Testing a Fast, Efficient Method for Multidimensional Scaling

Michael C. Hout, Ryan W. Ferguson, Donald Homa & Stephen D. Goldinger

Arizona State University

50th Annual Meeting of the Psychonomic Society

ABSTRACT

The classic methods used in multidimensional scaling, although useful, suffer from several shortcomings. Specifically, as the stimulus set increases, possible pairwise comparisons grow exponentially. This leads to lengthy experimental protocols for participants, or procedures that involve scaling only subsets of stimuli. In the present project, we examined a method proposed by Goldstone (1994) in which scaling is accomplished by presenting many stimuli at once. The participant moves the stimuli around the screen, placing them at distances from one another that are proportional to the user's subjective similarity ratings. This method takes advantage of the spatial nature of similarity, and provides a fast, efficient method for obtaining an MDS space. We provide evidence that the Goldstone method works well on controlled 2- and 3-D visual stimuli and on non-visual stimuli with less well defined dimensions. We also make our software (written in E-Prime) available to the public through the first author's Web site.

What is Multidimensional Scaling (MDS)?

⌘ MDS is a set of statistical techniques for constructing representations of the psychological structure of a set of stimuli (Shepard, 1980). The output is a spatial configuration wherein stimuli are plotted in locations that best represent their perceived similarity to other stimuli (i.e., the greater distance between two stimuli, the larger the dissimilarity).

⌘ The notion that stimuli can be modeled in such a way that perceived similarities are represented by spatial proximities dates back to Isaac Newton (who suggested that spectral hues be represented on a circle).

⌘ MDS algorithms use matrices of item-to-item similarities (or dissimilarities), typically obtained through the use of rating, sorting, or perceptual confusion tasks.

⌘ The researcher decides on the number of dimensions that the algorithms generate (i.e., the number of coordinate values used to locate a point in space). Since MDS is often a descriptive model for understanding data, decisions about dimensionality are made according to interpretability, ease of use, and stability (Kruskal & Wish, 1978).

⌘ With more dimensions, there is a better statistical fit to the data, but the results are harder to interpret. Data plotted in a sufficiently low dimensionality permit a visual examination of the underlying psychological structure.

Drawbacks of Classic MDS

⌘ Classic MDS involves obtaining a similarity rating for every possible pairwise combination of stimuli (typically via Likert Scales). Drawbacks include…

⌘ *Inefficiency*: many judgments must be made to obtain a solution. The number of comparisons increases as a quadratic function of the number of compared objects.

⌘ *Protocol duration*: subjects may change their strategies over time, become fatigued, or simply disengage and rate arbitrarily.

⌘ *Memory*: recollection of ratings made to previous stimuli may influence future similarity ratings.

⌘ *Low resolution*: discrete ratings narrow the range of responses that people can make, limiting resolution on individual trials (also, sorting and confusion methods average over binary outcomes).

The *Apperceptive* Method

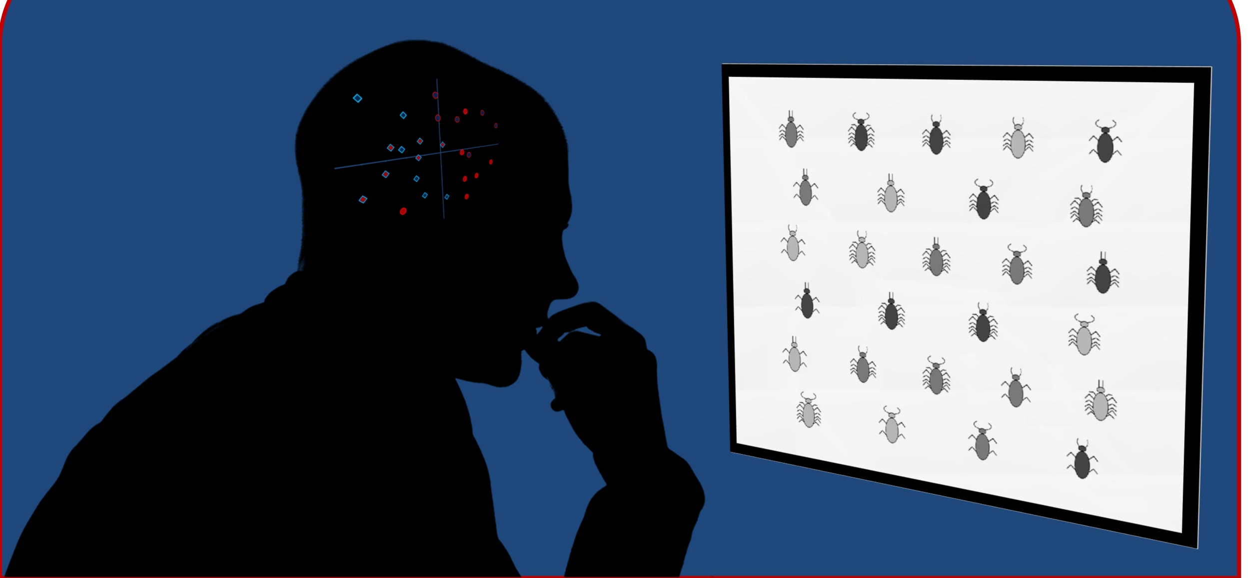
⌘ We investigated a novel approach to obtaining MDS data, introduced by Goldstone (1994).

⌘ This technique involves having people arrange stimuli on a computer screen spatially, by dragging and dropping individual items. Items that are perceived as similar are placed close together, and dissimilar items are moved farther away.

⌘ This can be thought of as having people create their own MDS space.

⌘ Stimuli are first presented in discrete rows, with randomized item placement. The participant then uses the mouse to organize the stimuli according to their perceived similarities.

⌘ The output is a matrix of item-to-item Euclidean distances (dissimilarities).



Discrete, Visual Stimuli

⌘ We generated 4 stimulus sets, created to possess discrete dimensions. The goal was to determine if the methods "discovered" the intended dimensions. Stimulus sets were either 2- or 3-dimensional; the latter included those of the original 2-D stimuli, and one additional dimension.

⌘ *Spokes*: 2-D varied in line thickness, and angle of the spoke. 3-D included a dimension of background color.

⌘ *Bugs*: 2-D varied in body color, and number of legs. 3-D included a dimension of antannae curvature.

Non-Visual Stimuli

⌘ We also employed 2 sets of non-visual stimuli; participants were asked to indicate the similarity of various animals. People were presented with the names of the animals, rather than pictures.

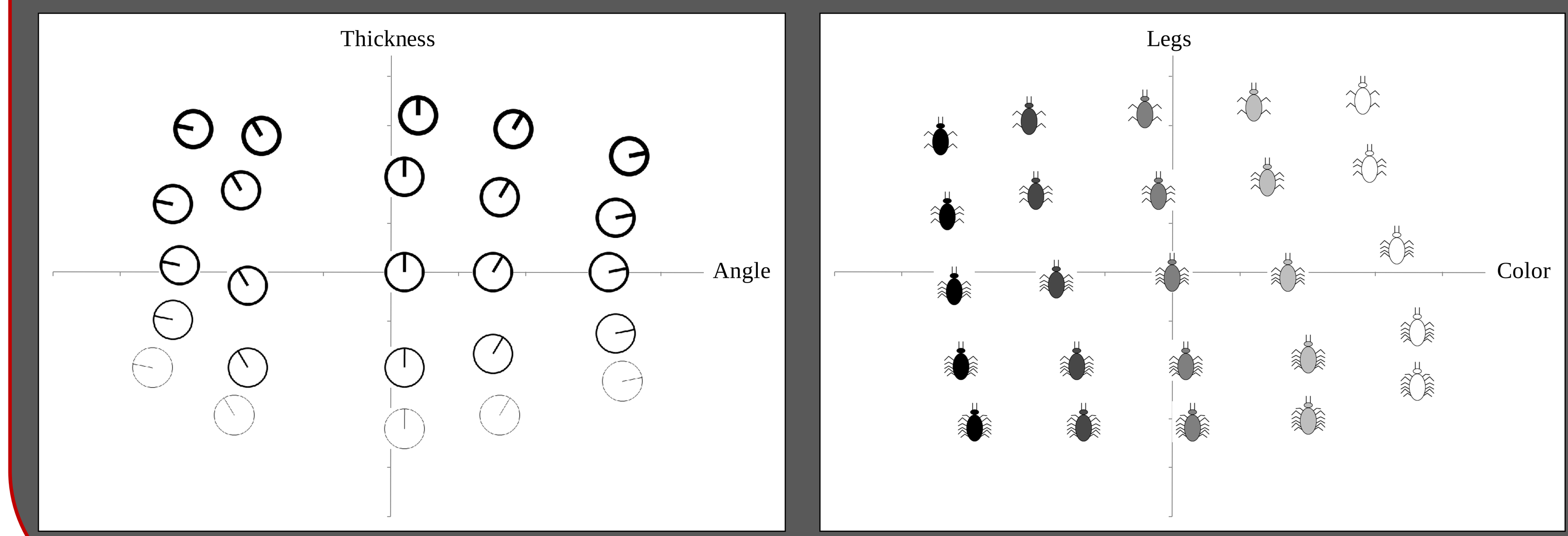
⌘ *Categorical*: the first set of animals varied categorically on two specific dimensions. All animals were either birds or non-birds, and primarily resided in either water or on land. (Adapted from Hornberger, et al., 2009).

⌘ *Continuous*: the second set of animals were chosen to vary continuously on two dimensions: size and domesticity. (Adapted from Henley, 1969).

Software

⌘ The Apperceptive method was programmed by the first author in E-Prime v1.2 (Psychology Software Tools, 2006). Data was analyzed in SPSS, using the Proxscal procedure (Data Theory Scaling Systems Group, 2006).

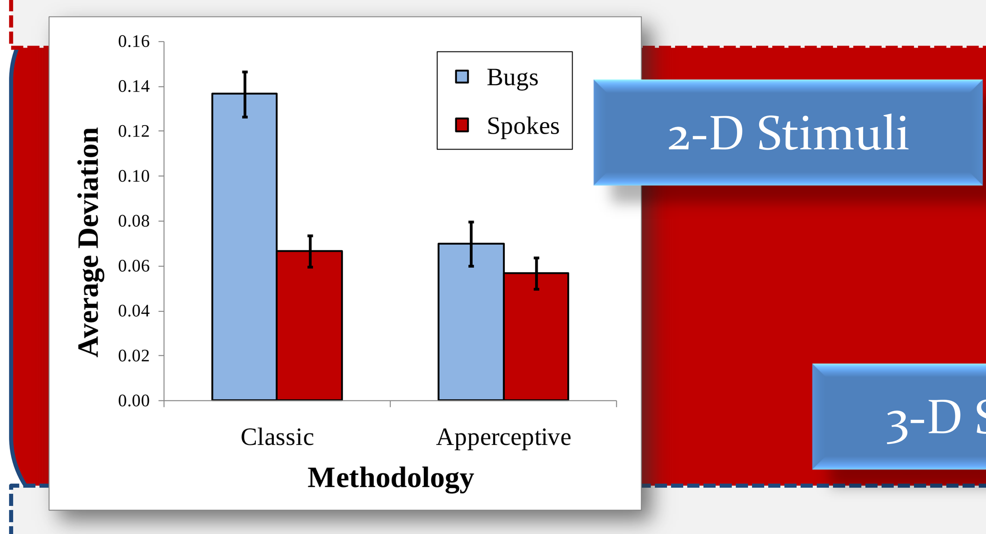
Apperceptive Solutions (2-D stimuli)

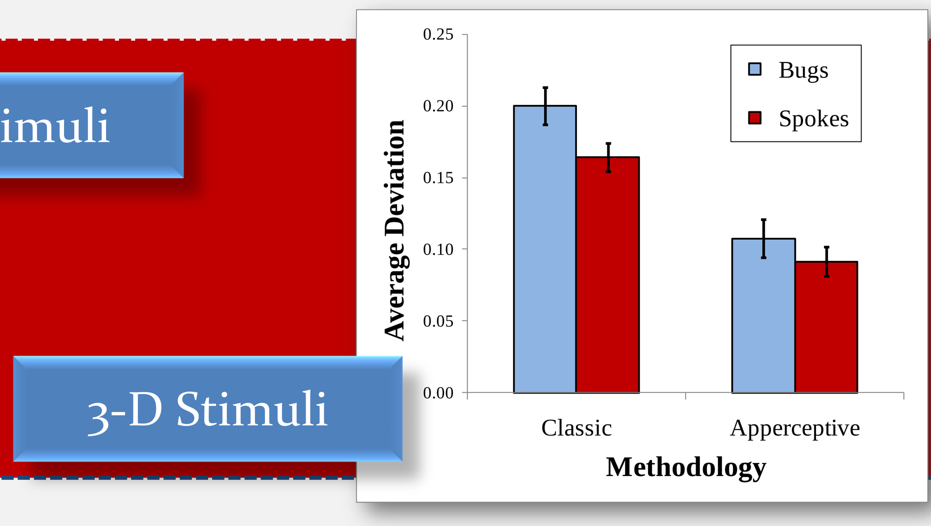


Solution Assessment

⌘ There is no test-independent method for revealing "true similarity" (Goldstone & Medin, 1994). Therefore, in order to objectively assess the quality of the solutions produced by the Apperceptive method, we compared the solutions to those generated by the classic MDS technique.

⌘ With respect to the animal stimuli, we also compared Apperceptive solutions to those generated by Latent Semantic Analysis (Landauer & Kintsch, 2003).





Results (Discrete Stimuli)

⌘ To assess the quality of the discrete solutions, we compared the Proxscal coordinates to those of "ideal" solutions, scaled to the maximum values on the methods' respective dimensions. We then calculated deviation-from-ideal values for each item (per dimension), and entered them into a 3-way ANOVA: Methodology (Classic, Apperceptive) x Stimuli (bugs, spokes) x Dimension (primary, secondary, tertiary).

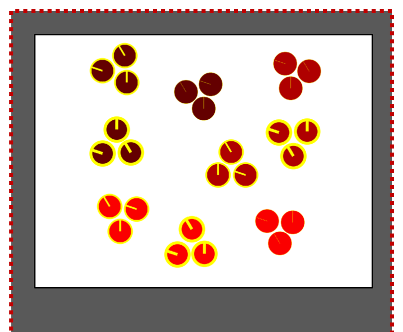
⌘ ANOVA results for the 2-D stimuli showed smaller mean deviations for the Apperceptive method (M = .06), relative to Classic (M = .10), $F(1, 48) = 13.59$, $p < .001$, $\eta_p^2 = .22$. ANOVA results for 3-D stimuli also showed an advantage for the Apperceptive method (M = .10), relative to Classic (M = .18), $F(1, 52) = 34.14$, $p < .001$, $\eta_p^2 = .40$.

⌘ Following Goldstone (1994), we performed correlations on the inter-point distances to ascertain measures of internal consistency. Each technique was compared to its "ideal" solution, and the methods were also compared to one another.

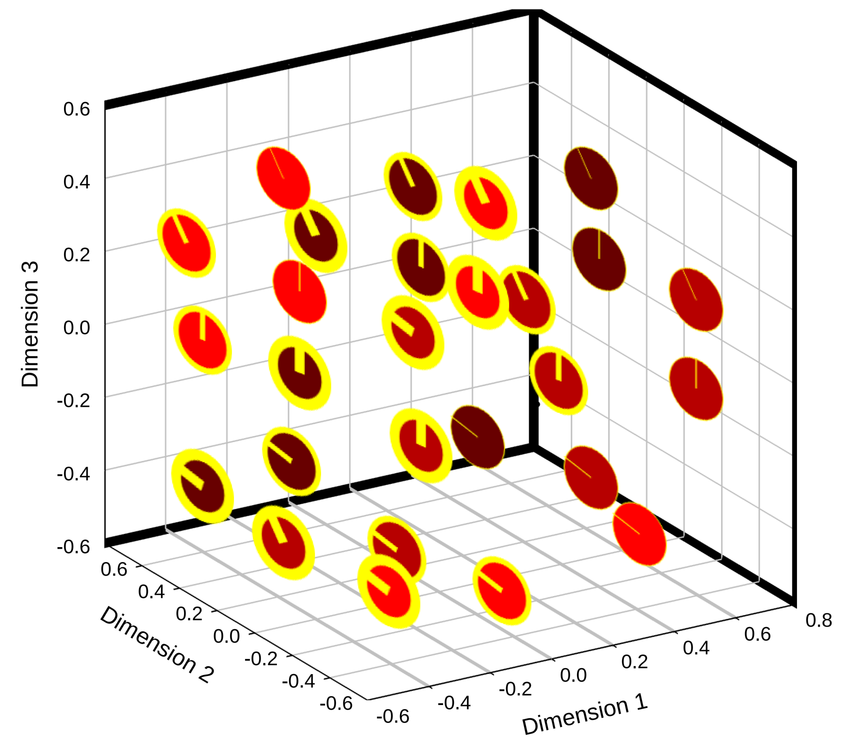
⌘ Results showed that all 2-D solutions were correlated with their "ideal" solutions, all $ps < .001$. On average, correlations were quite high (Apperceptive, $r = .98$; Classic, $r = .96$). Also, all 3-D solutions were significantly correlated with "ideal", all $ps < .001$ (Apperceptive, $r = .97$; Classic, $r = .81$).

⌘ Moreover, the solutions were consistent between methodologies, all $ps < .001$. Correlations were again quite high for both 2-D ($r = .96$, $r = .94$, for bugs and spokes, respectively), and 3-D stimuli ($r = .74$, $r = .95$, for bugs and spokes).

Apperceptive Solution (3-D spokes)



Sample solution obtained by the Apperceptive Method



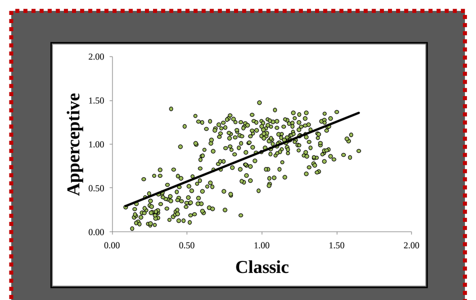
Results (Non-Visual Stimuli)

⌘ For the *categorical* animals, we performed a chi-square analysis on successful categorization (per dimension) by each of the techniques. For the Bird/NonBird dimension, the Apperceptive method (24) had the greatest number of successful categorizations, followed by Classic (23), and LSA (20), $\chi^2(2) = 3.64$, *n.s.* For the Land/Water dimension, the Apperceptive method (24) again had the greatest number of successes, followed by Classic and LSA (both 17), $\chi^2(2) = 7.45$, $p < .05$.

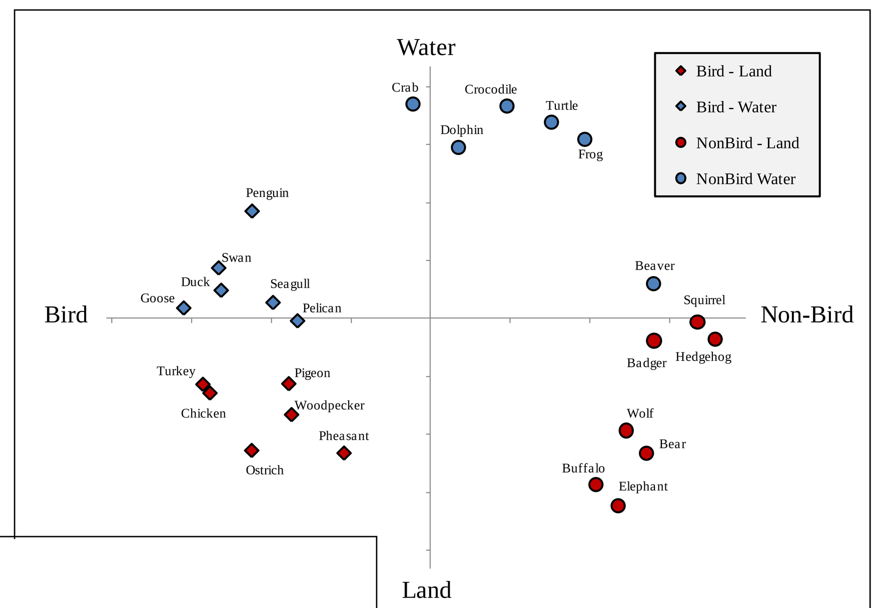
⌘ Moreover, inter-point distances from the Apperceptive and Classic solutions were significantly correlated ($r = .69$), and each method was correlated with those of LSA ($r = .44$, $r = .36$ for Apperceptive and Classic, respectively), all $ps < .001$.

⌘ For the *continuous* animals, inter-point distances from the Apperceptive and Classic solutions were correlated ($r = .78$) and Apperceptive was correlated with LSA ($r = .26$), both $ps < .001$. Classic and LSA were not significantly correlated ($r = .06$).

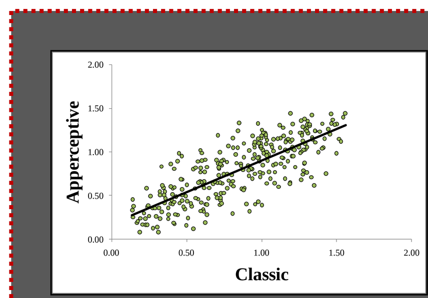
Categorical Animals



Correlation of Inter-Point Distances



Continuous Animals



Correlation of Inter-Point Distances

Conclusions

⌘ The Apperceptive method is efficient; on average, scaling 25-27 stimuli takes 5 minutes (compared to 20-25 minutes for Classic). With more stimuli, this disparity grows. Also, it obtains high resolution similarity estimates, and prevents interference by previously remembered responses.

⌘ For 2- and 3-D perceptual stimuli, the Apperceptive method produces solutions that are well-ordered and consistent with Classic methodology.

⌘ The method also works well with non-visual stimuli, both categorical and continuous in nature.

⌘ The Apperceptive method uses an intuitive interface, to take advantage of people's natural tendency to think about similarity spatially.

References

Data Theory Scaling Systems Group. (2006). Proxscal (version 1.0). Faculty of Social and Behavioral Sciences: Leiden University, the Netherlands.

Goldstone, R. L. (1994). An efficient method for obtaining similarity data. *Behavior Research Methods, Instruments, & Computers, 26,* 381-386.

Goldstone, R. L., & Medin, D. L. (1994). The time course of comparison. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 20,* 29-50.

Henley, N. M. (1969). A psychological study of the semantics of animal terms. *Journal of Verbal Learning and Verbal Behavior, 8,* 176-184.

Hornberger, M., Bell, B., Graham, K. S., & Rogers, T. T. (2009). Are judgements of semantic relatedness systematically impaired in Alzheimer's disease? *Neuropsychologia, 47,* 3084-3094.

Kruskal, J. B., & Wish, M. (1978). Multidimensional Scaling. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-011. Beverly Hills and London: Sage Publications.

Landauer, T., & Kintsch, W. (2003). *Latent Semantic Analysis.* Utilized October, 2009 from: http://lsa.colorado.edu.

Psychology Software Tools, Inc. (2006). E-Prime (version 1.2). Pittsburgh, Pa.

Shepard, R. N. (1980). Multidimensional Scaling, Tree-Fitting, and Clustering. *Science, 210,* 390-398.

For a copy of the software, email: michael.hout@asu.edu

1047