Developing new practices for increasing transparency in social science research

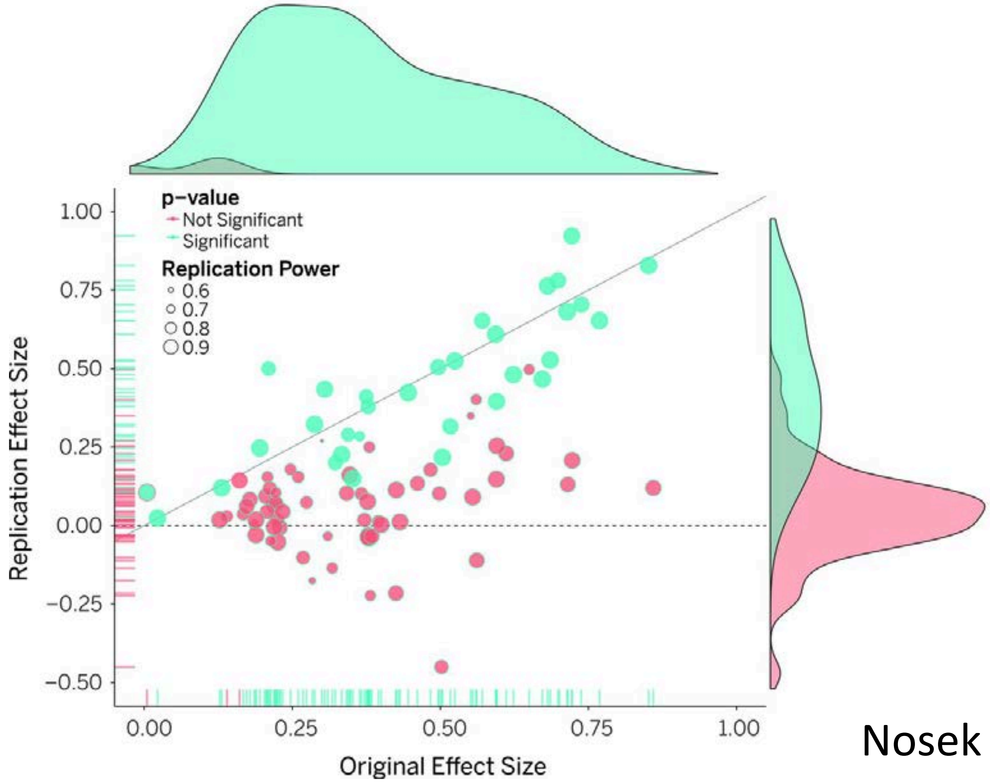
Emily H Ho
Department of Psychology

Outline

- Reproducibility crisis
- Study 1: Theoretical
- Study 2: Empirical (subjective judgment)
- Conclusion

Reproducibility Crisis

- Nosek et al. (2015) found that the results from about 100 studies, using original data, could not be replicated
  - 97% percent of original studies had significant results ($P < .05$).
  - Only 36% percent of replications had significant results



Nosek et al. (2015)

How better to evaluate design and results of studies?

- Empirical distributions, or data occurring ‘in the wild’, is often non-normal
- Yet there is no statistically robust way of quantifying the extent of non-normality between two such distributions
- Additionally, simulation studies often attempt to investigate the extent of non-normality on a certain phenomenon

Conflicting definitions of non-normality

	Values	Method of simulation
Woods, 2008	Skewness = 1.57 Kurtosis = 3.52	Mixture of two normal distributions (M1 = −0.253, M2 = 2.192, S1 = 0.609, S2 = 1.045, mp1 = .897, mp2 = .103)
Preston & Reise, 2014	Skewness = 1.75 Kurtosis = 6.75	Mixture of normal distributions Bimodal (M1 = 21.5, M2 = 3.0, S1 = 0.7, S2 = 1.5, mp1 = 1.0, mp2 = 0.7)
Savalei, 2010	Skewness = 2 Kurtosis = 7	Fleishman (1978) method
Enders, 2001	Moderate nonnormality: Skewness=1.25 Kurtosis = 7.0 Extreme nonnormality: Skewness=3.25 Kurtosis= 20.0	(Fleishman, 1978; Vale & Maurelli, 1983)

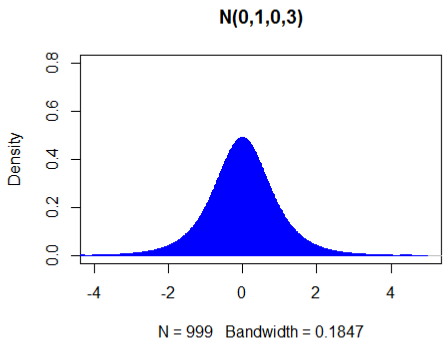
Theoretical

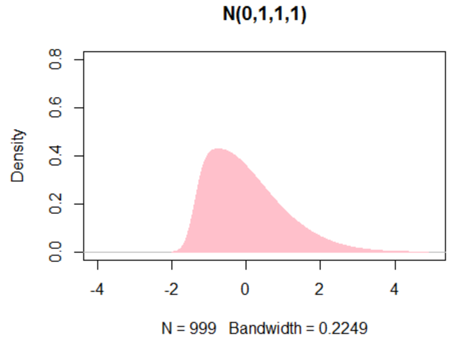
- Goal: to evaluate a suite of theoretical distances, as applied to constructed non-normal distributions

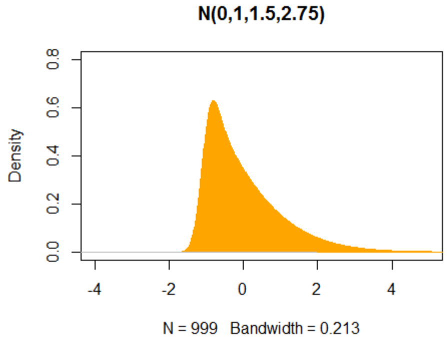
- Using the Fleishman (1978) method of constructing non-normal distributions, we simulate 107 distributions of the form F~(0,1,x,y))

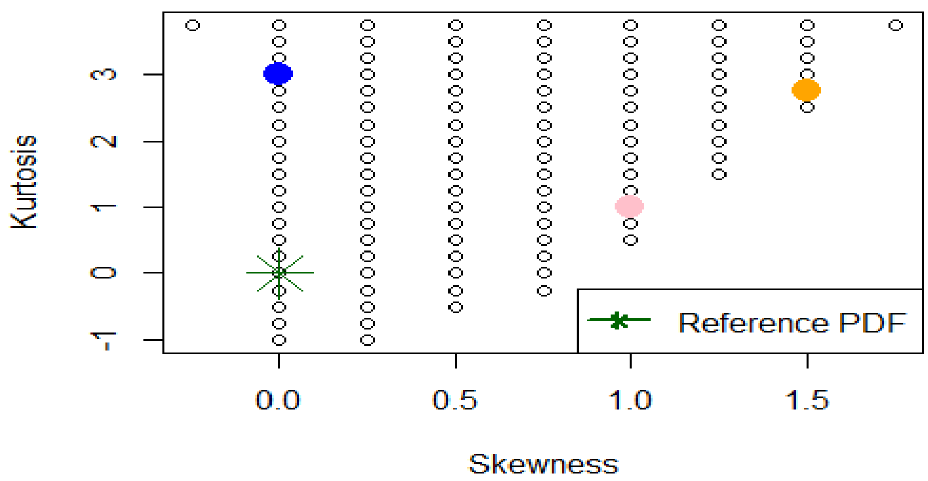
- x = Skewness: ranged from [-0.25, 1.75]
- y = Kurtosis: ranged from [-1, 3.75]

Theoretical





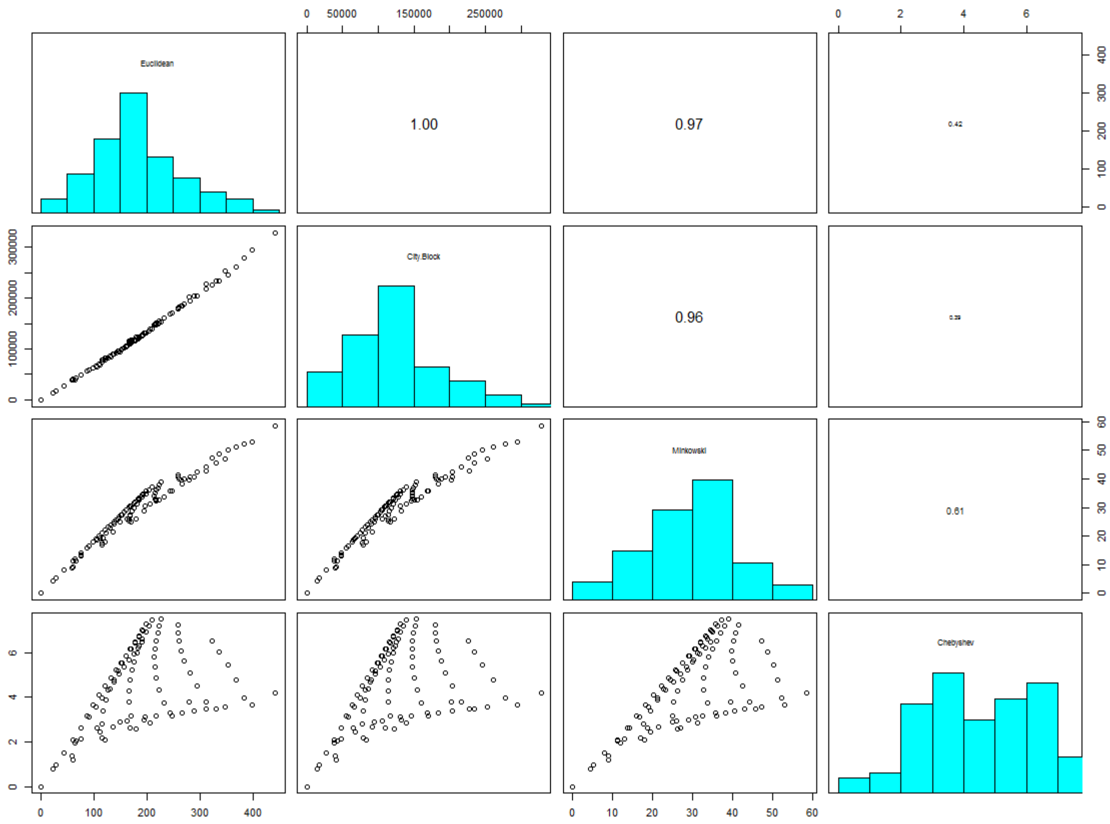




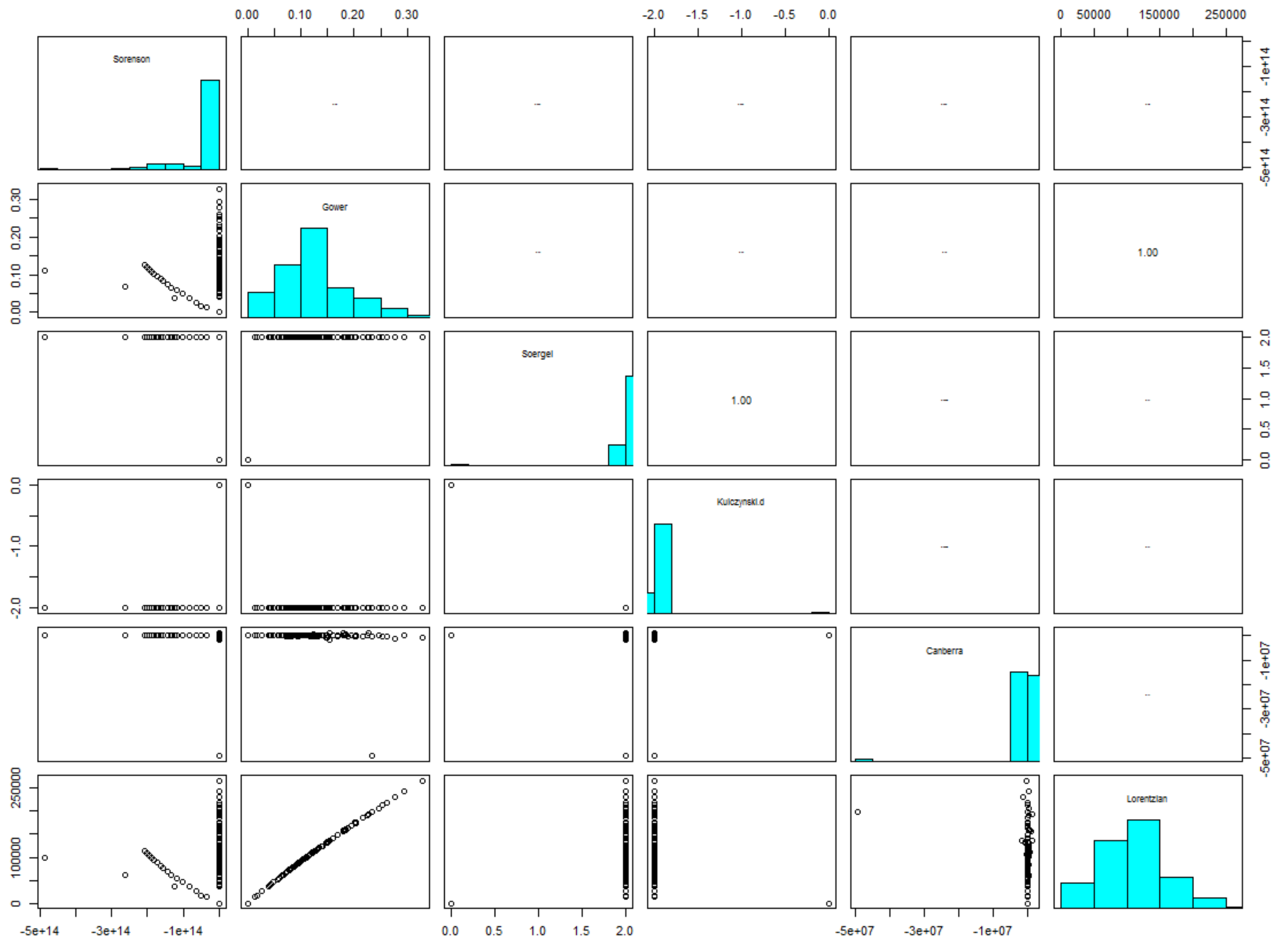
Theoretical

- Using 17 measure s from 3 classes, we calculated the distance between 107 non-normal distributions and the standard normal, X~N(0,1,0,0).
  - Minkowski (includes Euclidean, Chebyshev distance)
  - L1 (includes Sorenson)
  - Intersection (includes WavesHedges)

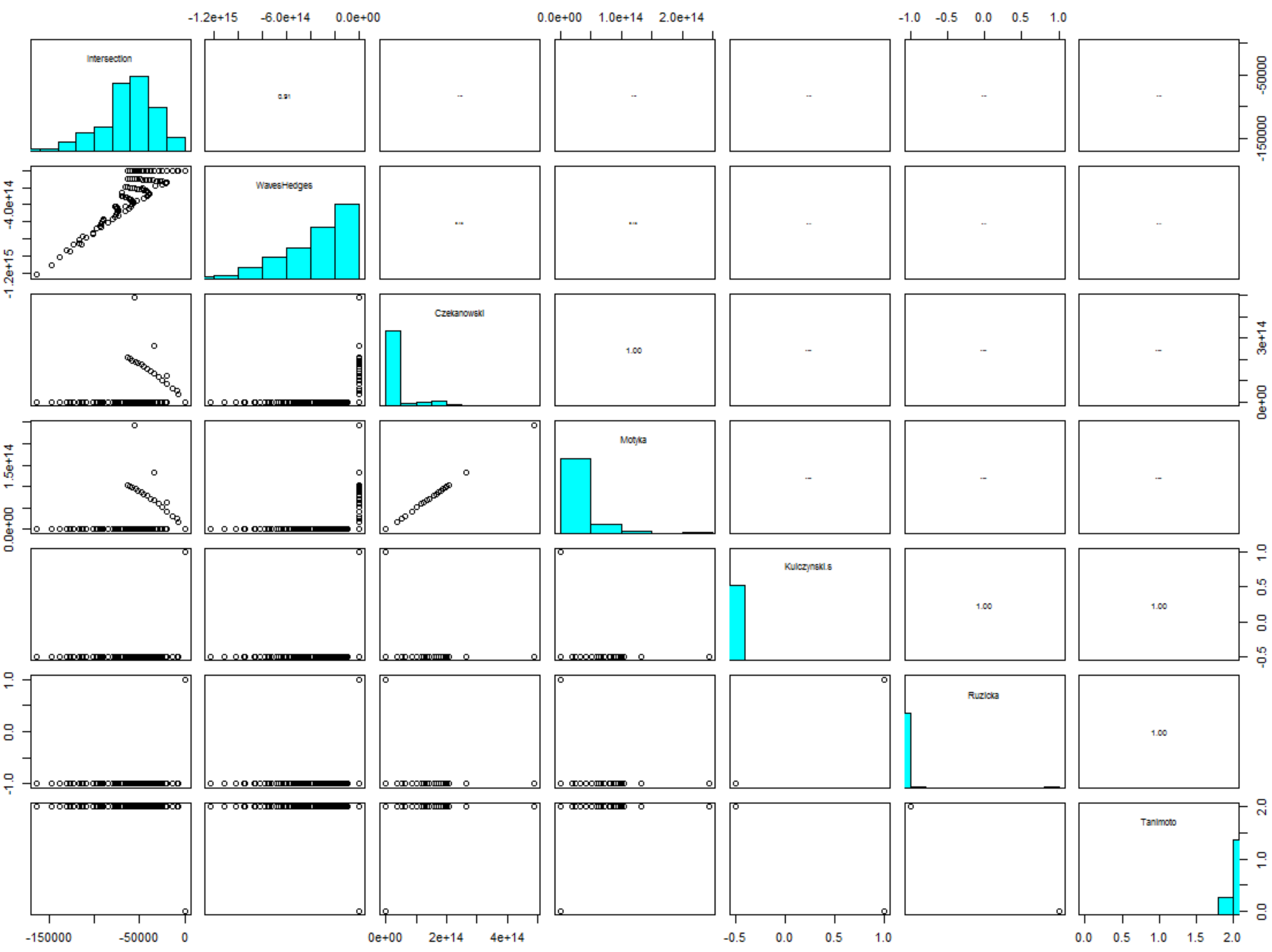
Theoretical (Minkowski)



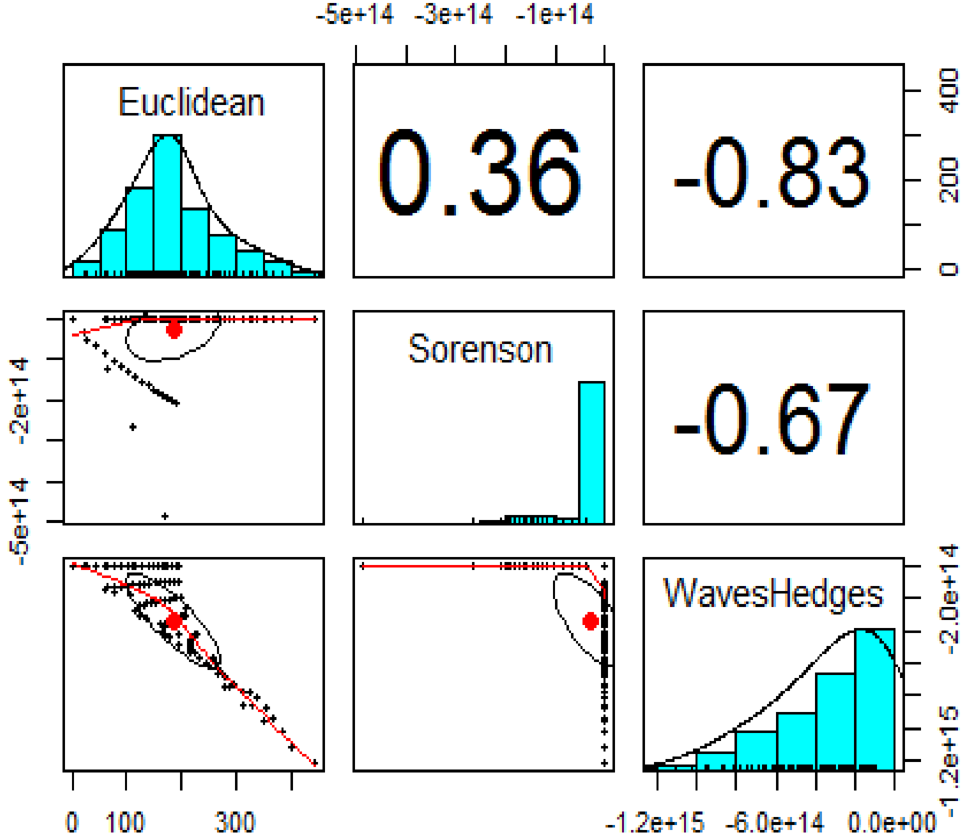
Theoretical (Lp family)



Theoretical (Intersection Family)



Theoretical (summary)



Theoretical

- The intra-class distance measures exhibit little agreement with each other, with some measures appearing to be minimally affected by skewness and kurtosis
- To more clearly delineate a ‘new’ grouping of distances based on their sensitivity to skewness and kurtosis, polynomial regression was conducted:

$$Distance = Intercept + Skewness + Kurtosis + Skewness^2 + Kurtosis^2 + Skewness * Kurtosis$$

Prescribed Taxonomy

Distance	Int	Skew	Kurt	Skew^2	Kurt^2	Skew*Kurt	R^2
Euclidean	1	1	0	1	1	1	0.93
City.Block	1	1	0	1	1	1	0.94
Minkowski	1	1	0	1	1	1	0.9
Chebyshev	1	1	1	1	1	0	0.92
Sorenson	1	1	0	1	0	0	0.42
Gower	1	1	0	1	1	1	0.94
Soergel	1	0	0	0	0	0	0.06
Kulczynski.d	1	0	0	0	0	0	0.06
Canberra	0	0	0	0	0	0	0.58
Lorentzian	1	1	0	1	1	1	0.93
Intersection	1	1	0	1	1	1	0.94
WavesHedges	0	1	1	1	1	1	0.98
Czekanowski	1	1	0	1	0	0	0.42
Motyka	1	1	0	1	0	0	0.42
Kulczynski.s	1	0	0	0	0	0	0.06
Ruzicka	1	0	0	0	0	0	0.06
Tanimoto	1	0	0	0	0	0	0.06

Variance Explained Taxonomy

x	Int	Skew	Kurt	Skew^2	Kurt^2	Skew*Kurt	R^2
Euclidean	1	1	0	1	1	1	0.93
City.Block	1	1	0	1	1	1	0.94
Minkowski	1	1	0	1	1	1	0.9
Gower	1	1	0	1	1	1	0.94
Lorentzian	1	1	0	1	1	1	0.93
Intersection	1	1	0	1	1	1	0.94
Sorenson	1	1	0	1	0	0	0.42
Czekanowski	1	1	0	1	0	0	0.42
Motyka	1	1	0	1	0	0	0.42
Chebyshev	1	1	1	1	1	0	0.92
Canberra	0	0	0	0	0	0	0.58
WavesHedges	0	1	1	1	1	1	0.98

The Soergel, Kulczynski.d, Kulczynski.s, Ruzicka, and Tanimoto were predicted only by the intercept, $R^2 = 0.06$

Theoretical: Summary

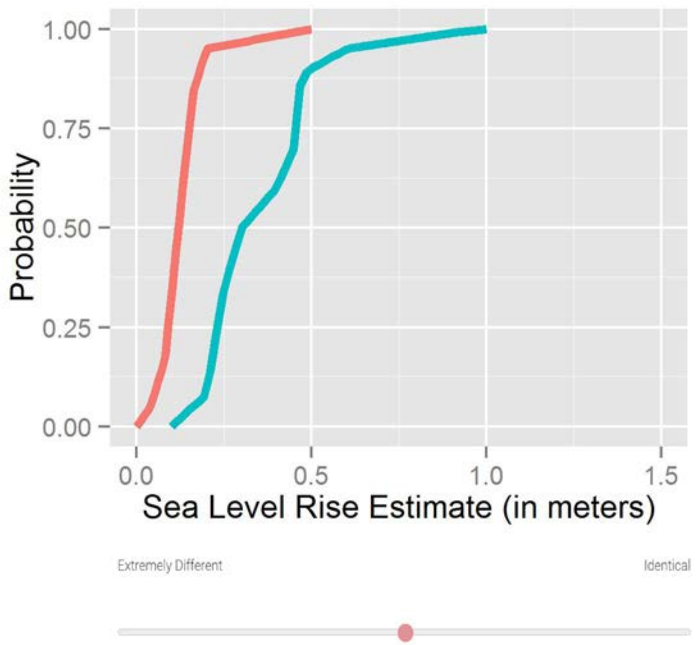
- Theoretical distances do not cleanly map onto the non-normal distributions of varying skewness and kurtosis
- The distances are, in some cases, completely insensitive to these moments
- A variance explained taxonomy of distances reveals very different groupings than the theoretical classes

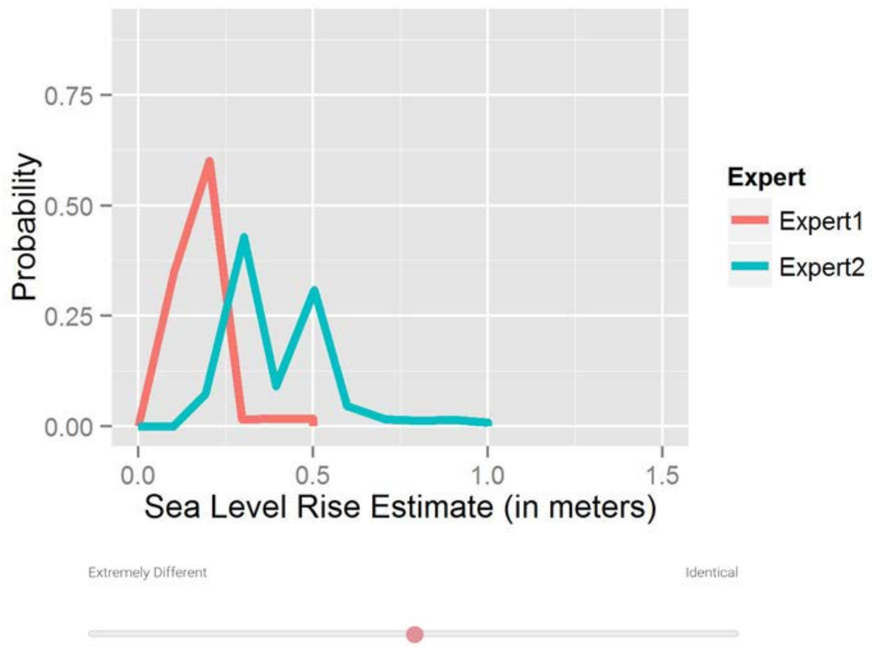
- Future work: more distance classes, different ways of producing non-normal distributions (e.g., mixture proportions), different distribution parameters (hyper-moments, tail weights)

Empirical (psychophysical)

- How well can laypeople distinguish differences between distributions?
- Thomas et al (2015) asked 6 experts their assessment of sea level rise in the year 2050.
- Using two display formats (cumulative distribution functions and probability density functions), we asked laymen to evaluate their perceived distance between two experts

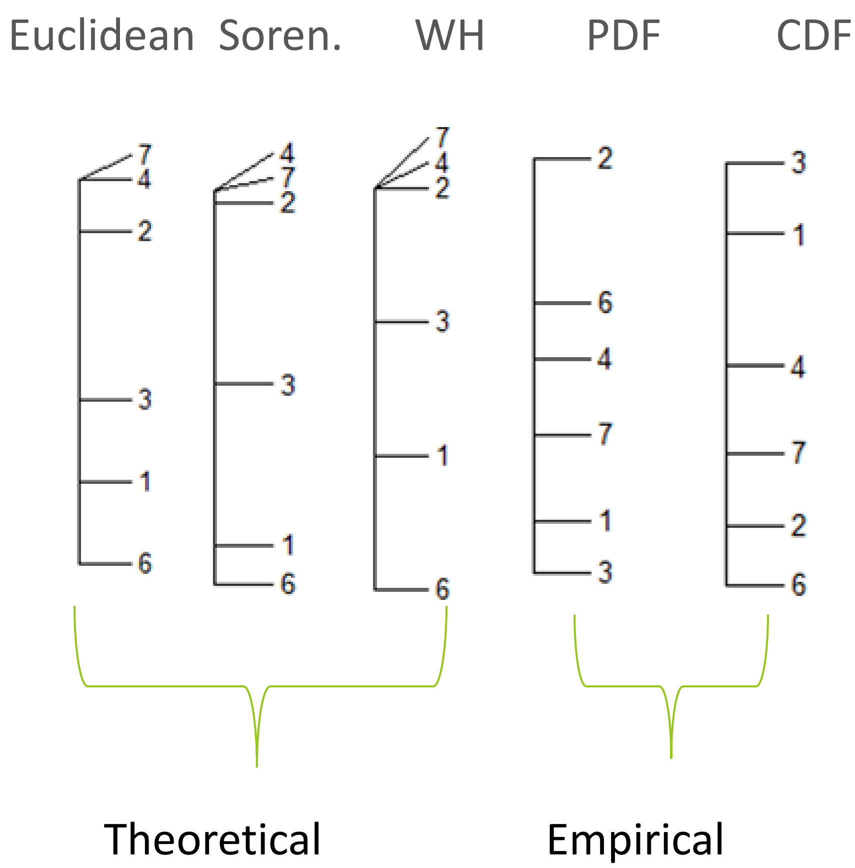
Experiment Display





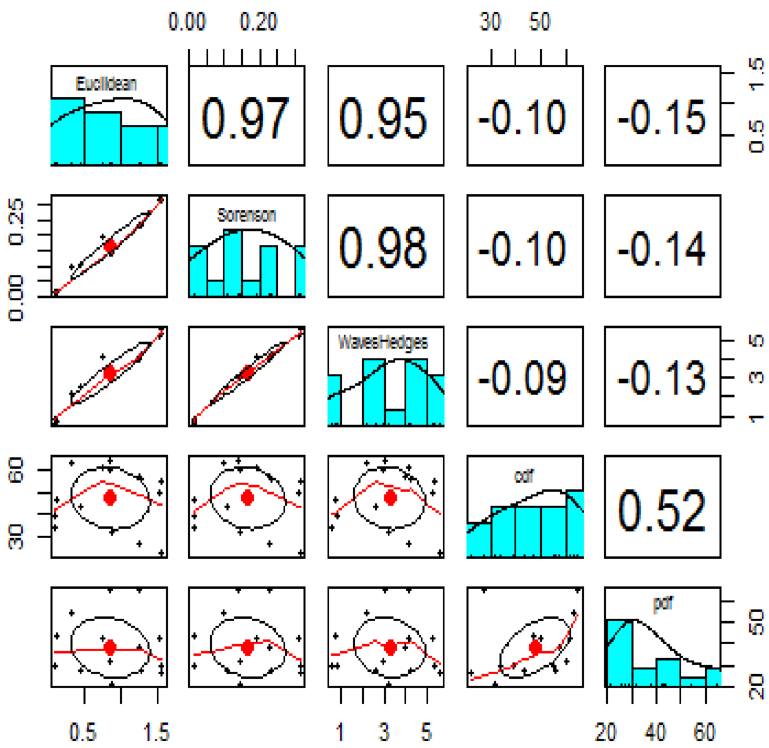
Empirical

- Unidimensional scaling of the theoretical distances (same used in Study 1) and the median subjective judgments (pdf and cdf)



Empirical

- The theoretical metrics agree on the clustering of expert judgments
- The empirical metrics have moderate agreement



Conclusion

- Theoretical metrics perform non-uniformly when assessing the extent of systematic deviance from a standard normal distribution
- In the empirical study, we found (a) high agreement between theoretical metrics, (b) moderate agreement (r=0.5) between empirical judgments based on pdfs and cdfs, and (c) low agreement between empirical judgments and the theoretical metrics
- Psychophysical methods do not agree with purely statistical methods
- Future experiments will use theoretical distributions (as in those generated by the Fleishman method) to compare more directly psychophysical and statistical judgments

References

- Deza, M., & Deza, E. (2009). *Encyclopedia of Distances*. Springer Berlin Heidelberg.
- Enders, C. K. (2001). The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. *Psychological Methods*, *6*(4), 352.
- Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika*, *43*(4), 521–532.
- Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, *105*(1), 156.
- Nosek, et al (2015). Estimating the reproducibility of psychological science. Science. doi: 10.1126/science.aac4716.
- Preston, K. S. J., & Reise, S. P. (2014). Estimating the Nominal Response Model Under Nonnormal Conditions. *Educational and Psychological Measurement*, *74*(3), 377–399. http://doi.org/10.1177/0013164413507063
- Savalei, V. (2010). Expected versus observed information in SEM with incomplete normal and nonnormal data. *Psychological Methods*, *15*(4), 352–367. http://doi.org/10.1037/a0020143
- Thomas, M., Pidgeon, N., Whitmarsh, L., & Ballinger, R. (2015). Expert judgements of sea-level rise at the local scale. *Journal of Risk Research*, 1–22. http://doi.org/10.1080/13669877.2015.1043568
- Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, *48*(3), 465–471.
- Woods, C. M. (2008). Likelihood-Ratio DIF Testing: Effects of Nonnormality. *Applied Psychological Measurement*, *32*(7), 511–526. http://doi.org/10.1177/0146621607310402

Theoretical Distance Formulas

- Euclidean (of the Minkowski family)- "straight-line" metric distance

$$d_{Euc} = \sqrt{\sum_{i=1}^{d} | P_i - Q_i |^2}$$

- Sorenson (of the L_1 family) – a proportion coefficient of overlap that is semi-metric (can have two distinct points with a distance of 0); more sensitive to outliers

$$d_{sor} = \frac{\sum_{i=1}^{d} | P_i - Q_i |}{\sum_{i=1}^{d} (P_i + Q_i)}$$

- WavesHedges (of the Intersection family) – metric distance used as a measure of overlap; sensitive to histogram binning
-

$$d_{WH} = \sum_{i=1}^{d} \left(1 - \frac{\min(P_i, Q_i)}{\max(P_i, Q_i)}\right)$$

$$= \sum_{i=1}^{d} \frac{| P_i - Q_i |}{\max(P_i, Q_i)}$$