multivariate Normal distribution with the observed inter-correlations of the Big Five traits, three-cluster solutions would suggest the RUO personality prototypes. In this case, the highly-replicable prototypes would be seen as a matter of re-expressing (categorizing) the data by discarding information rather than as a matter of “carving personality description at its joints” (Alessandri et al., 2014; Asendorpf et al., 2001). Note, however, that our analysis does not indicate that categorical prototypes could not exist—only that more care is needed in their analysis to avoid confounding with simpler factorial structures (cf. last paragraph of Introduction).

Since Normal distribution is a maximum entropy distribution for a given mean vector and covariance matrix, and therefore contains less *a priori* structure than alternative distributions (Cover & Thomas, 2006), other things being equal, it is the simplest explanation of Big Five correlations. Therefore, one should be hesitant about alternative explanations of Big Five correlations without

being able to show that they actually represent the data better than the usual Big Five model, or that some non-statistical arguments favor the alternative. Thus far, the proponents of the RUO typology have mainly concentrated on demonstrating the replicability and consistency of the typologies across samples (Alessandri et al., 2014; Asendorpf et al., 2001). We demonstrated here that the types can be a consequence of the replicable and consistent correlations between the Big Five traits and of the standard practice of cluster re-labeling (DeYoung et al., 2007; van der Linden et al., 2010).

A bimodal trait distribution would be a finding that would suggest a typology underlying the trait. Asendorpf et al. (2001) sought for these but concluded that “[t]he expected bimodal pattern for overcontrol and undercontrol when the resilient type was dropped from analysis was found only for the Q-sort data. In all other 22 cases, the distribution was unimodal.” Also more formal taxometric methods have generally failed to find support for any personality typologies, excluding schizotypal personality (Haslam et al., 2012). Psychometricians have called for more rigor in tests of the dimensional (or factorial) structure of personality (Borsboom, 2006), and it is only appropriate to extend that concern to typologies. All the more so because the discussion on typologies *versus* continua is topical for the oncoming revisions to clinical personality-disorder diagnoses, wherein several researchers argue for the removal of diagnoses based on typologies (Kotov, et al., 2017; Krueger & Markon, 2014; Tyrer et al., 2015; Widiger & Mullins-Sweatt, 2009).

Although not directly pertinent to the present argument, it is of interest to note that modern psychometric developments are not restricted to kinds *versus* continua, but can cover ground between these extremes, as well as go beyond that debate (Borsboom et al., 2016; De Boeck, Wilson, & Acton, 2005; Kotov, et al., 2017; Masyn, Henderson, & Greenbaum, 2010). For example, (i) factor mixture models make assumptions that can be seen as intermediate to older cluster (or latent class) models and factor models, (ii) graded membership models assume typologies but allow varying degrees of membership in these, and (iii) network models posit dynamic systems that allow

sudden, typology-like state transitions in some individuals to co-exist with gradual, continuous transitions in others (Borsboom et al., 2016). The Big Five traits were derived from cross-sectional between-individual differences, which may not correspond to personality processes within individuals (Asendorpf, 2002; Cramer et al., 2012; Molenaar & Campbell, 2009). While novel psychometric tools and emerging longitudinal data hold much promise for personality research when properly applied, they do not lift the risk of confounded assumptions, and in isolated applications, may even increase it. Risk of misinterpretation can be minimized (*i*) by understanding the distributional implications of the models, (*ii*) by comparing sufficiently many models so that a range of implied distributions are covered, and (*iii*) by examining absolute fit of the model. An anonymous reviewer also suggested a solution to the cluster re-labeling problem: if a systematic re-labeling method is applied, it can be also repeatedly applied to simulated data without true clusters thereby creating a ‘null’ distribution; then any ‘true’ cluster means are required to differ from that null distribution (an idea reminiscent of Horn’s parallel analysis criterion; Horn, 1965).

Much of discussion on psychometric choices is actually about natural order of things or about etiology, which is probably why Asendorpf et al. (2001) entitled their RUO-typology paper using the phrase “carving personality description at its joints”; an obvious reference to Plato’s classic metaphor of “carving nature at its joints”. For purposes of natural science, “Occam’s razor” is a frequently applied ‘carving tool’. The maxim states that if one can explain a phenomenon without assuming this or that hypothetical entity, then there is no grounds for assuming it (Russell, 1945, pp. 472). Although most would intuitively agree that multiple distributions (correlated or not) represent more entities than a single correlated distribution, rigorous mathematical (Bayesian and information theoretic) justifications exist as well (Cover & Thomas, 2006; MacKay, 2003). In this sense, typologies should be discarded in favor of dimensional representation when exchangeable in terms of fit. This does not deny potentially useful purposes for typologies, such as their cognitive ergonomics when treating high-dimensional phenomena, or their utility in making a decision between

treating *versus* not treating a patient. However, the potential benefits of using typologies should be weighed against potential losses in the form of frustrating the scientific progress (Borsboom et al., 2016; Kotov, et al., 2017) and in terms of statistical bias and loss of power (Maxwell & Delaney, 1993; Royston, et al., 2006; van der Sluis, et al., 2013).

In conclusion, this study demonstrated that a simple multivariate normal distribution with the Big Five correlation structure is sufficient to generate the RUO types when a researcher forces for three clusters and re-labels them according to their resemblance with previous findings, as is commonly done. It was also shown that the clusters replicable in the IPIP open-access data via such means could not be found after applying a linear transformation that removes correlations but no other types of clustering. Formal comparison of alternative model-based clustering solutions would be a principled way to detect genuine normally-distributed clusters despite correlations in the data, but even there a careful power analysis for establishing sufficient sample size is advisable. Sometimes researchers have forced solutions with other number of clusters than that supported by model-selection criteria, and we showed that in such case, even model-based clustering may support replicable but non-existent RUO types. It remains to be shown by explicit model comparison and/or theoretical argumentation if personality types provide any practical advantages over continuous scores as models of natural reality (methods for making a choice between treating and not treating are likely to remain important in clinical work).

### **Declaration of Conflict of Interest**

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### **Acknowledgements**

We thank Dr. Fartein Ask Torvik for valuable feedback, and the Emil Aaltonen Foundation for financial support of the manuscript preparation. The sponsor had no role in writing of the manuscript, design of analyses, or decision to submit the manuscript for publication.

### **Appendix**

Among other methods, Akaike's Information Criterion (AIC) has been used to select for the correct number of clusters in a cluster analysis (Alessandri, et al., 2014). Although selecting number of clusters was not a topic of this study, we demonstrate here using AIC as an example that the model-based and adhoc clustering methods diverge with respect to it, despite providing qualitatively similar results when using the same number of clusters. Alessandri et al. (2014) defined AIC for cluster models according to Burnham & Anderson (2004). They examine re-scaled AIC values

$$\Delta AIC_i = AIC_i - \min_i(AIC_i),$$

that represent the difference between AIC for model with  $i$  clusters and minimum AIC over all the tested cluster numbers. Because smaller AIC values indicate better models, the favored number of clusters according to this method is  $i$  such that  $\Delta AIC_i = 0$ . The heuristic methods with spherically symmetric clusters can define AIC as

$$AIC = n \log(\sigma^2) + 2 K,$$

where  $\sigma^2$  is the mean-Euclidian distance of data points from their nearest estimated cluster center,  $n$  is the number of independent observations, and  $K$  is the number of estimated parameters (in the  $k$ -

means algorithm  $K = k*d$ , where  $d$  is the dimension of the data; e.g., 5 for the Big Five traits). In model-based clustering, AIC is defined as

$$AIC = -2 \log(L) + 2 K,$$

where  $L$  is the value of the model's likelihood function at the maximum-likelihood parameter values.

To illustrate the differences of model-based clustering and the heuristic, spherical-cluster methods ( $k$ -means and Ward's clustering), we simulated one more dataset exactly as in the Figure 2, only with 10 000 observations to minimize sampling variation in AIC estimates. Figure 3 shows AIC for heuristic and model-based methods. As evident, the heuristic methods would suggest two clusters to accommodate the spherical-cluster assumption with the elliptic/correlated data, whereas the model-based clustering method that estimates also the cluster correlations correctly suggests only a single cluster model.

## References

- Alessandri, G., Vecchione, M., Donnellan, B. M., Eisenberg, N., Caprara, G. V., & Ciecuch, J. (2014). On the cross-cultural replicability of the Resilient, Undercontrolled, and Overcontrolled personality types. *Journal of Personality, 82*(4), 340–353. <http://doi.org/10.1111/jopy.12065>
- Allport, G. W. (1937). *Personality: A Psychological Interpretation*. New York, Holt: Rinehart & Winston.
- Asendorpf, J. B. (2002). Editorial: The puzzle of personality types. *European Journal of Personality, 16*, S1-S5. <http://doi.org/10.1002/per.446>
- Asendorpf, J. B., Borkeanu, P., Ostendorf, F., & Van Aken, M. A. G. (2001). Carving personality description at its joints: Confirmation of three replicable personality prototypes for both children and adults. *European Journal of Personality, 15*(3), 169–198. <http://doi.org/10.1002/per.408>
- Asendorpf, J. B., & van Aken, M. A. (1999). Resilient, overcontrolled, and undercontrolled personality prototypes in childhood: Replicability, predictive power, and the trait-type issue. *Journal of Personality and Social Psychology, 77*(4), 815–832. <http://doi.org/10.1037/0022-3514.77.4.815>
- Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika, 71*(3), 425–440. <http://doi.org/10.1007/s11336-006-1447-6>
- Borsboom, D., Rhemtulla, M., Cramer, A. O. J., van der Maas, H. L. J., Scheffer, M., & Dolan, C. V. (2016). Kinds versus continua: a review of psychometric approaches to uncover the structure of psychiatric constructs. *Psychological Medicine, 46*(8):1567–1579. <http://doi.org/10.1017/S0033291715001944>
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research, 33*(2), 261–304. <https://doi.org/10.1177/0049124104268644>

- Cover, T. M., & Thomas, J. A. (2006). *Elements of Information Theory* (2nd ed.). Hoboken, US-NJ: John Wiley & Sons, Inc.
- Cramer, A. O. J., van der Sluis, S., Noordhof, A., Wichers, M., Geschwind, N., Aggen, S. H., Kendler, K. S., Borsboom, D. (2012). Dimensions of normal personality as networks in search of equilibrium: You can't like parties if you don't like people. *European Journal of Personality*, *26*(4), 414–431. <http://doi.org/10.1002/per.1866>
- De Boeck, P. Wilson, M., & Acton, G. S. (2005). A conceptual and psychometric framework for distinguishing categories and dimensions. *Psychological Review*, *112*(1):129–158. <http://doi.org/10.1037/0033-295X.112.1.129>
- Dempster, A. P., Lair, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, *39*(1):1–38.
- DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology*, *93*(5), 880–896. <http://doi.org/10.1037/0022-3514.93.5.880>
- Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, *40*(1), 417–440. <http://doi.org/10.1146/annurev.ps.41.020190.002221>
- Efron, B. & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Boca Raton, US-FL: Chapman & Hall/CRC.
- Fraley, C., & Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, *97*(458), 611–632. <http://doi.org/10.1198/016214502760047131>
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Cough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40*(2006), 84–96. <http://doi.org/10.1016/j.jrp.2005.08.007>
- Haslam, N., Holland, E., & Kuppens, P. (2012). Categories versus dimensions in personality and psychopathology: a quantitative review of taxometric research. *Psychological Medicine*, *42*, 903–920. <http://doi.org/10.1017/S0033291711001966>

- Herzberg, P. Y. & Roth, M. (2006). Beyond resilient, undercontrollers, and overcontrollers? An extension of personality prototype research. *European Journal of Personality, 20*, 5–28. <http://doi.org/10.1002/per.557>
- Horn, J. L. (1965). A rationale and test for the number of factors in Factor Analysis. *Psychometrika, 30*(2), 179–185.
- Hyvärinen, A., Karhunen, J., & Oja, E. (2001). *Independent Component Analysis*. NY, New York: Wiley.
- John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), (3rd ed., pp. 114–158). New York, NY: The Guilford Press.
- Klenke, A. (2008). *Probability Theory: A Comprehensive Course*. London, UK: Springer-Verlag.
- Klimstra, T. A., Hale, W. W., Raaijmakers, Q. A. W., Branje, S. J. T., & Meeus, W. H. J. (2010). A developmental typology of adolescent personality. *European Journal of Personality, 24*(4), 309–323. <http://doi.org/10.1002/per.744>
- Kotov, R., Krueger, R.F., Watson, D., Achenbach, T.M., Althoff, R.R., Bagby, R.M., ... Zimmerman, M. (2017). The hierarchical taxonomy of psychopathology (HiTOP): a dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*. Advance online publication. <http://doi.org/10.1037/abn0000258>
- Krueger, R.F., & Markon, K. E. (2014). The role of the DSM-5 personality trait model in moving toward a quantitative and empirically based approach to classifying personality and psychopathology. *Annual Review of Clinical Psychology, 10*, 477-501. <http://doi.org/10.1146/annurev-clinpsy-032813-153732>
- MacKay, D. J. C. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge, UK: Cambridge University Press. Retrieved from <http://www.inference.phy.cam.ac.uk/mackay/itila/>
- Markon, K. E., Chmielewski, M., & Miller, C. J. (2011). The reliability and validity of discrete and continuous measures of psychopathology: A quantitative review. *Psychological Bulletin, 137*(5), 856–879. <http://doi.org/10.1037/a0023678>
- Masyn, K. E., Henderson, C. E., & Greenbaum, P. E. (2010). Exploring the latent structures of

- psychological constructs in social development using the dimensional-categorical spectrum. *Social Development*, 19(3):470–493. <http://dx.doi.org/10.1111/j.1467-9507.2009.00573.x>
- Maxwell & Delaney (1993). Bi-variate median splits and spurious statistical significance. *Psychological Bulletin*, 113(1):181–190. <http://dx.doi.org/10.1037/0033-2909.113.1.181>
- McCrae, R. R., & Costa, P. T. (2003). *Personality in Adulthood: A Five-Factor Theory Perspective* (2nd ed.). New York, USA: The Guilford Press.
- McLahlan, G., & Peel, D. (2000). *Finite Mixture Models*. New York, NY-USA: John Wiley & Sons, Inc.
- Meeus, W., Van de Schoot, R., Klimstra, T., & Branje, S. (2011). Personality types in adolescence: Change and stability and links with adjustment and relationships: A five-wave longitudinal study. *Developmental Psychology*, 47(4), 1181–1195. <http://doi.org/10.1037/a0023816>
- Molenaar, P. C. M., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science*, 18(2), 112–117.
- Murtagh, F. & Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: Which algorithms implement Ward's criterion? *Journal of Classification*, 31, 274–295. <http://doi.org/10.1007/s00357-014-9161-z>
- Pesonen, A.-K., Räikkönen, K., Heinonen, K., Andersson, S., Hovi, P., Järvenpää, A.-L., ... Kajantie, E. (2008). Personality of young adults born prematurely: the Helsinki study of very low birth weight adults. *Journal of Child Psychology and Psychiatry*, 49(6), 609–617. <http://doi.org/10.1111/j.1469-7610.2007.01874.x>
- Robins, R. W., John, O. P., Caspi, A., Moffitt, T. E., & Stouthamer-Loeber, M. (1996). Resilient, overcontrolled, and undercontrolled boys: Three replicable personality types. *Journal of Personality and Social Psychology*, 70(1), 157–171. <http://doi.org/10.1037/0022-3514.70.1.157>
- Royston, P., Altman, D. G., & Sauerbrei, W. (2006). Dichotomizing continuous predictors in multiple regression: a bad idea. *Statistics in Medicine*, 25(1)127–141. <http://doi.org/10.1002/sim.2331>
- Russell, B. (1945). *A History of Western Philosophy: And its connection with political and social circumstances from the earliest times to the present day*. New York, NY: Simon & Schuster

- Specht, J., Luhmann, M., & Geiser, C. (2014). On the consistency of personality types across adulthood: Latent profile analyses in two large-scale panel studies. *Journal of Personality and Social Psychology, 107*(3), 540–556. <http://doi.org/10.1037/a0036863>
- Tyrer, P., Reed, G. M., & Crawford, M. J. (2015). Classification, assessment, prevalence, and effect of personality disorder. *Lancet, 385*(9969), 717–726. [http://doi.org/10.1016/S0140-6736\(14\)61995-4](http://doi.org/10.1016/S0140-6736(14)61995-4)
- van der Linden, D., te Nijenhuis, J., & Bakker, A. B. (2010). The General Factor of Personality: A meta-analysis of Big Five intercorrelations and a criterion-related validity study. *Journal of Research in Personality, 44*(3), 315–327. <http://doi.org/10.1016/j.jrp.2010.03.003>
- van der Sluis, S., Posthuma, D., Nivard, M. G., Verhage, M., & Dolan, C. V. (2013). Power in GWAS: lifting the curse of the clinical cut-off. *Molecular Psychiatry, 18*(1):2–3. . <http://doi.org/10.1038/mp.2012.65>
- Widiger, T. A., & Mullins-Sweatt, S. N. (2009). Five-factor model of personality disorder: A proposal for DSM-V. *Annual Review of Clinical Psychology, 5*, 197–220. <http://doi.org/10.1146/annurev.clinpsy.032408.153542>

**Table****Table 1. Meta-analytic (above diagonal) and community-sample (below diagonal) correlations used in simulation of multivariate normal datasets**

	O	C	E	A	N
O	-	.14	.31	.14	-.12
C	.15	-	.21	.31	-.32
E	.29	.25	-	.18	-.26
A	.03	.25	.10	-	-.16
N	-.11	-.27	-.18	-.38	-

Abbreviations denote the Big Five traits: “O” = Openness to experience; “C” = Conscientiousness; “E” = Extraversion; “A” = Agreeableness; “N” = Neuroticism. The numbers above the diagonal are from van der Linden et al. (2010) and the numbers below the diagonal from DeYoung et al. (2007).

## Figure captions

*Figure 1. Illustration of properties of clustering algorithms in uncorrelated data. The upper-left panel shows 1000 observations of spherical (uncorrelated) example data. The remaining panels plot cluster means arising when 2-cluster clustering solutions are estimated for 50 simulated datasets of the same kind using three different clustering methods. It is unlikely that the 50 simulated researchers using any of the three methods would end up in a theory of 2 robust personality clusters when the data is spherical, or uncorrelated. The situation changes dramatically for the correlated data (see Figure 2).*

*Figure 2. Illustration of properties of clustering algorithms in correlated data. The upper-left panel shows 1000 observations of correlated data. The remaining panels plot cluster means arising when 2-cluster clustering solutions are estimated for 50 simulated datasets of the respective kind using three different clustering methods. All three methods systematically place the cluster centers in only 2 out of the 4 quadrants.*

*Figure 3. Akaike's information criterion for heuristic and model-based clustering. A dataset of 10 000 observations was simulated similarly as in Figure 2. Clustering models with 1 through 5 clusters were estimated either using k-means or model-based clustering algorithm. The k-means cluster presentation is similar to that in the Ward's method (i.e., spherical). Although the cluster centers are derived differently, both the heuristic methods would favour 2 clusters (minimum AIC achieved for 2-cluster model), whereas the model-based method correctly detects that there is just a single cluster in the data.*

*Figure 4. Clusters from previous empirical study and simulated k-means cluster means. Bars give simulated means and cluster sizes, with whiskers denoting 95% Wald confidence intervals. The dashed-line whiskers give confidence intervals for the cross-national empirical study of RUO types ( $n = 4$  nations). Pattern distance (“dist”) is the Euclidian distance between observed and simulated cluster centers.*

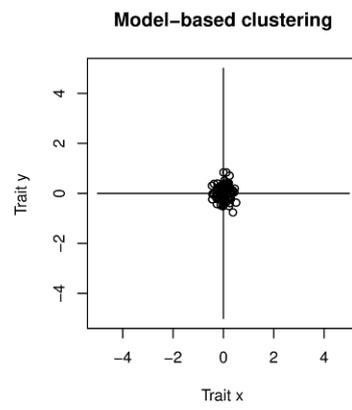
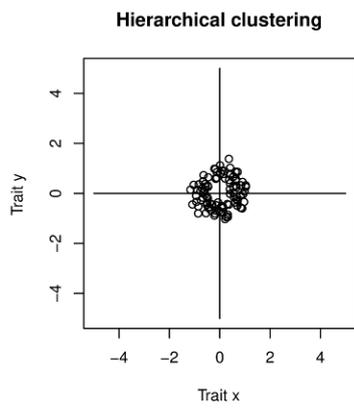
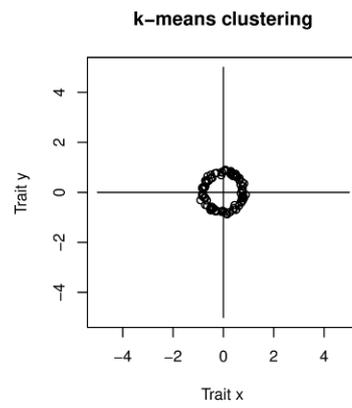
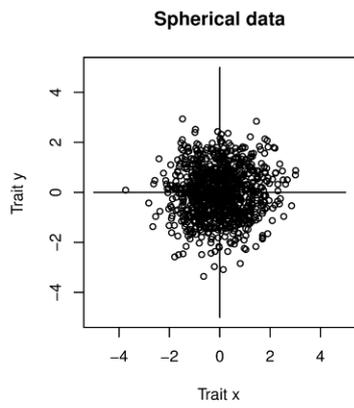
*Figure 5. Clusters from previous empirical study and simulated hierarchical agglomerative cluster means. This analysis replicates the analysis of the Figure 4, demonstrating that the results are similar for the hierarchical agglomerative clustering as for the k-means algorithm.*

*Figure 6. Clusters from previous empirical study and simulated model-based cluster means. This analysis replicates the analysis of the Figure 4, demonstrating that the results are similar for the model-based clustering as for the k-means algorithm.*

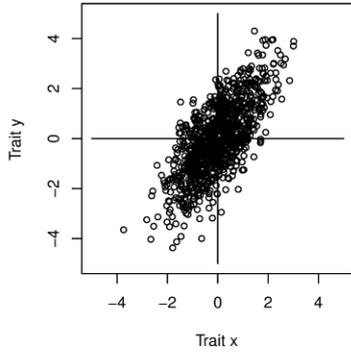
*Figure 7. Simulated k-means cluster means for correlations drawn from the Eugene-Springfield study. This replicates the findings of the Figure 4 to verify that they do not derive from differences between correlation structures of the study populations in the meta-analysis, and to provide a sensitivity analysis. Only correlations of the simulated data were changed in comparison to Figure 4.*

*Figure 8. The k-means cluster means for attenuated correlations. The first row shows the analysis of Figure 2 when all trait correlations were multiplied by  $\frac{1}{2}$ , the second row when multiplied by  $\frac{1}{10}$ , and the last row without any trait correlations. Whiskers denote 95% Wald confidence intervals.*

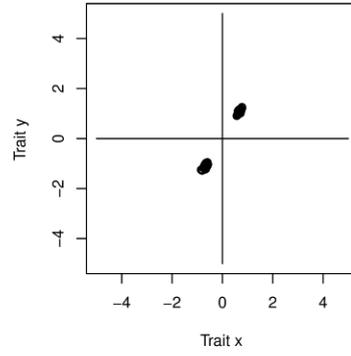
*Figure 9. Three-cluster k-means analysis of the International Personality Item Pool (IPIP) open-access data on Big-Five-Factor markers personality traits. Top row of panels: Clusters and 95% bootstrap percentile confidence intervals as derived from the original IPIP data. Bottom row of panels: Clusters and CIs based on IPIP data after a linear whitening transformation; i.e., after removing trait correlations.*

**Figures**

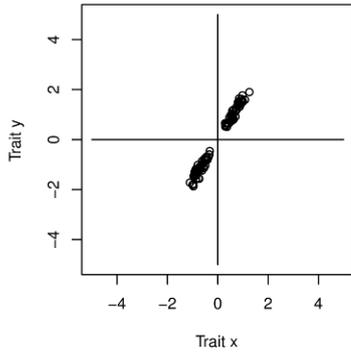
**Correlated data**



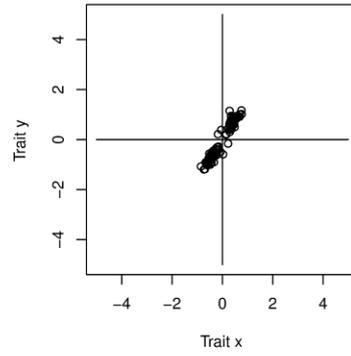
**k-means clustering**

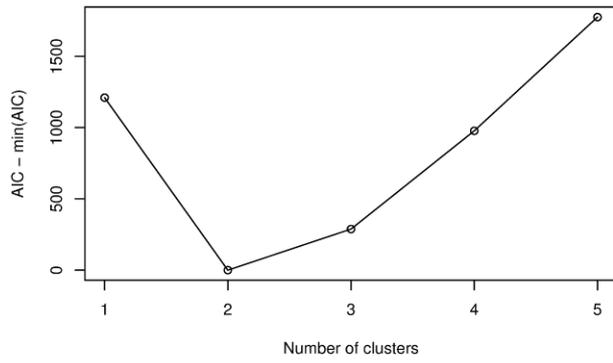


**Hierarchical clustering**



**Model-based clustering**



**Heuristic clustering****Model-based clustering**