Strength in numbers: Testing the fidelity of multidimensional scaling for large datasets

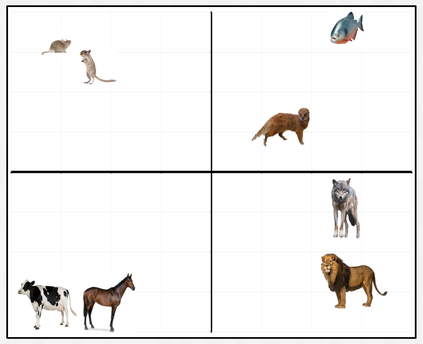
Arryn Robbins [1], Corbin A. Cunningham [2], Justin MacDonald [1], & Michael C. Hout [1]

[1] New Mexico State University; [2] Johns Hopkins University

MULTIDIMENSIONAL SCALING

- Multidimensional scaling (MDS) is a statistical technique that is used to model the psychological similarity among stimuli (see Hout, Papesh, & Goldinger, 2012; Hout et al., 2015).
- Similarity is important in many research areas of psychology (e.g. visual search) and having access to stimuli that are organized by similarity is a valuable tool. Databases of objects cataloguing similarity relationships do exist (e.g. Hout, Goldinger, & Brady, 2014), but the size of such item sets are currently somewhat restrictive. Importantly, in order to develop large sets of stimuli (with measured similarity among items) – be it for databases or experimentation – it is necessary to first examine the fidelity of MDS when used with large item sets.
- MDS has often been conducted on stimuli in the range of 10-30 items. There are three potential reasons for the adoption of these relatively small stimulus set sizes: Computational effort, data collection time, and concerns about the fidelity of the data.
- As the stimulus set increases, the number of comparisons required to perform an analysis grows dramatically, increasing the computational effort.
  - Specifically, for *n* stimulus items, *n(n-1)/2* ratings are required, such that each item is compared with every other at least once.
  - E.g. For as few as 30 items, 435 pairwise ratings are required; adding another 5 items to the stimulus set (35 items) brings the required comparisons up to 595, and a doubling of the original set (60 items) requires 1,770 ratings.
- However, conducting MDS analyses on many item sets is now logistically feasible, due to the adoption of spatial approaches (e.g. the spatial arrangement method; Hout, Goldinger, & Ferguson, 2013), and the advent of "big data" collection procedures that can spread out pairwise ratings across many people.
- Left to be understood is how the MDS analyses themselves hold up to large item sets (though see Young, 1970, for an approach similar to our own).

In a set of simulations, we tested the ability of MDS to accurately quantify the similarity of stimulus sets larger than 30 items (i.e. up to 1024 items).



A hypothetical 2-dimensional MDS solution for 8 animals. The vertical dimension appears to represent size, and the horizontal dimension might represent ferocity.

METHOD

Experiment 1

- Simulated 10 participants by creating "true" spaces and adding noise to distances between points
- Also simulated spaces that were pure noise
- 3x4x9 design
  - Dimensionality (2, 3, 4)
  - Error (5%, 25%, 50%, pure noise +/- to item distances)
  - Item Set Size (4, 8, 16, 32, 64, 128, 256, 512, 1024).
- PROXSCAL scaling algorithm

Experiment 2

- To show results of Experiment 1 were not idiosyncratic to using PROXSCAL, we used ALSCAL
- Identical to Experiment 1, except with the use of the ALSCAL scaling algorithm
- ALSCAL can only use up to 100 items, so we used 4-64 items.

Experiment 3

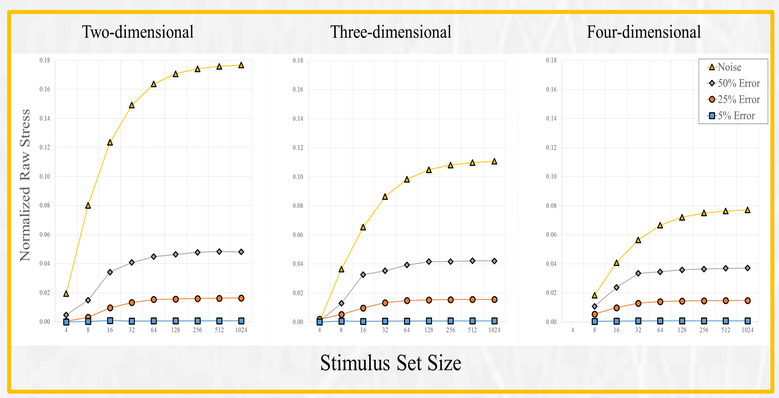
- Same as Experiment 1
- Used mdscale in MATLAB for scaling
- Dimensionality was also underestimated and overestimated
- Fixed error of 50%
- 500 iterations of each simulation

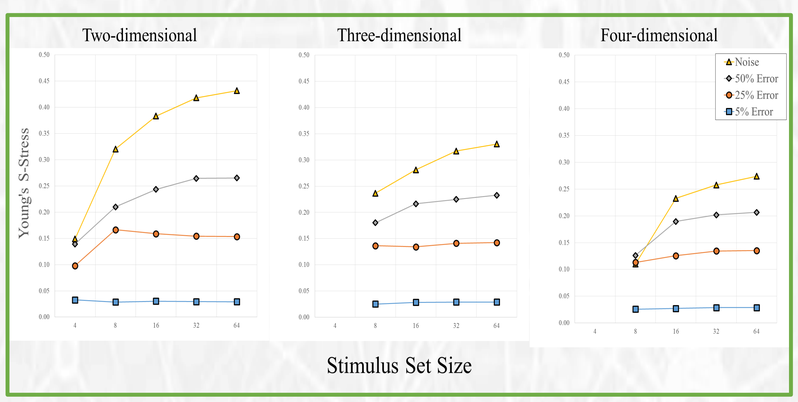
Determinacy: The degree to which the MDS solutions *determined* the correct organization of points

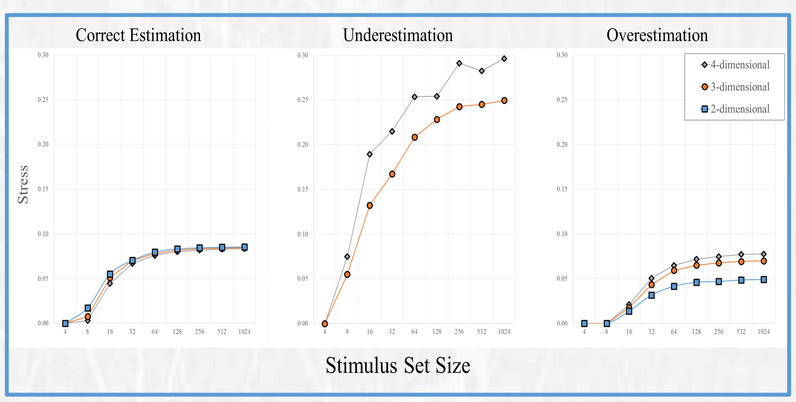
Correspondence: The degree to which the raw simulated data points *correspond* with the "true" multidimensional distances

SIMULATION RESULTS

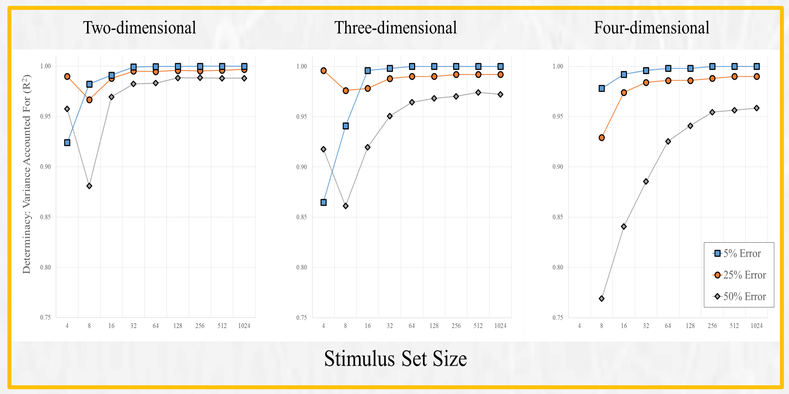
STRESS

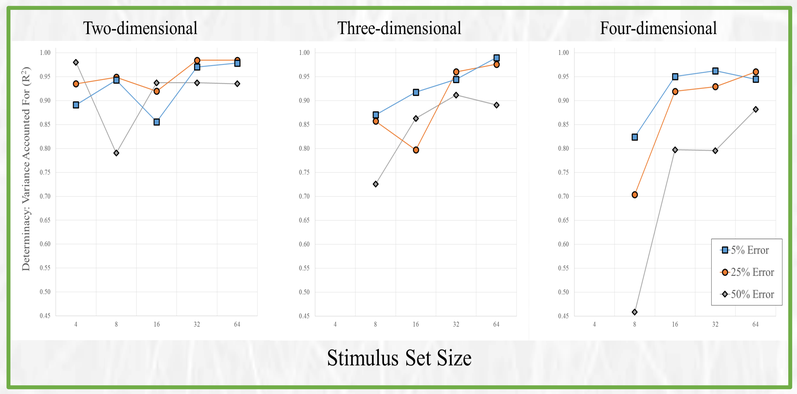


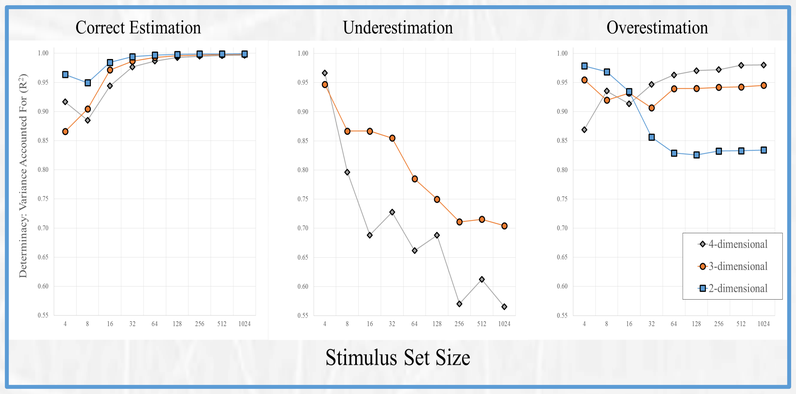




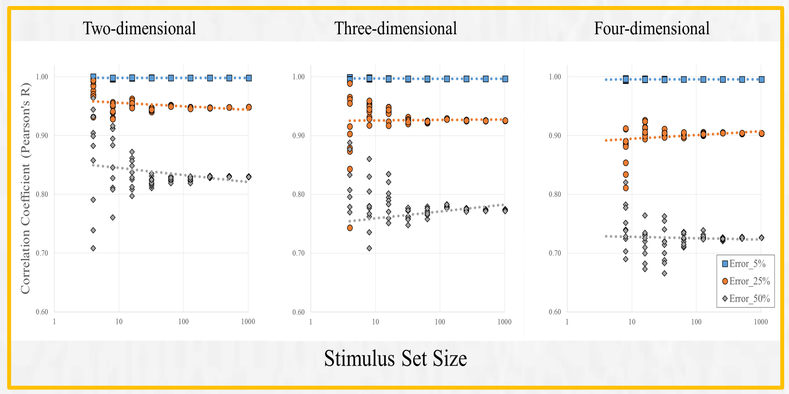
DETERMINACY







CORRESPONDENCE



CONCLUSIONS

- Higher set sizes decrease model fit (i.e., they produce increased "stress"), but increased the determinacy of the MDS spaces.
- We also showed (not pictured) that this finding disappears when purely random data are generated, demonstrating that the benefit of using large stimulus set sizes comes from the "signal" in the data, not the "noise."
- These results are consistent across different MDS scaling algorithms, and across thousands of simulated iterations, provided the data are scaled in the appropriate dimensionality.
- We show that this large set benefit is contingent on the proper selection of dimensionality for the space, and that when the dimensionality of the solution is over- or underestimated, larger set sizes can be detrimental to the quality of the solution.
- We argue that is it not only reasonable to adopt large stimulus set sizes, but advantageous to do so. However, we also stress that it is crucially important to correctly estimate the dimensionality of the MDS space, and to err on the side of overestimation, rather than underestimation of dimensionality.