nature human behaviour

MATTERS ARISING

https://doi.org/10.1038/s41562-019-0722-3

Reply to: Four personality types may be neither robust nor exhaustive

Martin Gerlach[1], William Revelle[2] and Luís A. Nunes Amaral[1,3,4,5*]

REPLYING TO Freudenstein, J. P., Strauch, C., Mussel, P. & Ziegler, M. *Nature Human Behaviour* https://doi.org/10.1038/s41591-019-0721-4 (2019)

We thank Freudenstein et al.[1] for their thoughtful remarks on our study[2] and for the opportunity to clarify our published results. We share the authors' concern about the premature use of the uncovered types in personality assessment, but must emphasize that the empirical findings of the computational analysis reported in our study remain valid. Indeed, Freudenstein et al. "concur that [their] analyses reveal four meaningful types", highlighting the methods' "notable improvement to previous approaches". We supplement the opinion of Freudenstein et al. on the interpretation of our results along four main directions.

First, we fully and wholeheartedly agree with the statement of Freudenstein et al. urging practitioners to avoid the use of typologies in applied settings. Indeed, while we suggested that personality types are "potentially useful in applied contexts", we also made clear that our study is just a step toward a better understanding of personality in view of the numerous open challenges.

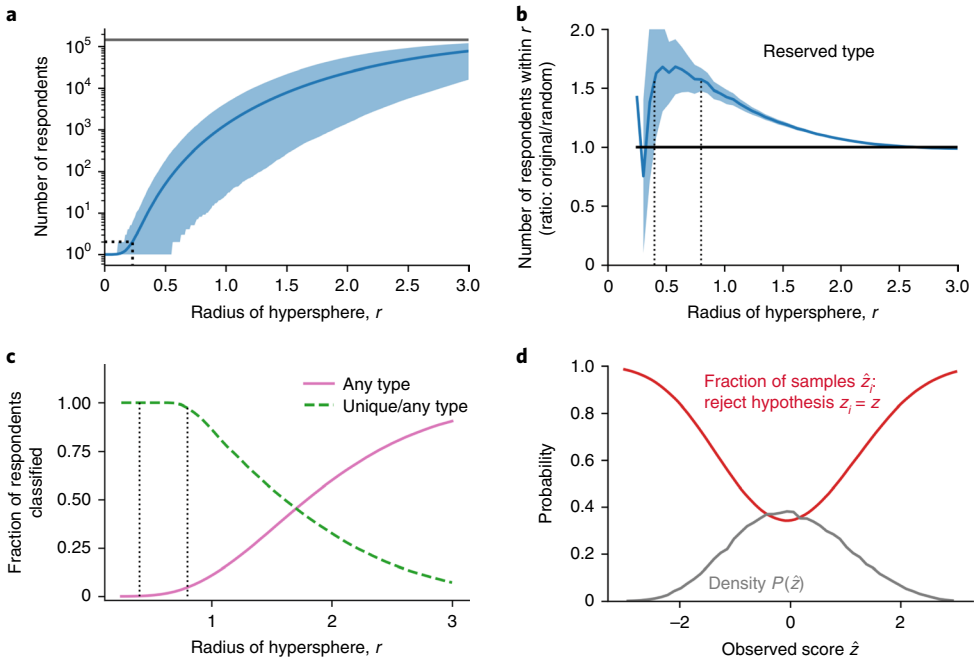

Fig. 1 | Unintuitive properties of the five-dimensional space of personality traits. a, The number of respondents within a hypersphere of radius r with the coordinates of a randomly selected respondent at the centre. The curve shows the mean and the 5th and 95th percentile confidence interval over 1,000 realizations. The solid grey line shows the saturation level of the total number of respondents. The dotted black line shows the radius $r \approx 0.2$ for which one finds the nearest neighbour on average. **b**, The ratio between the number of respondents within a hypersphere radius of r of the 'Reserved' type and the number expected in a randomized dataset. The shaded areas denote the 5th and 95th percentile confidence interval from 100 random realizations. The dotted lines indicate $d = 0.4$ and $d = 0.8$, respectively. **c**, The fraction of respondents within a hypersphere of radius of r of any of the four types ('any type') and the fraction of those within only one type ('unique'/'any type'). The dotted lines indicate the location of $d = 0.4$ and $d = 0.8$, respectively. **d**, Given an observed score $\hat{z}$, the fraction of respondents with observed scores $\hat{z}_i$ for which we reject the hypothesis that the true score $z_i = z$ at the 5% significance level (that is, $P(|z_i - z| > 0.78)$, assuming a reliability of $r = 0.8$) (red curve). The empirical density of the observed scores $P(\hat{z})$ is shown in light grey. The data are from the neuroticism domain of the Johnson-300 dataset.

[1]Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL, USA. [2]Department of Psychology, Northwestern University, Evanston, IL, USA. [3]Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL, USA. [4]Department of Physics and Astronomy, Northwestern University, Evanston, IL, USA. [5]Department of Medicine, Northwestern University, Evanston, IL, USA. *e-mail: amaral@northwestern.edu

Second, we disagree with the statement of Freudenstein et al. that "personality models should be exhaustive" and believe that the underlying premise is conceptually flawed. In fact, exhaustiveness does neither imply that the model is supported empirically nor that it is useful. For example, previous typologies for classifying individuals such as the four ancient temperaments may be exhaustive, but their empirical status is weak[3]. Every model has a limited range of validity and not all predictions of a model may be verifiable due to detectability thresholds. We also want to emphasize that nowhere in the original study do we state that the goal is to assign every individual to a type. Instead, we show compelling empirical evidence for the existence of several robust clusters with higher-than-expected density. Moreover, and of greater consequence, selectively imposing a criterion of exhaustiveness on typologies introduces a double standard in the evaluation of personality models that overlooks the uncertainty and limitations of trait-based approaches.

Third, we agree with the observation of Freudenstein et al. that only 42% of respondents were associated with one and only one of the four personality types. This confirms the difficulty in identifying the typological structure in the space of personality, in great part due to the associated measurement error of the underlying personality traits. To gain intuition on the challenges overcome by our analyses, we investigate the distribution of scores in a five-dimensional space of personality traits. Even with the size of the samples considered (>100,000), resolution is limited. Indeed, the average Euclidean distance to the nearest neighbour of a randomly selected respondent is about 0.2 (Fig. 1a). Therefore, when comparing the actual and expected number of respondents within a hypersphere of radius r of any cluster, for small radii, there simply are not enough data points to obtain any meaningful calculations. For intermediate radii, however, the enrichment is as high as 75% (Fig. 1b). The data in the figure suggest that a more appropriate radius for type assignment should be in the range $0.4 \leq d \leq 0.8$. For example, choosing $d = 0.8$, the fraction of respondents assigned to any cluster is 5%, with almost all of those (98%) assigned uniquely to one of the types (Fig. 1c). Increasing the radius will lead to a larger coverage at the expense of greater overlap across clusters. As one considers these values, it is helpful to reiterate the typical uncertainty expected from trait-based measurements. Given an observed normalized single-dimension trait score $\hat{z}$ for a test with reliability r (typically not larger than 0.8), we know that there is a 95% chance that the true score is in the range $z = r \times \hat{z} \pm 1.96\sqrt{r(1-r)}$ (ref. [4]). Thus, when attempting to measure individual differences, the true score of a respondent with an observed score $\hat{z} = 0$ is statistically significantly different from only 40% of the respondents in the dataset (Fig. 1d). Only for the more extreme scores $|\hat{z}| > 2$ (for which there are far fewer available data) does this fraction surpass 80%. Clearly, one must also be careful when using trait-based approaches in applied settings.

Fourth, Freudenstein et al. question the robustness of our results due to the larger number of "meaningful" clusters appearing in their analysis of the additional datasets. We emphasize that the authors' findings exactly replicate the results reported in our original study. In fact, we explicitly showed how this phenomenon results from larger uncertainty due to the decrease in the number of items administered in some of the datasets. Indeed, we observed that the reduction in the number of items in the reference dataset led to a similar increase in the number of meaningful clusters (Supplementary Figs. 11 and 12 in ref. [2]). As a result, our original study reports the existence of at least four distinct personality types, namely those that could be reproduced across all datasets.

To construct rigorous type assignment procedures, we will need to build a much deeper understanding of the nature of the uncertainty in the estimation of the values of the individual traits for an individual respondent. More generally, we want to reaffirm that we do not see the existence of personality types as a negation of the fundamental importance of personality traits[5]. In fact, our findings suggest that there is a typological structure within the paradigm of personality traits. It remains an open challenge for future studies to unify this dichotomy in the search for a consensual paradigm for personality.

Data availability

Data are available from https://osf.io/tbmh5/.

Code availability

Code is available in a GitHub repository at https://github.com/amarallab/personality-types.

Received: 3 July 2019; Accepted: 7 August 2019;
Published online: 16 September 2019

References

1. Freudenstein, J.-P., Strauch, C., Mussel, P. & Ziegler, M. Four personality types may be neither robust nor exhaustive. *Nat. Hum. Behav.* https://doi.org/10.1038/s41591-019-0721-4 (2019).
2. Gerlach, M., Farb, B., Revelle, W. & Amaral, L. A. N. A robust data-driven approach identifies four personality types across four large data sets. *Nat. Hum. Behav.* **2**, 735–742 (2018).
3. Revelle, W., Wilt, J. & Condon, D. M. in *The Wiley-Blackwell Handbook of Individual Differences* (eds Chamorro-Premuzic, T. et al.) (Wiley-Blackwell, 2013).
4. Revelle, W. & Condon, D. M. in *The Wiley Handbook of Psychometric Testing: A Multidisciplinary Reference on Survey, Scale and Test Development* (eds Irwing, P., Booth, T, & Hughes, D.J.) ch. 23 (Wiley, 2019).
5. McCrae, R. R. & Costa, P. T. in *The SAGE Handbook of Personality Theory and Assessment* Vol. 1 (eds Boyle, G. J. et al.) 273–294 (SAGE, 2008).

Acknowledgements

L.A.N.A. acknowledges a John and Leslie McQuown Gift to NICO and support from the Department of Defense Army Research Office under grant number W911NF-14-1-0259. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

M.G., W.R. and L.A.N.A. designed research; M.G., W.R. and L.A.N.A. performed research; M.G. analysed data; M.G., W.R. and L.A.N.A. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to L.A.N.A.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.