# Reply to Katahira et al. "Distribution of personality: Types or skewness?"

Martin Gerlach, William Revelle, Luis A. N. Amaral

Northwestern University

Katahira et al.[1] investigated the potential impact of skewness in the marginal distributions of personality trait on the findings reported by us in Gerlach et al.[2]. We concur with Katahira et al.'s finding in synthetic 2-dimensional data that there *exists a mechanism* by which skewness can induce detection of "meaningful clusters" using our proposed methodology. Here, we argue that skewness cannot fully account for the deviations from the null model for the data studied in our original study.

Katahira et al. consider a 2D unimodal distribution with skew and show that the procedure described in our study[2], incorrectly "identifies" the existence of several clusters. Unarguably, this is an unavoidable consequence of model misspecification – i.e. using a mixture of Gaussians to approximate a non-Gaussian distribution. However, from a purely practical perspective, the two models might be statistically indistinguishable given only a finite amount of data. Indeed, the empirical joint distributions of data of the same sample size generated from the skew model and the fitted GMM are extraordinarily similar (Fig. 1). In fact, we barely reject the hypothesis that the data from the "true" model was generated by the GMM with a 0.05-significance threshold (empirical 1-sided p-value=0.012 from 1000 synthetic datasets using the likelihood as a goodness-of-fit statistic).

Katahira et al. thus suggest that the same misspecification "could have happened in [our] analysis" citing the fact that the empirical marginal distributions are skewed. However, below we show that skewness alone cannot account for the latter observation by counting the number of respondents in each of the 32 orthants (Fig. 2). We approximate the contribution from skewness by drawing trait scores from the marginal distributions observed in the data but assuming that trait scores are uncorrelated. We observe that the data sampled from the skewed marginals does lead to substantial and significant differences in the number of respondents across orthants. Most importantly, however, it cannot account for the much larger number of respondents in the orthants containing the detected clusters – the 3 orthants which can be unambiguously associated

32  with one of the clusters are among the 5 most over-populated orthants. Interestingly, the number

33  of respondents from the skewness model is compatible with the number of respondents of the

34  third-most populated orthant – a orthant for which we did not detect a cluster despite it containing

35  a very large number of respondents.

36

37    We disagree with Kathira et al.'s statement that "The categorical view posits that there are

38  multiple clusters (dense regions) in personality space (i.e. the distribution is multimodal) and

39  each individual can be classified into one of these clusters. In personality theory, such clusters are

40  referred to as personality 'types'."    Rather we discuss higher levels of density in parts of

41  distribution, but do not claim that each individual can or should be classified into one of these

42  regions. In fact, we want to reaffirm that we do not see the existence of personality types as a

43  negation of the fundamental importance of personality traits.    Our findings, which are

44  strengthened by this analysis, suggest that there is a typological structure within the paradigm of

45  personality traits. In fact, through the null models used in our analysis we have firmly established

46  that the degree of concentration of respondents in certain orthants cannot be accounted for by

47  neither the dimensional view nor the skewness. It remains an open challenge for future studies to

48  gain a better understanding of the multivariate distribution of personality traits in order to

49  establish a consensual paradigm for personality. Further, we caution that a dichotomous view on

50  either dimensions or categories is likely to lead to a sterile debate.

51

52  Data and code are available in the github-repository https://github.com/amarallab/personality-

53  types.

54

55  [1] Katahira, K., Kunisato, Y., Yamashita, Y., & Suzuki, S. Distribution of personality: Types or

56  skewness? *PsyArXiv* (2019). Preprint available at https://psyarxiv.com/wua9j/

57  [2] Gerlach, M., Farb, B., Revelle, W. & Amaral, L. A. N. A robust data-driven approach

58  identifies four personality types across four large data sets. *Nat. Hum. Behav.* **2,** 735–742 (2018).
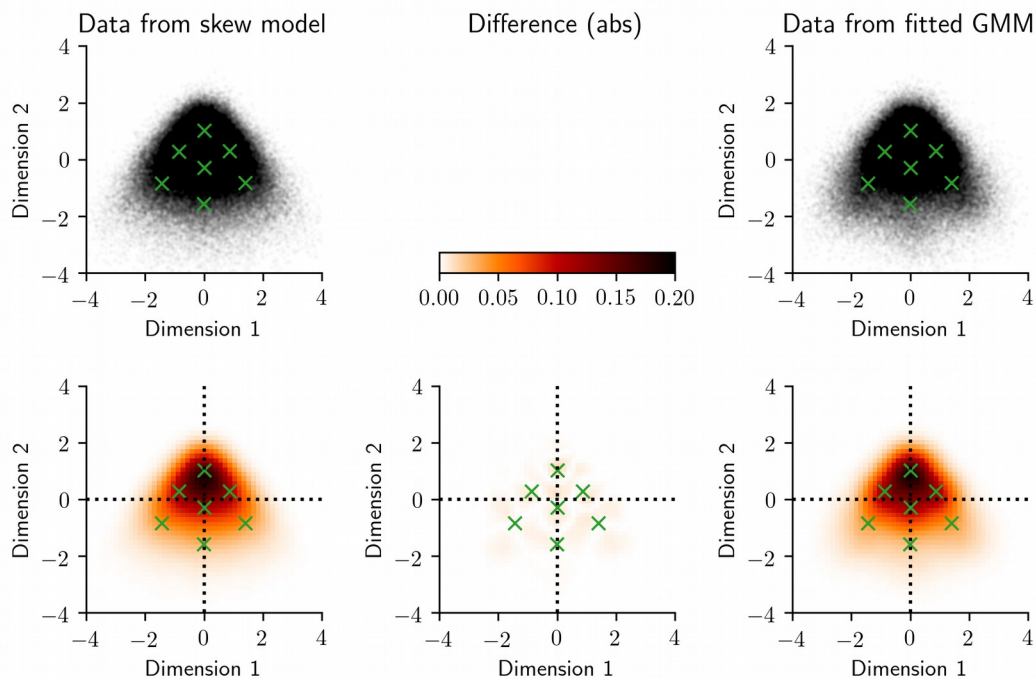
**Figures**



*Figure 1. Data from skew model are statistically indistinguishable from data generated with a fitted Gaussian mixture model.*

*Left: Scatter plot (top) and empirical density (bottom) generated from Katahira et al.'s procedure (N=145,388).*

*Right: Scatter plot (top) and empirical density (bottom) generated from fitted GMM with 7 clusters (N=145,388).*

*Middle: Absolute difference between the empirical densities from the two models.*

*Green crosses indicate positions of 7 cluster centers from the fitted GMM.*

*Testing the hypothesis that the data from the skew model was generated by a GMM with 7 clusters yields an empirical p-value=0.012 (one-sided) from 1000 synthetic datasets using the likelihood as a goodness-of-fit statistic (z-score = 2.12).*
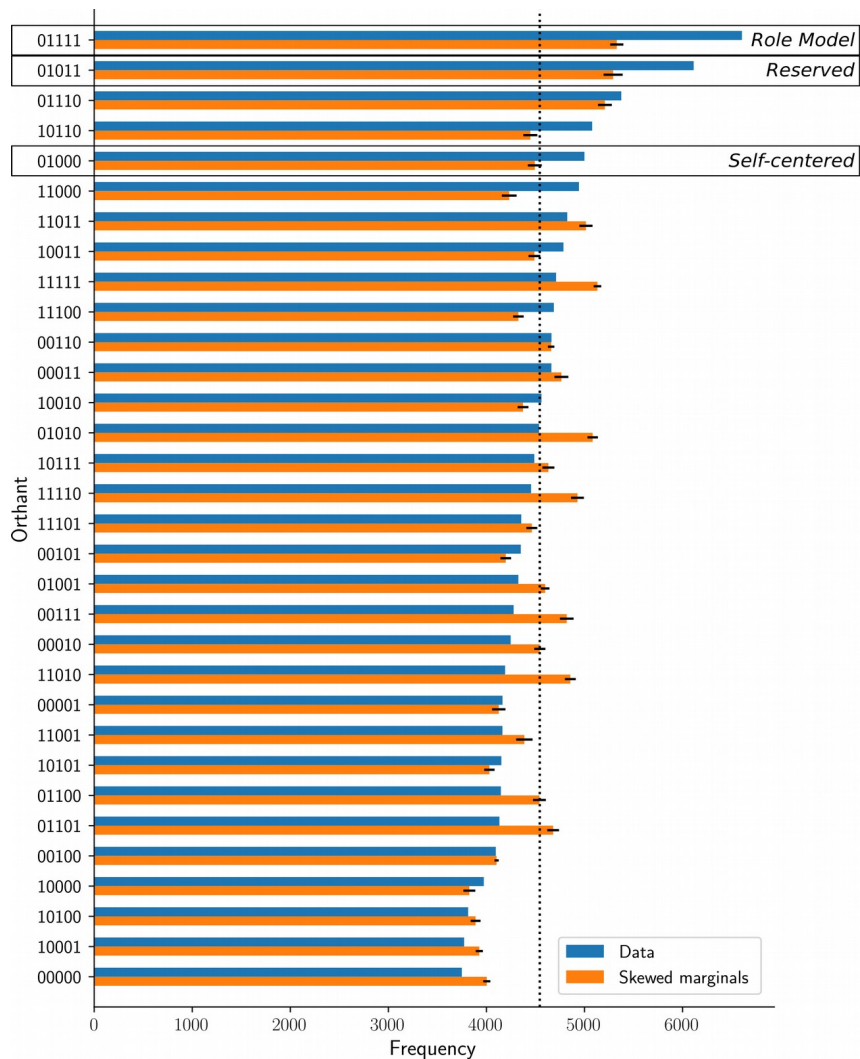
*Figure 2. Skewness alone is not sufficient to account for excess in number of respondents in orthants containing personality types.*

*For each orthant in the 5D space of personality traits ('0': z<0, '1':z>0) we count the number of respondents (blue) and compare with a null model drawing traits independently from the skewed marginals (orange).*

*We mark the orthants containing 3 personality types; note that one of the identified types in our original study was 'Average' which cannot be unambiguously associated with a orthant.*