Journal of Experimental Psychology: General
2016, Vol. 145, No. 3, 376–382

© 2016 American Psychological Association
0096-3445/16/$12.00 http://dx.doi.org/10.1037/a0039758

COMMENT

Caveats for the Spatial Arrangement Method: Comment on Hout, Goldinger, and Ferguson (2013)

Steven Verheyen, Wouter Voorspoels, Wolf Vanpaemel, and Gert Storms
University of Leuven

The gold standard among proximity data collection methods for multidimensional scaling is the (dis)similarity rating of pairwise presented stimuli. A drawback of the pairwise method is its lengthy duration, which may cause participants to change their strategy over time, become fatigued, or disengage altogether. Hout, Goldinger, and Ferguson (2013) recently made a case for the Spatial Arrangement Method (SpAM) as an alternative to the pairwise method, arguing that it is faster and more engaging. SpAM invites participants to directly arrange stimuli on a computer screen such that the interstimuli distances are proportional to psychological proximity. Based on a reanalysis of the Hout et al. (2013), data we identify three caveats for SpAM. An investigation of the distributional characteristics of the SpAM proximity data reveals that the spatial nature of SpAM imposes structure on the data, invoking a bias against featural representations. Individual-differences scaling of the SpAM proximity data reveals that the two-dimensional nature of SpAM allows individuals to only communicate two dimensions of variation among stimuli properly, invoking a bias against high-dimensional scaling representations. Monte Carlo simulations indicate that in order to obtain reliable estimates of the group average, SpAM requires more individuals to be tested. We conclude with an overview of considerations that can inform the choice between SpAM and the pairwise method and offer suggestions on how to overcome their respective limitations.

Keywords: multidimensional scaling, similarity, reliability, dimensionality, pairwise rating

Supplemental materials: http://dx.doi.org/10.1037/a0039758.supp

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Many investigations in psychology are concerned with mental representation. Through the use of statistical procedures, researchers attempt to infer the latent structure that underlies participants' perception and cognition of stimulus materials. Multidimensional scaling (MDS; Borg & Groenen, 2005; Kruskal & Wish, 1978) is perhaps the best-known of these procedures. MDS conveys the structure in terms of one or more latent dimensions that organize the stimuli. MDS has the potential to reveal to researchers interested in mental representation the number, relative importance, and nature of latent dimensions, as well as the relative positions of individual stimuli along these dimensions. It achieves this by representing proximities between stimuli as distances between points in a multidimensional space. That is, MDS takes as its input indices of the psychological (dis)similarity between the stimuli (proximities for short), as most statistical procedures in this domain do. A variety of proximity data collection methods exist. The quintessence of this comment is that the choice for a particular proximity data collection method is not trivial. Each method has both strengths and limitations that one needs to consider in order to select the optimal method for one's research purposes. There is no one method that is to be preferred in every application.

The gold standard for obtaining proximity data is the pairwise method: having participants judge the (dis)similarity of all stimulus pairs via Likert scales. An important drawback of the pairwise method is the rapid increase of pairs with the number of stimuli. This makes for lengthy data collection protocols, which in turn may cause participants to change their rating strategy over time, become fatigued, or disengage altogether. As stimuli are presented in pairs, the extent and nature of the interstimuli differences may only become apparent after a considerable number of trials, rendering the early trials of the pairwise method less informative. An alternative method for gathering proximity data is the Spatial Arrangement Method (SpAM). SpAM requires participants to arrange stimuli on a computer screen such that the interstimuli distances are proportional to psychological proximity. The method can be conceived

Steven Verheyen, Wouter Voorspoels, Wolf Vanpaemel, and Gert Storms, Faculty of Psychology and Educational Sciences, University of Leuven.

Steven Verheyen and Wouter Voorspoels are both postdoctoral fellows at the Research Foundation–Flanders. We thank Michael Hout for generously providing the data from the target article. We are in debt to Alena Anishchanka for a number of insightful remarks about SpAM. We also thank Eef Ameel, Simon De Deyne, Farah Mutiasari Djalal, Tom Heyman, Loes Stukken, and Bram Van Rensbergen for helpful suggestions.

Correspondence concerning this article should be addressed to Steven Verheyen, Faculty of Psychology and Educational Sciences, University of Leuven, Tiensestraat 102, Box 3711, BE-3000 Leuven, Belgium. E-mail: steven.verheyen@ppw.kuleuven.be

of as inverse MDS: the physical distance between stimuli in a two-dimensional space is taken as a proxy of interstimulus dissimilarity. In a recent paper in *Journal of Experimental Psychology: General* Hout, Goldinger, and Ferguson (2013) make a case for the method by arguing that SpAM is free of the drawbacks the pairwise method has and is on even par with the pairwise method in other respects. Specifically, they argue that since in SpAM, changing the position of a single stimulus involves fixing its proximity to all the other stimuli simultaneously, the method is more efficient and thus faster than the pairwise method is. In SpAM, all stimuli are on the computer screen from the onset of the data collection protocol, and as a consequence, the extent and nature of the interstimuli differences are immediately clear to the participants. Finally, Hout et al. (2013) point out that SpAM is more engaging and user-friendly than the pairwise method is. The publication appears to have spurred adoption of the method, with more applications in the past year (Berman et al., 2014; Canini, Griffiths, Vanpaemel, & Kalish, 2014; Godwin, Hout, & Menneer, 2014; Hout, Goldinger, & Brady, 2014; Jenkins, Samuelson, Smith, & Spencer, 2015; Montez, Thompson, & Kello, 2015), than in the years before combined (Aronoff et al., 2006; Migo, Montaldi, & Mayes, 2013; Moelter et al., 2005; Ober & Shenaut, 1999).

In light of SpAM's potential for wide adoption due to its practical advantages and dissemination through *Journal of Experimental Psychology: General*, we note a number of caveats for the method. We argue that SpAM favors spatial over featural representations, is a crude instrument for measuring more than two dimensions, and yields less reliable average data than the pairwise method does. We provide evidence for each of these claims through a reanalysis of the data from Experiments 1 and 2 in Hout et al. (2013), which include four perceptual and two conceptual stimulus sets. The former stimuli are artificially constructed visual stimuli varying along two or three perceptual dimensions. The latter comprise existing verbal stimuli (animal names). The evidence in support of the caveats is summarized in nontechnical terms in the main text. We refer interested readers to the Supplemental Materials for details of the data and the analyses. In the General Discussion, we offer some recommendations for choosing between SpAM and the pairwise method, and we conclude with an overview of manners in which both methods can be adapted to overcome some of their respective shortcomings.

Caveat 1: SpAM Favors Spatial Over Featural Representations

Background

MDS is not the only means of representing proximity data. There exist alternatives to MDS that represent proximities in terms of trees with (binary) features rather than spaces with (continuous) dimensions (Van Mechelen & Storms, 1995). These alternative representations have proven their worth in various areas of psychology (e.g., Johnson & Tversky, 1984; Kemp & Tenenbaum, 2008; Shepard, 1980).

A positive skew (many small distances and few large ones; Sattath & Tversky, 1977) and low centrality (the absence of focal stimuli that are the nearest neighbors of many other stimuli; Tversky & Hutchinson, 1986) characterize spatial representations of the kind provided by MDS. In order to determine whether a spatial representation is appropriate of a particular set of stimuli, researchers therefore investigate to what extent proximity data present with these characteristics (e.g., Dry & Storms, 2009; Ghose, 1998; Giordano et al., 2011).

Claim

There is a close correspondence between the manner in which proximities are collected in SpAM and the way proximities are represented in MDS: in SpAM, stimuli are spatially arranged, and MDS represents proximities in a spatial manner. SpAM's spatial nature might impose structure on the resulting proximity data, biasing researchers toward a spatial means of representation.

Evidence

For a particular stimulus set, the participants in Hout et al. (2013) provided proximities either through SpAM or through the pairwise method. We computed the skewness and centrality of each participant's proximity data and conducted independent samples *t* tests of the hypothesis that the mean skewness and centrality of the SpAM and pairwise proximity data are different. The results, reported in Tables 1 and 2 in the Supplemental Materials, show that the characteristics of the proximities are affected by the manner in which they were obtained. The SpAM proximities tend to be more positively skewed and have lower centrality values than the pairwise proximities. The former difference is significant for three out of six stimulus sets, whereas the latter difference is significant for all six stimulus sets. The differences are most pronounced for the two conceptual stimulus sets.

The properties of the SpAM proximity data are found to be inherent to the spatial manner in which they were obtained. The SpAM proximity data present with positive skewness and low centrality values, in line with known distributional characteristics of distances obtained from spatial representations. Whereas the pairwise proximity data only display these properties for the perceptual stimulus sets and yield proximity data with a negative skew and high centrality for the conceptual stimulus sets, SpAM yields proximity data with a positive skew and low centrality independent of the stimulus sets (see Tables 1 and 2 in the Supplemental Materials). Although the distributional characteristics of the pairwise proximity data thus support a representational distinction between perceptual and conceptual stimuli, the SpAM proximity data do not. SpAM favors a spatial representation for both types of stimuli.

Implication

The constraints that SpAM's spatial nature places on proximity data can obscure interesting phenomena from a researcher. For instance, because of its low-dimensional spatial nature, SpAM does not allow participants to communicate that there are focal stimuli in a set (Tversky & Hutchinson, 1986). This in turn may lead to different conclusions altogether regarding the way to represent certain types of stimuli. The representation of conceptual versus perceptual stimuli is a case in point. The pairwise data from Hout et al. (2013) suggest that conceptual and perceptual stimuli

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

are best represented by featural and spatial representations, respectively (for similar claims, see Dry & Storms, 2009; Pruzansky, Tversky, & Carroll, 1982; Tversky & Hutchinson, 1986), but the SpAM data suggest both are best represented in a spatial manner. The pairwise method appears to allow a more diverse pattern of outcomes than SpAM does. Although the SpAM data consistently display the characteristics of spatial representations (irrespective of the stimuli), the characteristics of the pairwise data are determined by the stimuli, not the method. Until the kinds of stimuli for which the choice of method affects the results have been identified, the adoption of SpAM across varied areas of psychology is not recommended.

Caveat 2: SpAM Is a Crude Instrument for Measuring Many Dimensions

Background

When one decides to represent proximities in a spatial manner using MDS, the number of dimensions of the space always needs to be determined (Lee, 2001; Verheyen, Ameel, & Storms, 2007). In the rare cases where MDS is used to psychophysically scale known or obvious dimensions (e.g., Nosofsky, 1985, 1986), determining the dimensionality is trivial. Generally, however, researchers are unaware of the (number of) underlying dimensions and would like to infer them from the scaling. Establishing the number of dimensions is of particular importance in so-called individual-differences MDS (Carroll & Chang, 1970; Takane, 2007; Takane, Young, & De Leeuw, 1977), where the goal is to establish meaningful individual differences in the importance attached to every dimension and relate them to external information (such as group membership) one might have about the individuals.

Claim

The low-dimensional nature of SpAM might invoke a bias against high-dimensional representations. The number of dimensions of variation that can effectively be conveyed in a single two-dimensional arrangement is limited. When stimuli are more than two-dimensional, participants might decide to restrict themselves to the communication of the two most important dimensions or might come up with idiosyncratic ways to communicate the extra dimensions in their two-dimensional arrangement.

Evidence

Analysis of the three-dimensional perceptual data sets from Hout et al. (2013) with individual-differences MDS yielded three dimension weights and a badness-of-fit value for every participant who contributed proximities. The former values give an indication of the importance attached by every participant to each of the three manifest dimensions along which the stimuli are organized. The latter gives an indication of the extent to which a participant's proximities agree with this three-dimensional organization. Figures 1 and 2 in the Supplemental Materials depict the resulting dimensions weights for the two three-dimensional data sets available in Hout et al. (2013).

Levene's tests for equality of variance of the SpAM and pairwise dimension weights revealed that SpAM participants were more variable in the dimensions they communicated than the pairwise participants were. The null hypothesis that there is no difference in the variances of the dimension weights for the two methods was rejected for five of the 2×3 dimensions. Notably, for a considerable number of SpAM participants, the weight attached to one of the three dimensions was estimated to be close to 0, indicating that these participants conveyed only two dimensions in their arrangements. This was not the case for the pairwise participants. They invariably received positive dimension weight estimates, indicating that they employed all three dimensions when providing judgments of interstimulus dissimilarity.

SpAM participants with positive weights for all three dimensions conveyed the three dimensions suboptimally in their two-dimensional arrangements. Their badness-of-fit values were found to be significantly higher than those for the pairwise participants who also communicated three dimensions and those for the SpAM participants who communicated only two, for each of the two three-dimensional data sets.

Implication

SpAM's two-dimensional nature can make one underestimate the number of dimensions and/or wrongly assess the relative importance of the dimensions. By using SpAM, one might mistake differences that reflect the demands of the task for fundamental differences in proximity perception.

The requirements that have to be met for SpAM to produce a representative spatial representation of high-dimensional stimuli are very strong: Assuming that the participants share a high-dimensional representation, the different manners in which they collapse it into a two-dimensional spatial arrangement have to be systematically related to the original representation (i.e., reflecting the distinct importance of the various dimensions). The importance of salient dimensions is likely to be overestimated and the importance of less salient dimensions underestimated if participants can only communicate two dimensions. If there are marked saliency differences between dimensions, one will require many participants to ensure that a sufficient number of "odd" participants are included who communicate the less salient dimensions. Without such participants, the MDS solution will not reveal the nondominant dimensions.

The low-dimensional nature of SpAM might lead to unwarranted conclusions regarding the importance attached to different dimensions by different individuals. It might be a matter of fact that an individual used only two dimensions to arrange a particular set of stimuli, but this does not necessarily imply they were the only two that were available to the participant. The observation might just be a result of the method, which demands participants to select two dimensions along which to arrange the stimuli. Similarly, the observation that one participant uses dimensions X and Y for his or her arrangement, whereas another participant uses Y and Z, does not necessarily reflect a fundamental difference in perception, when X, Y, and Z are equally salient, but the task basically demands one to select two at random.

This is less of a concern with pairwise judgments where participants are not restricted to employ only two dimensions.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

That is not to say that pairwise participants will necessarily employ all available dimensions. As the number of dimensions becomes large, it might become impossible to take them all into account, and similar concerns might start to arise for the pairwise method as well.

Caveat 3: Burdens on the Reliability of Average SpAM Data

Background

It is common in the MDS literature to scale the average dissimilarity data of several individuals. The resulting representation is intended to be a good reflection of the proximity perception of the group of individuals as a whole, which then becomes subject to interpretation. The rationale behind this procedure is that differences between individuals are considered random error, and the averaging is intended to reduce this error. The reliability of the average is an indication of the success of the procedure. MDS representations of unreliable proximity data are not a good reflection of the shared structure among the participants (Ashby, Maddox, & Lee, 1994; Lee & Pope, 2003) and are known not to reproduce well (Sturidsson et al., 2006; Voorspoels et al., 2014; White, Voorspoels, Storms, & Verheyen, 2014).

Claim

SpAM's proclaimed efficiency rests primarily on the fact that fixing a single stimulus' position effectively establishes its proximity to all other stimuli. This should render the positioning of each stimulus a taxing problem, which seems at odds with the speed with which participants solve the task. SpAM participants might not give all proximities due consideration when positioning a stimulus, rendering some proximities less informative than others. This is less likely to be the case for the pairwise method where participants have to consider only two stimuli per trial.

SpAM is approached in a variety of ways as Hout et al. (2013) already indicated. Some participants might treat SpAM more like a discrete sorting task than as a continuous proximity measurement task. When participants choose to sort stimuli together in cluster-like groups, the within-cluster distances are not very informative. Conversely, when participants focus primarily on the local arrangement of stimuli within clusters, the between-cluster distances are rendered less meaningful. Other individual differences might arise because the low-dimensional spatial nature of SpAM forces participants to select two dimensions from several dimensions they feel are important, to convey additional dimensions in a manner that does not come natural to them (Caveat 2) or to convey perceived proximities in a format that does not match theirs (Caveat 1). Although the pairwise method is presumably not completely free of these influences, it seems fair to say that they apply to a lesser extent.

Inaccurate measurements and large individual differences affect the reliability of average proximity data negatively (Verheyen et al., 2015). Because individual indications of interstimuli proximity might be less informative in SpAM and SpAM is suspect to a variety of individual differences, estimates of the reliability of the average SpAM data may turn out lower than similar estimates for the pairwise method.

Evidence

For each of the stimulus sets from Hout et al. (2013), we obtained MDS solutions for both the individual and average proximity data. The correlations between the individual MDS distances and the group MDS distances were then subjected to independent samples t tests of the hypothesis that the correlations resulting from the pairwise method are greater than the correlations resulting from SpAM. Table 3 in the Supplemental Materials lists the results. For continuous stimulus sets and perceptual stimulus sets with more than two dimensions, the MDS representations of individual dissimilarity data corresponded significantly better with the MDS representation of the average dissimilarity data when the data were obtained with the pairwise method compared to SpAM. This suggests more pronounced individual differences among the data provided by the SpAM participants.

To quantify the reliability of the average proximity data resulting from a proximity data collection method, we randomly split the participants in two halves and calculated the correlation between the average proximities across each half. Table 4 in the Supplemental Materials lists the resulting values for the SpAM and pairwise data from Hout et al. (2013). Monte Carlo simulations, in which the number of participants providing proximity data was varied, showed that when the number of SpAM and pairwise participants is the same, the reliability of the average is significantly smaller for SpAM than for the pairwise method, for both continuous stimulus sets and perceptual stimulus sets with more than two dimensions. The simulations also indicated that across stimulus sets, about two times the number of SpAM participants are required to arrive at the reliability level of the pairwise method.

Implication

The above results indicate a trade-off between speed and reliability of the average. Spatially arranging stimuli on a screen according to similarity might be less time-consuming than judging all pairs of stimuli, but to achieve comparable reliability, more participants are needed. Following the law of large numbers, the reliability of the group average is to improve as more individuals are tested (Lord & Novick, 1968). When too few participants are tested, average SpAM data might not be representative of individual data and difficult to replicate.

A final caveat thus pertains to how the efficiency of proximity data collection methods should be assessed: in terms of time or number of participants required? Ideally, we would have an empirical indication of how speed and reliability trade off as a function of such parameters as the number of stimuli, the nature of the stimuli, and the number of dimensions. Practical considerations such as access to participants (or lack thereof) and the time available for data collection can then inform researchers about the preferred method.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

General Discussion

Whereas Hout et al. (2013) advocated the merits of SpAM compared to the pairwise method, we have identified a number of caveats for the method. This comment should not be mistaken for a plea to consistently use the pairwise method, however. We agree with Hout et al. (2013) that it is time-consuming, is often tedious, and holds the risk that participants change their rating strategy in the course of the task. We believe neither method is suited for adoption across the entire range of areas of investigation in psychology, and realizing the strengths and the weaknesses of both methods allows researchers to select the proximity data collection method that best serves their needs.

Recommendations

Following the investigation of Hout et al. (2013) and the one we presented here, we see merit in the use of SpAM for low-dimensional or factorial perceptual stimuli of the kind that are often used in experimental psychology, where psychophysical scaling of known or obvious dimensions is the goal. That is, those situations where the cognitive impressions of similarity are known to align with the low-dimensional spatial nature of SpAM. In our reanalyses of the Hout et al. (2013) data, we observed few differences between SpAM and the pairwise method for these stimuli, and SpAM is clearly the faster method. We advise that one restricts the MDS analyses of SpAM proximity data to the group level, provided the average data are sufficiently reliable. In situations where the cognitive abilities of the participants (children, patients) or the vast number of stimuli prohibit the use of the pairwise method, SpAM may be used as well, providing its limitations are recognized. For other applications, such as those involving conceptual or high-dimensional stimuli (for which the dimensions are generally not known by the investigator), we advise against the use of SpAM because of the constraints the method places on the proximity data. For such stimuli, we observed significant differences between the pairwise and the SpAM data in our reanalyses of the Hout et al. (2013) data.

Alternatives

Researchers who feel caught between a rock and a hard place might consider using variations on SpAM and the pairwise method that may take away some of their limitations. There exist, for instance, more efficient ways of gathering pairwise proximities. Many pairs provide redundant information (Young & Cliff, 1972). This can be capitalized on through the use of incomplete proximity tasks in which only a subset of pairs is presented to participants for rating. Because the location of a single point in an MDS representation is determined by $n - 1$ proximities (with n the number of stimuli), MDS is very robust against randomly distributed missing data (Graef & Spence, 1979; MacCallum, 1979; Malhotra, Jain, & Pinson, 1988; Robinson & Lissitz, 1977). One can leave out a substantial percentage of the pairs for rating and still recover the original representations quite well. This can constitute a considerable time gain when collecting proximities with the pairwise method. The use of a cyclic design (a particular kind of partially balanced incomplete block design; Burton, 2003; Clatworthy, 1973; Spence & Domoney, 1974) or interactive MDS (the real-time determination of a subset of stimuli that is critical for arriving at the dimensional representation; Baker & Young, 1975; Girard & Cliff, 1976) would probably speed up the data collection even more. To allow participants to employ the same strategy throughout the task instead of having to adjust it as they encounter more stimuli, Hout et al. (2013) propose using the total-set pairwise method, which involves presenting all the stimuli simultaneously and highlighting two of them on every trial for rating.

In order for participants to be able to convey more dimensions using SpAM, one could consider providing them with an interactive display that allows them to rotate items in perspective, so that they could visualize three-dimensional relations, or one could have them build multiple organizations, as in Goldstone (1994). It would also be interesting to see how SpAM fares when adaptive selection of stimulus subsets during measurement is introduced, as in Kriegeskorte and Mur (2012). The latter two suggestions involve the aggregation of different arrangements from the *same* participant rather than the aggregation of arrangements from *different* participants as we have been discussing in this comment. It remains to be seen how SpAM and the pairwise method compare under these circumstances.

Here, we have approached the comparison between SpAM and the pairwise method from the perspective of appropriateness for proximity data collection. An alternative approach altogether would be to develop more sophisticated psychological theories about how people transform their perceptions or knowledge of stimuli into the distinct answer formats of the tasks (e.g., Nosofsky, 1986, 1991).

References

Aronoff, J. M., Gonnerman, L. M., Almor, A., Arunachalam, S., Kempler, D., & Andersen, E. S. (2006). Information content versus relational knowledge: Semantic deficits in patients with Alzheimer's disease. *Neuropsychologia, 44,* 21–35. http://dx.doi.org/10.1016/j.neuropsychologia.2005.04.014

Ashby, F. G., Maddox, W. T., & Lee, M. D. (1994). On the dangers of averaging across subjects when using multidimensional scaling or the similarity-choice model. *Psychological Science, 5,* 144–151. http://dx.doi.org/10.1111/j.1467-9280.1994.tb00651.x

Baker, R. F., & Young, F. W. (1975). A note on an empirical evaluation of the ISIS procedure. *Psychometrika, 40,* 413–415. http://dx.doi.org/10.1007/BF02291766

Berman, M. G., Hout, M. C., Kardan, O., Hunter, M. R., Yourganov, G., Henderson, J. M., . . . Jonides, J. (2014). The perception of naturalness correlates with low-level visual features of environmental scenes. *PLoS ONE, 9,* e114572. http://dx.doi.org/10.1371/journal.pone.0114572

Borg, I., & Groenen, P. J. F. (2005). *Modern multidimensional scaling: Theory and applications*. New York, NY: Springer.

Burton, M. L. (2003). Too many questions? The uses of incomplete cyclic designs for paired comparisons. *Field Methods, 15,* 115–130. http://dx.doi.org/10.1177/1525822X03015002001

Canini, K. R., Griffiths, T. L., Vanpaemel, W., & Kalish, M. L. (2014). Revealing human inductive biases for category learning by simulating cultural transmission. *Psychonomic Bulletin & Review, 21,* 785–793. http://dx.doi.org/10.3758/s13423-013-0556-3

Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of Eckart-Young decomposition. *Psychometrika, 35,* 283–319. http://dx.doi.org/10.1007/BF02310791

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Clatworthy, W. H., Cameron, J. M., & Speckman, J. A. (1973). Tables of two-associate-class partially balanced designs. Washington, DC: National Bureau of Standard Applied Mathematics Series, 63.

Dry, M. J., & Storms, G. (2009). Similar but not the same: A comparison of the utility of directly rated and feature-based similarity measures for generating spatial models of conceptual data. *Behavior Research Methods, 41,* 889–900. http://dx.doi.org/10.3758/BRM.41.3.889

Ghose, S. (1998). Distance representations of consumer perceptions: Evaluating appropriateness by using diagnostics. *Journal of Marketing Research, 35,* 137–153. http://dx.doi.org/10.2307/3151843

Giordano, B. L., Guastavino, C., Murphy, E., Ogg, M., Smith, B. K., & McAdams, S. (2011). Comparison of methods for collecting and modeling dissimilarity data: Applications to complex sound stimuli. *Multivariate Behavioral Research, 46,* 779–811. http://dx.doi.org/10.1080/00273171.2011.606748

Girard, R. A., & Cliff, N. A. (1976). A Monte Carlo evaluation of interactive multidimensional scaling. *Psychometrika, 41,* 43–64. http://dx.doi.org/10.1007/BF02291697

Godwin, H. J., Hout, M. C., & Menneer, T. (2014). Visual similarity is stronger than semantic similarity in guiding visual search for numbers. *Psychonomic Bulletin & Review, 21,* 689–695. http://dx.doi.org/10.3758/s13423-013-0547-4

Goldstone, R. L. (1994). An efficient method for obtaining similarity data. *Behavior Research Methods, Instruments & Computers, 26,* 381–386. http://dx.doi.org/10.3758/BF03204653

Graef, J., & Spence, I. (1979). Using distance information in the design of large multidimensional scaling experiments. *Psychological Bulletin, 86,* 60–66. http://dx.doi.org/10.1037/0033-2909.86.1.60

Hout, M. C., Goldinger, S. D., & Brady, K. J. (2014). MM-MDS: A multidimensional scaling database with similarity ratings for 240 object categories from the Massive Memory picture database. *PLoS ONE, 9,* e112644. http://dx.doi.org/10.1371/journal.pone.0112644

Hout, M. C., Goldinger, S. D., & Ferguson, R. W. (2013). The versatility of SpAM: A fast, efficient, spatial method of data collection for multidimensional scaling. *Journal of Experimental Psychology: General, 142,* 256–281. http://dx.doi.org/10.1037/a0028860

Jenkins, G. W., Samuelson, L. K., Smith, J. R., & Spencer, J. P. (2015). Non-Bayesian noun generalization in 3- to 5-year-old children: Probing the role of prior knowledge in the suspicious coincidence effect. *Cognitive Science, 39,* 268–306. http://dx.doi.org/10.1111/cogs.12135

Johnson, E. J., & Tversky, A. (1984). Representations of perceptions of risks. *Journal of Experimental Psychology: General, 113,* 55–70. http://dx.doi.org/10.1037/0096-3445.113.1.55

Kemp, C., & Tenenbaum, J. B. (2008). The discovery of structural form. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 105,* 10687–10692. http://dx.doi.org/10.1073/pnas.0802631105

Kriegeskorte, N., & Mur, M. (2012). Inverse MDS: Inferring dissimilarity structure from multiple item arrangements. *Frontiers in Psychology, 3,* 245. http://dx.doi.org/10.3389/fpsyg.2012.00245

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling*. Beverly Hills, CA: Sage.

Lee, M. D. (2001). Determining the dimensionality of multidimensional scaling representations for cognitive modeling. *Journal of Mathematical Psychology, 45,* 149–166. http://dx.doi.org/10.1006/jmps.1999.1300

Lee, M. D., & Pope, K. J. (2003). Avoiding the dangers of averaging across subjects when using multidimensional scaling. *Journal of Mathematical Psychology, 47,* 32–46. http://dx.doi.org/10.1016/S0022-2496(02)00019-6

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison Wesley.

MacCallum, W. (1979). Recovery of structure in incomplete data by ALSCAL. *Psychometrika, 44,* 69–74. http://dx.doi.org/10.1007/BF02293785

Malhotra, N. K., Jain, A. K., & Pinson, C. (1988). The robustness of MDS configurations in the case of incomplete data. *Journal of Marketing Research, 25,* 95–102. http://dx.doi.org/10.2307/3172929

Migo, E. M., Montaldi, D., & Mayes, A. R. (2013). A visual object stimulus database with standardized similarity information. *Behavior Research Methods, 45,* 344–354. http://dx.doi.org/10.3758/s13428-012-0255-4

Moelter, S. T., Hill, S. K., Hughett, P., Gur, R. C., Gur, R. E., & Ragland, J. D. (2005). Organization of semantic category exemplars in schizophrenia. *Schizophrenia Research, 78,* 209–217. http://dx.doi.org/10.1016/j.schres.2005.06.011

Montez, P., Thompson, G., & Kello, C. T. (2015). The role of semantic clustering in optimal memory foraging. *Cognitive Science*. Advance online publication. http://dx.doi.org/10.1111/cogs.12249

Nosofsky, R. (1985). Overall similarity and the identification of separable-dimension stimuli: A choice model analysis. *Perception & Psychophysics, 38,* 415–432. http://dx.doi.org/10.3758/BF03207172

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115,* 39–57. http://dx.doi.org/10.1037/0096-3445.115.1.39

Nosofsky, R. M. (1991). Stimulus bias, asymmetric similarity, and classification. *Cognitive Psychology, 23,* 94–140. http://dx.doi.org/10.1016/0010-0285(91)90004-8

Ober, B. A., & Shenaut, G. K. (1999). Well-organized conceptual domains in Alzheimer's disease. *Journal of the International Neuropsychological Society, 5,* 676–684. http://dx.doi.org/10.1017/S1355617799577102

Pruzansky, S., Tversky, A., & Carroll, J. D. (1982). Spatial versus tree representations of proximity data. *Psychometrika, 47,* 3–24. http://dx.doi.org/10.1007/BF02293848

Robinson, E. J., & Lissitz, R. W. (1977). The approximation of a group stimulus space by averaging responses to selected subsets of the stimuli. *Psychometrika, 42,* 447–450. http://dx.doi.org/10.1007/BF02293663

Sattath, S., & Tversky, A. (1977). Additive similarity trees. *Psychometrika, 42,* 319–345. http://dx.doi.org/10.1007/BF02293654

Shepard, R. N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science, 210,* 390–398. http://dx.doi.org/10.1126/science.210.4468.390

Spence, I., & Domoney, D. W. (1974). Single subject incomplete designs for nonmetric multidimensional scaling. *Psychometrika, 39,* 469–490. http://dx.doi.org/10.1007/BF02291669

Sturidsson, K., Långström, N., Grann, M., Sjöstedt, G., Åsgård, U., & Aghede, E. M. (2006). Using multidimensional scaling for the analysis of sexual offence behaviour: A replication and some cautionary notes. *Psychology, Crime & Law, 12,* 221–230. http://dx.doi.org/10.1080/10683160500126227

Takane, Y. (2007). Applications of multidimensional scaling in psychometrics. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics: Psychometrics* (Vol. 26, pp. 359–400). Amsterdam, the Netherlands: Elsevier.

Takane, Y., Young, F., & De Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika, 42,* 7–67. http://dx.doi.org/10.1007/BF02293745

Tversky, A., & Hutchinson, W. (1986). Nearest neighbor analysis of psychological spaces. *Psychological Review, 93,* 3–22. http://dx.doi.org/10.1037/0033-295X.93.1.3

Van Mechelen, I., & Storms, G. (1995). Analysis of similarity data and Tversky's contrast model. *Psychologica Belgica, 35,* 85–102.

Verheyen, S., Ameel, E., & Storms, G. (2007). Determining the dimensionality in spatial representations of semantic concepts. *Behavior Research Methods, 39,* 427–438. http://dx.doi.org/10.3758/BF03193012

Verheyen, S., Voorspoels, W., Longenecker, J., Weinberger, D. R., Elvevåg, B., & Storms, G. (2015). Invalid assumptions in clustering analyses of category fluency data: Reply to Sung, Gordon and Schretlen

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

(2015). *Cortex*. Advance online publication. http://dx.doi.org/10.1016/j.cortex.2015.05.012

Voorspoels, W., Storms, G., Longenecker, J., Verheyen, S., Weinberger, D. R., & Elvevåg, B. (2014). Deriving semantic structure from category fluency: Clustering techniques and their pitfalls. *Cortex, 55,* 130–147. http://dx.doi.org/10.1016/j.cortex.2013.09.006

White, A., Voorspoels, W., Storms, G., & Verheyen, S. (2014). Problems of reliability and validity with similarity derived from category fluency. *Psychiatry Research, 220,* 1125–1130. http://dx.doi.org/10.1016/j.psychres.2014.10.001

Young, F. W., & Cliff, N. (1972). Interactive scaling with individual subjects. http://dx.doi.org/10.1007/BF02291217. *Psychometrika, 37,* 385–415.

Received June 28, 2013
Revision received August 7, 2015
Accepted August 7, 2015 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Psychological Bulletin Call for Papers: Replication and Reproducibility: Questions Asked and Answered via Research Synthesis

The primary mission of *Psychological Bulletin* is to contribute a cohesive, authoritative, theory-based, and complete synthesis of scientific evidence in the field of psychology.

The editorial team is currently interested in the contribution of meta-analysis and systematic reviews to answering questions about the replication and reproducibility of psychological findings. This call is designed to encourage authors to consider the degree to which multi-study primary research papers, systematic replications efforts (e.g., Open Science Framework and Many Labs Replications), and meta-analysis can contribute to substantial psychological knowledge.

The editors of *Psychological Bulletin* announce interest in publishing Replication and Reproducibility: Questions Asked and Answered via Research Synthesis. Papers may be published as a special issue or as a series of special sections. Manuscripts will be evaluated and may be published as they come in but must be **submitted by June 30, 2017** for consideration for the special sections/issue.

We invite authors to submit single papers that address a clearly defined question related to replication and reproducibility of psychological findings using systematic-review or meta-analytic perspectives. Purely methodological papers, such as tests or criticisms of particular approaches, or commentaries on replication and reproducibility, are not appropriate for *Psychological Bulletin*. Primary-level research reports are also outside of the interest domain of *Psychological Bulletin*. Examples of possible manuscripts include but are not limited to meta-analyzing replication efforts, comparing evidence from different replication methods, synthesizing models of replication determinants, meta-analyzing predictors of questionable research practices and retraction, and comparing methods to gauge publication bias.

For guidance in proposing a submission, interested authors may contact the Editor, Dolores Albarracín (dalbarra@illinois.edu), or Associate Editor, Blair T. Johnson (blair.t.johnson@uconn.edu). Alternatively, authors are free to submit unsolicited manuscripts and should indicate that the submission targets this call in the cover letter.