Multidimensional scaling

Michael C. Hout[1]*, Megan H. Papesh[2] and Stephen D. Goldinger[1]

The concept of similarity, or a sense of 'sameness' among things, is pivotal to theories in the cognitive sciences and beyond. Similarity, however, is a difficult thing to measure. Multidimensional scaling (MDS) is a tool by which researchers can obtain quantitative estimates of similarity among groups of items. More formally, MDS refers to a set of statistical techniques that are used to reduce the complexity of a data set, permitting visual appreciation of the underlying relational structures contained therein. The current paper provides an overview of MDS. We discuss key aspects of performing this technique, such as methods that can be used to collect similarity estimates, analytic techniques for treating proximity data, and various concerns regarding interpretation of the MDS output. MDS analyses of two novel data sets are also included, highlighting in step-by-step fashion how MDS is performed, and key issues that may arise during analysis. © 2012 John Wiley & Sons, Ltd.

INTRODUCTION

Similarity is a pervasive concept, one that is central to understanding and engaging in the world around us. A sense of 'sameness' between two things allows us to appropriately generalize and discriminate: We can detect when two situations (or stimuli) are similar enough to be acted upon as the same, and when they are dissimilar enough to require different actions.[1–2] For example, a basset hound and a dachshund will both chase a thrown toy, but a cat will respond by staring. Brake lights on a semi-truck and a sedan may be visually dissimilar, but both signal stopping behavior. Conversely, the sedan's brake lights and turn signal may look comparatively more alike, but signal very different actions. As a theoretical construct, similarity is ubiquitous, with broad implications in fields such as memory,[3,4] language,[5,6] categorization,[7–9] and perception,[10,11] to name but a few. As researchers, it is often of benefit to have not just a sense, but a quantitative estimate of the similarity between two items (or among a larger set of items). A language scientist may wish

to know precisely how alike two different phonemes are perceived, and a music researcher may seek to estimate the degree to which Mozart is more like Beethoven than he is to Chopin, Vivaldi, or Bach. But similarity is difficult to quantify. Psychological measurements may be imprecise or 'noisy', and perceived similarity relations change dramatically given different environments for comparison.[12,13] For example, a rabbit and a salmon may seem fairly dissimilar upon context-free consideration, but they appear more alike when both are featured on a chef's tasting menu.

Multidimensional scaling (MDS) is a tool by which to quantify similarity judgments. Formally, MDS refers to a set of statistical procedures used for exploratory data analysis and dimension reduction.[14–21] It takes as input estimates of similarity among a group of items; these may be overt ratings, or various 'indirect' measurements (e.g., perceptual confusions), and the stimuli may be perceptual or conceptual in nature. The outcome of MDS is a 'map' that conveys, spatially, the relationships among items, wherein similar items are located proximal to one another, and dissimilar items are located proportionately further apart. From this map, one may infer the underlying dimensions of a data set (or confirm prior hypotheses) by subjectively examining the organization of the space.[22] For example, if one were to apply MDS to pairwise similarity ratings among a set of color patches (e.g., 'On a scale from

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: michael.hout@asu.edu

[1]Department of Psychology, Arizona State University, Tempe, AZ, USA

[2]Department of Psychology, Louisiana State University, Baton Rouge, LA, USA

1 to 9, how similar are these two colors?'), the outcome would likely resemble Newton's famed 'color wheel',[23] with blue being located close to purple, but far from yellow, and so on. MDS maps are valuable because they reduce potentially complex data sets to the primary dimensions along which the items differ, and because they permit visual appreciation of the relationships that are present. Moreover, by assessing the distance between items on the map, one can obtain a quantitative measurement of their perceived similarity, relative to other items in the space. The applicability of MDS is broad, with potential utility across many disciplines, such as cognitive,[24] social,[25] developmental,[26] and clinical psychology,[27] as well as psychophysics,[28] neuroscience,[29] marketing,[30] political science,[31] sociology,[32] ecology,[33] and others. The current article provides a tutorial for performing MDS (using two novel data sets for a step-by-step demonstration), and discusses important practical considerations for those with interest in this technique.

THE DATA: FORM AND METHODS OF COLLECTION

To conduct MDS, a 'proximity matrix' is required; that is, a collection of similarity estimates between each pair of items in the stimulus set. For a set composed of k items, $(k * (k - 1))/2$ proximities must be acquired, such that each item is compared to every other at least once. This means that the number of comparisons grows rapidly as a function of the stimulus set size. For a set of 10 items, 45 comparisons must be made. For sets of 20, 40, and 80, the comparisons grow to 190, 780, and 3160 (respectively), and so on. With large set sizes, it becomes impractical to collect a completed matrix from each person, so data may be concatenated across people to form a single, aggregate matrix. Ideally, the complete matrix will be collected from multiple participants,[34] so that each pair of items can be rated several times (to safeguard against measurement noise). Commonly, the data are organized into a half-triangular (or square-symmetric) matrix, with the names of each of the stimuli ordered across rows and columns. The similarity rating for each pair of items is then placed at the intersection of the appropriate row and column (although some software packages provide an option to compute proximities from raw data). This matrix (or matrices, for multiple participants) is then analyzed in statistical software packages, such SPSS.[35]

Proximity data may take the form of similarities or dissimilarities (e.g., 'How alike/different are these two items?'), and can be collected in a variety of ways. Broadly, methods for data collection can be categorized as *direct* or *indirect* (see Ref 36 for a review). In direct methods, people knowingly assess the items. The simplest technique is to provide people with two items at a time, and ask them to rate how similar they are to one another, using a Likert scale or a slide bar. People do this many times—at least once for every pairwise combination of items—and the mean ratings that then comprise a proximity matrix for each participant. Alternative direct methods have people categorize items according to some criteria,[37] or sort them into 'piles' based on their similarity.[38] Proximities are then calculated by counting how often items are sorted into the same category across participants. Higher degrees of similarity are indicated by items being categorized together more often.

In indirect methods, proximities are derived from secondary empirical sources. For instance, stimulus confusions or generalizations are commonly obtained from speeded *same–different* judgment tasks. Here, proximities can be estimated by identifying the proportion of times that items are mistakenly identified as the same,[39] or by the speed of accurate discriminations.[40] Higher degrees of similarity are indicated by higher rates of confusion, or slower discrimination times.

A DEMONSTRATION OF MDS

To provide clarity, we demonstrate the process of performing MDS, using data from two new experiments. Volunteers rated the similarity of either a set of perceptual or conceptual stimuli: Our perceptual items were 16 door-knockers, presented as pictures (obtained from the 'Massive Memory' database[41]). The conceptual stimuli were 12 different crimes, presented as words. There were 22 and 26 participants in each experiment, respectively.

The data were obtained using a *spatial arrangement method*, originally proposed by Goldstone[42] (see also Ref 43,44). This technique involves presenting all stimuli to the participant at once, randomly arranged on a computer screen. The participant uses the computer mouse to arrange the items (which are manipulated using simple drag-and-drop) on the screen in such a way that all inter-item distances reflect the participant's perception of object similarity. Items that are considered similar are placed closer together; those considered dissimilar are placed proportionately further apart. Conceptually, it is as if participants project their own 'psychological spaces' onto a two-dimensional plane (i.e., the computer monitor). After the organization is complete, a proximity matrix is derived from the location of the items on the screen.

Each proximity value for a pair of stimuli is measured by taking the item-to-item Euclidean distance, measured in pixels. The results of these experiments are presented below, following a discussion of the analytic process for MDS.

ANALYSIS AND INTERPRETATION OF THE SPACE

MDS data can be analyzed in various ways, using any of the numerous instantiations of scaling algorithms, such as *PROXSCAL*,[45] or *ALSCAL*.[46] Each of these algorithms treat the proximity data in a subtly different way, and some—such as *INDSCAL*[47] and *PREFSCAL*[48]—achieve secondary purposes, as well. What is shared among them is their primary goal: They seek to create geometric representations of the data, wherein the distances between each pair of items respects (as well as possible) the proximity data. That is, points in the output space should faithfully reflect the rated similarity between each pair of items.

To visualize how this is achieved, imagine that you possess a table with the distances among four cities: Los Angeles, New York, Chicago, and Dallas (see Refs 36,49 for a more thorough treatment of this example). This is a proximity matrix, telling you how far each city is from every other (it is approximately 2500 miles from Los Angeles to New York, 1500 miles from Dallas to New York, and so on). Using this set of proximities, an MDS algorithm will attempt to spatially recreate the original configuration of the objects. For someone familiar with American geography, this is fairly simple: One may start by placing Los Angeles on the far left of the map, and New York on the right. But scaling algorithms do not possess information about where to start placing the items. In fact, many begin with completely random starting configurations. Los Angeles may start somewhere near the center of the map, for example. The process of 'moving' the points is an iterative procedure. In each iteration, the distance between each pair of items in the space is compared to their intended distance specified by the proximity matrix. Going back to our example, assume that Dallas, Chicago, and New York are all where they should be. If Los Angeles is in the center of the map, then it is too close to New York, and must be moved farther away, but in what direction? Los Angeles could be placed 2500 miles from New York simply by moving it to the top of the map, but this would also change its relationship to other cities in the space: It would still be closer to Chicago than it should be, and even farther from Dallas than is appropriate. Thus, the challenge of 'placing' the items is in finding locations that appreciate all individual relationships among the points (it is not surprising, therefore, that this process often requires many 'moves').

Of course, in this example, a 'true' solution exists wherein the distances between each pair of cities perfectly matches the proximity data. But with experimental measurements, this is rarely the case. Psychological estimates are often noisy or imprecise. Participants may not agree with one another, and at times, a participant's own ratings may be in conflict (see Ref 50 for a review). Suppose someone is using a Likert scale to rate the similarity of three items: A, B, and C. They rate the A–B pair as 1 (*most similar*), and the A–C pair as 9 (*least similar*). Assuming the points are plotted on a two-dimensional plane, the B–C rating should therefore also be highly dissimilar (i.e., B is like A, which is unlike C). But, because participants usually only see two items at a time, they may not consider their earlier ratings, and may therefore create a conflict, such that the relationships among all the items are not perfectly accommodated. To quantify the amount of conflict that is present in the data, a *stress function* is calculated (the particular function varies across different scaling algorithms), which measures the agreement between the estimated distances and the input proximities. Lower stress values indicate a better fit; thus, the algorithms attempt to increase the fidelity to the input data by minimizing this stress function. The iterative process (i.e., moving the points in space) is repeated until the configuration has reached some optimization criterion. Again, the criteria vary across algorithms, and can often be modified by the researcher: For example, the iterations may cease once the stress has reached a pre-determined value, once it ceases to change significantly across iterations, or once a specific number of iterations have been completed. Finally, because a 'true' solution rarely exists, each scaling attempt may result in a different configuration of points. A good solution is stable, such that the relationships among points are consistent across scaling attempts.

MODEL SELECTION

Selection of the specific MDS 'model' to be used can be complex; it should be tailored to the particular circumstances under which the proximity data were collected, and to the researcher's analytic goals. (A full discussion of model selection is beyond the scope of this brief article, but see Refs 34,37,49 for thorough treatment.) The models vary in many ways, including (but not limited to) the geometry used to map the data, the calculations of—and algorithms used to optimize—model 'fit', and the ability to

handle singular or multiple proximity matrices. At a gross level, a distinction is made between models that implement *metric* versus *nonmetric* algorithms for scaling the original proximities (similarity or dissimilarity data) into the 'disparities' that are used for comparison to the distances derived from the analysis.[34] Metric MDS uses a linear function to map the proximities onto disparities,[51] whereas nonmetric models use any positive monotonic function, without the constraint of linearity.[18–19] For quantitative data (i.e., interval or ratio level values), metric MDS is used, but for qualitative data (i.e., ordinal level values), nonmetric MDS is performed.

To be more precise, models can be identified by their specific properties and sub-goals. *Classical* MDS uses a single matrix of raw (or averaged) data and is not especially robust. By comparison, *replicated* MDS accommodates multiple matrices; it is more robust due to using increased data. When the analyst is interested in between-subjects differences in judgments, *individual differences scaling* (or *weighted* MDS) can be implemented. This model treats data from multiple sources, and produces a secondary output in addition to the standard *common space plot* (i.e., the aggregate MDS plot using all data). This secondary output indicates the degree to which each participant (assuming each matrix represents a different person) weighted the dimensions that are revealed in the common space. For example, Janal et al.[52] applied electric stimulation to the skin of healthy volunteers who judged the similarity between 16 stimulations (which varied only in intensity), and also rated each on 16 different 'property scales' (e.g., 'On a scale from 0 to 9, how *brief* or *lasting* was the stimulus?'). Individual differences scaling suggested that the first two dimensions of the common space were *sensory magnitude* (e.g., slight sensation, strong sensation), and *pain intensity* (e.g., faint pain, intolerable), respectively. Moreover, the degrees to which participants weighted these dimensions mapped onto their choices of stimulus descriptors: For instance, participants with high weighting on the pain dimension frequently described the stimuli as 'cruel' or 'piercing'.

Another scaling approach that respects individual differences is *multidimensional unfolding*. This procedure is similar to individual differences scaling, but assumes that each observer has a different 'vantage point'. For example, Roskam[53] had 39 faculties in the University of Nijmegen each rank how relevant their nine shared research areas were for their own work. Unfolding techniques allow a common space to be derived, while remaining sensitive to wide individual variations in rank-orders. Papesh and Goldinger[10] used an unfolding technique

to assess similarity judgments of computerized faces in an experiment wherein different participants viewed different subsets of faces, and therefore had different perspectives on the full set. The technique is not commonly used in cognitive science, however, as most researchers assume normality in their sampled participants and objects, and usually obtain full (or randomly chosen subsets of) data from each observer. Several other techniques may be classified under the umbrella term of 'multidimensional scaling', such as *Procrustes' analysis* (which enables comparison of two configurations that were derived from different methods), *unidimensional scaling* (which configures the points along a single dimension), and *correspondence analysis* (which represents categorical data in a Euclidean space). This is not a comprehensive list of MDS models, but this sampling highlights an important point: Namely, that MDS procedures vary widely, and can be used to provide more than simple data visualization.

Finally, there are close correspondences (both mathematically and conceptually) between MDS and other methods used to reduce the dimensionality of complex data. *Principal components analysis* (PCA), for example, is a widely used method that achieves similar results. PCA is a technique that seeks orthogonal, linear fits to complex data, such that maximum variance is explained (with the first principal component explaining the most variance, followed in order by the second, third, etc.). The underlying approach of PCA is mathematically identical to metric MDS based on Euclidean distance (see Ref 53 for extensive discussion; also *related articles*). By comparison, nonmetric MDS is better able to preserve inter-point distances in a final configuration. In essence, the difference can best be described in terms of the research goals: PCA is more focused on the dimensions themselves, and fitting the variance as closely as possible, whereas MDS is more focused on relations among the scaled objects. In nonmetric MDS, we seek to discover the best orthogonal dimensions that describe the data, without losing too much information about point-by-point relations. For certain research purposes, such information may constitute mere noise, and PCA may be preferred. Also closely related to PCA is *exploratory factor analysis*; like MDS, factor analysis does not necessarily seek to maximize explained variance, but is intended to discover causal factors that organize the data.

INTERPRETING THE SOLUTION

Returning to the American cities example, interpretation of the output map is simple, because we already know the underlying solution. Some are located farther south, others farther east: The dimensions of

the map are clearly latitude and longitude. However, psychological data are rarely so straightforward, and interpretation of MDS solutions is subjective. The analyst must examine the organization of the space and attempt to infer the primary dimensions by which similarity estimates were given (or decides whether or not *a priori* hypotheses are confirmed). For instance, an MDS space for a group of animals may show dogs located close to cats, with both located far from wolves and tigers. From this, the analyst may infer that *domesticity* was a dimension by which the animals were rated. Alternatively, dogs and wolves may be close in space, both located far from cats and tigers. In this case, the analyst would not infer a dimension of domesticity, but perhaps canine versus feline.

A central point to remember is that MDS spaces convey information about relationships, not about particulars. First, the layout of the dimensions is unimportant. In the cities example, an MDS treatment might flip the dimensions, with east on the left or north on the bottom, but the relative positions among the cities should be consistent. Thus the interpretation of the dimensions (latitude, longitude) holds, regardless of the particular placement of the items in the space. Second, the units provided in an MDS space are arbitrary. Dogs may be located one unit away from cats, and four units away from wolves. But those units tell us little about perceived similarity without the context of the larger set. Similarity, as a construct, does not come with standard units of measurement. Rather, by inspecting the space, one can infer that dogs are perceived to be four times more similar to cats than they are to wolves.

CHOOSING THE RIGHT DIMENSIONALITY

When MDS is performed, the analyst must specify how many dimensions the algorithm should use when locating the points in space (i.e., how many coordinate values). By increasing the dimensions, more degrees of freedom are added to the movement of individual points. This has two consequences. First, it increases the information represented by the solution. For instance, if dogs, cats, wolves, and tigers were plotted on a line, one could only appreciate the animals' likeness based on domesticity or feline/canine nature (based on whichever emerged as the 'primary' dimension). If, however, a second dimension was added to the space, one could appreciate both aspects that were used for comparison. Second, because increasing the dimensionality adds degrees of freedom, it also decreases the stress of the solution. The lowest stress value will, of course, result from a space with as

many dimensions as there are items in the set. But the goal of conducting MDS is not to reduce the stress to zero, but rather to strike a balance between finding a good solution (i.e., one with a sufficiently low stress value) and one that is interpretable. Beyond three dimensions, interpretation of a space can become very difficult, so analysts are often conservative in choosing a dimensionality, such that visual examination of the data remains possible.

Scree plots are often used to determine how many dimensions are appropriate. To create a Scree plot, analysts scale the data several times (with higher dimensionality each time), and plot the stress values as a function of dimensions. A common heuristic is to look for an 'elbow' in the plot, which is the value at which added dimensions no longer improve the fit substantially.[36] Sometimes, an elbow is not immediately apparent. In these situations, analysts must rely on prior hypotheses about the likely dimensions of the space, and must consider how important visualization is to their goals. For instance, if visualization is relatively unimportant, then one may decide to plot the data in a high dimensionality (e.g., the point at which stress values plateau). In contrast, if the analysis is purely exploratory, a two- or three-dimensional plot may be more useful, as it allows simple visual inspection of the space. Recently, Bayesian approaches to dimension selection have been developed[54,55]; these take a more empirical approach by mathematically addressing the trade-off between the fit of the data and the complexity of the solution.

A MDS ANALYSIS OF DOOR-KNOCKERS AND CRIMES

To provide a concrete example of MDS analysis, and to highlight several of the issues already discussed, we now return to the experiments with crimes and door-knockers. (For interested readers, we provide the data and a step-by-step tutorial on Michael Hout's Web site (www.michaelhout.com). First, we scaled each data set six times, with increasing dimensionality, using the *ALSCAL* scaling algorithm provided in SPSS 20. This algorithm minimizes a stress function called Kruskal's S-Stress (specifically, the Takane-Young-de Leeuw formula[56];), and it provides an R^2 measure that indexes the proportion of variance of the scaled data explained by the fitted distances. The iterations of this algorithm terminate once they have converged on a layout of points with a stress value that improves by less than 0.001, relative to the previous iteration. Figure 1 shows the Scree plots for each data set, with similar plots of the R^2 values. For both data sets, stress values decrease as dimensionality increases, and
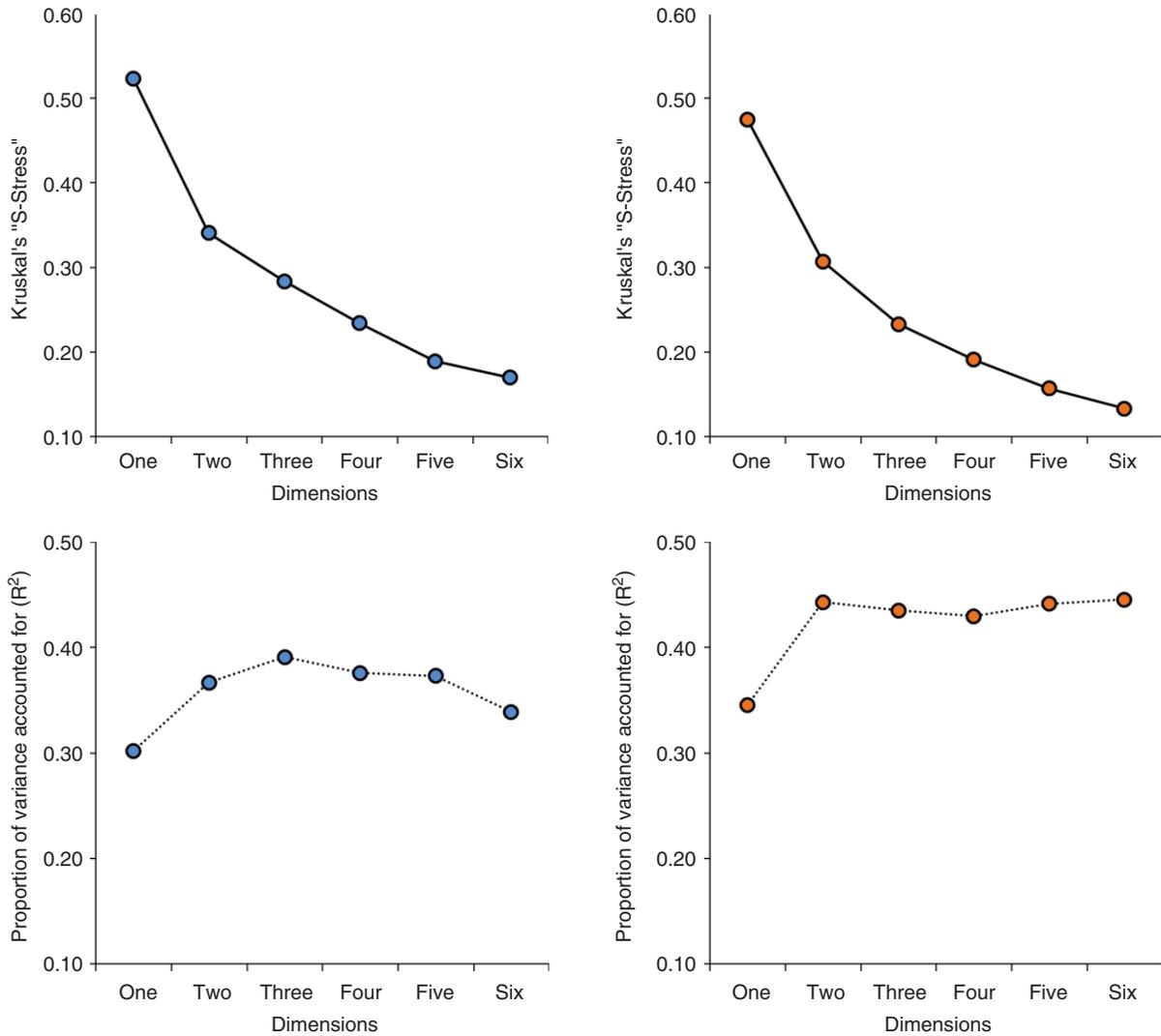
FIGURE 1 | Stress values (top panels) and explained variance (bottom panels), plotted as a function of dimensions used to locate the points in space, from door-knockers (left panels) and crimes (right panels), scaled using the *ALSCAL* algorithm.

R^2 values tend to rise. Note that these measures do not necessarily improve with each added dimension, however, because with higher dimensionality the algorithms may attempt to accommodate noise in the measurements (rather than true features, *per se*). The Scree plot for the door-knockers shows a likely elbow at dimension 2, but stress continues to fall as more dimensions are added to the solution.

For demonstration purposes, we plotted the door-knocker solutions on a two-dimensional plane. The left panel of Figure 2 shows the results of the *ALSCAL* analysis; the right panel shows a similar analysis, derived using the *PROXSCAL* scaling algorithm (also in SPSS 20). The organizations of the spaces are largely consistent with one another, although they are not in perfect agreement (e.g., the first solution shows tighter 'clustering' of the

objects; the second more evenly distributes them in space). Both solutions show (on the *x*-axes) a clear distinction between door-knockers with faces, and those without. There are also (on the *y*-axes) color dimensions, showing that metallic colors (silver, gray) tend to be removed from the softer colors (gold, bronze). Thus, although the particular inter-item distances differ across the two scaling attempts, the overall arrangement of the spaces is largely consistent.

One way to quantify the agreement across solutions is to correlate the inter-item distances from one solution to another. We therefore collected vectors of distances from each point to every other in each space, and calculated their correlation. To the extent that the spaces have similar arrangements of points, the correlation will be strongly positive.

FIGURE 2 | Two-dimensional MDS plots for the door-knocker stimuli. The left panel shows a solution derived using the *ALSCAL* scaling algorithm, and the right panel shows the same data set analyzed using the *PROXSCAL* scaling algorithm.

We conducted this analysis for each data set (door-knockers and crimes), comparing the coordinates derived from *ALSCAL* and *PROXSCAL*. Figure 3 shows the results: The correlations were strongly positive ($R^2 = 0.67$ and 0.54 for door-knockers and crimes, respectively). This technique can also be applied to repeated attempts of a single scaling algorithm (recall that each attempt results in a different configuration) to index the stability of a solution (see Ref 43).

For the crime data, we scaled the solutions in three-dimensions, shown in Figure 4. In this case, interpretation of the space is less straightforward, due to the conceptual nature of the stimuli. Dimension 1 is perhaps the clearest. Highly violent crimes (e.g., rape, murder) tend to cluster on one side of the space, far from less violent crimes (e.g., speeding,

jaywalking). Thus, we may infer a primary dimension of violence. However, this interpretation overlaps with other feasible explanations. The layout of crimes across dimension 1 could also reflect 'interpersonal interaction', because the violent crimes necessarily require personal interaction, whereas the 'victimless' crimes do not. Interpreting dimensions 2 and 3 is even more challenging. One may reflect the extent to which money factors into the crimes, as those committed for financial gain (e.g., forgery, fraud, kidnapping) tend to be located near each other, as are those that do not involve money (e.g., DUI, speeding). Moving away from dimensional interpretations (i.e., the extent to which points fall along continua), it is also useful to examine the space for 'clusters'. For instance, violent crimes tend to be grouped together, as are the 'financial' crimes, and the driving violations. Again,
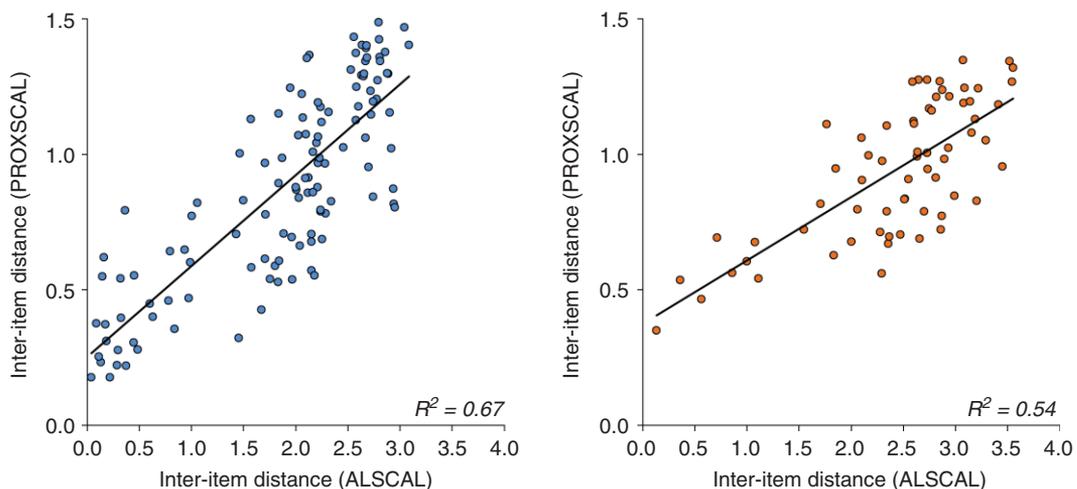
FIGURE 3 | Inter-item distance correlations from door-knocker (left panel) and crime (right panel) stimuli. The correlations show the agreement between the organizations of the spaces derived using *ALSCAL* and *PROXSCAL*.
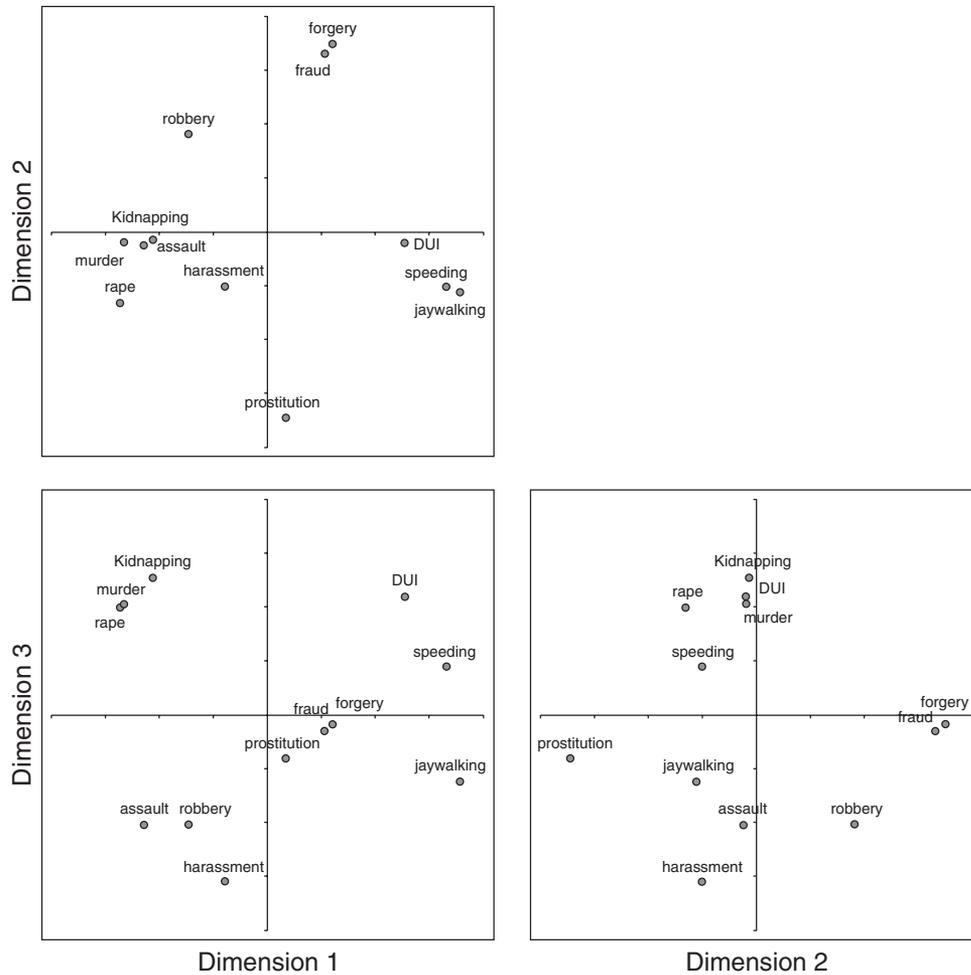
FIGURE 4 | Three-dimensional outcome of the crime data (scaled using *ALSCAL*), shown as three separate two-dimensional plots.

interpretation of the space is subjective and, although there may not be a 'correct' interpretation, inspection of the solutions provides the analyst with a wealth of information regarding the ways people perceive the similarity of these crimes.

Indeed, it should be noted that MDS may not be the optimal avenue for representing the similarity space of crimes. There are other methods for obtaining similarity structures that are not spatial in nature. For instance, it is possible to portray the crimes by their shared features, or in a hierarchical fashion. Shepard[57] suggests that *additive trees*[58,59] are especially well-suited to representing the structure of semantic stimuli. Items in an additive tree are not represented by meaningful coordinates; rather, the distance (i.e., similarity) between any two items in an additive tree is indicated by the sum of the length of the links between those two items. For example, Shepard[57] produced an additive tree for 30 species of animals.[60,61] One branch linked the carnivores together and further sub-divided canines from felines.

One may envision a tree structure for crimes that has different branches for financial crimes, violent crimes, and moving violations, and perhaps further branching delineating vehicle moving violations (e.g., speeding, DUI) from walking violations (e.g., jaywalking). When hierarchical structures are not desired or appropriate, *additive clustering* may be employed. Imagine, as before, that you are interested in representing the similarity of a dog, cat, wolf, and tiger. A hierarchical approach would necessarily obstruct classification based on one characteristic (e.g., feline vs. canine) at the expense of another (potentially equally salient) characteristic (e.g., size, domesticity). Additive clustering is not based on a distance construct, but assumes instead that similarity can be measured by the weighted sum of the shared properties between two items.[62] This analysis identifies clusters of similar items, allowing each item to be included in several clusters; moreover, each cluster is ranked according to its estimated weight. For example, Shepard[57] applied additive clustering to data for 20 names of body parts.

BOX 1

VERIFYING THE INTERPRETATION OF DIMENSIONS

When researchers use MDS for confirmatory purposes (i.e., when the analyst has prior hypotheses regarding the likely dimensions of the space) it is possible to indirectly verify the interpretation of the resulting dimensions by using linear regression.[49,63] For instance, suppose we are interested in the degrees of similarity among a group of animals. A reasonable hypothesis would be that people appreciate the animals based on domesticity (domestic vs. wild). To test this hypothesis, in addition to collecting pairwise similarity estimates, individual domesticity ratings would be obtained for each animal (e.g., 'On a scale from 1 to 9, how domesticated is this animal?'). These ratings would then be regressed over the coordinates derived (per item) from each dimension in the resulting MDS plot. A high regression weight could then be taken as evidence that a particular dimension reflects the hypothesized construct.

From this analysis, one can appreciate the similarity of an elbow to a hand (i.e., both are part of the arm), and can also appreciate the (albeit weaker) similarity between an elbow and a knee (i.e., both are joints). With respect to the crime stimuli, a potential additive clustering analysis may reveal a strong grouping of DUI with the other moving violations, but also a weaker grouping with murder (i.e., some DUIs result in wrongful death). Simply put, although MDS is a particularly useful tool for representing similarity structures, it is only one of many ways that analysts can reveal the underlying structure of data (Box 1).

CONCLUSION

MDS is a useful tool to help quantify the ubiquitous, but slippery, notion of similarity. Although we all know what it means for two things to share a sense of closeness, similarity is difficult to estimate empirically. An argument could be made that any two items are similar because they share a potentially infinite number of arbitrary features. It seems likely that most people judge a pencil and a remote control to be dissimilar. But how dissimilar, exactly, might they be? Both items are solid objects that exist on earth, can be held in your hand, and can be used to scratch between your shoulder blades. Only when these objects are placed in the context of a larger set of items (e.g., other writing utensils, other electronic devices) can one begin to appreciate the ways in which they differ. By subjecting similarity estimates to MDS, researchers acquire maps of the relationships among a set of stimuli. This map reduces the complexity inherent to a large table of proximities, and can be used to explore a space about which no prior hypotheses exist, or to confirm *a priori* notions about the organization of psychological space. MDS can be performed in various ways, using data obtained from overt ratings or indirect methods. Although there are important subtleties regarding the treatment of data, dimensional selection, and the interpretations of solutions, MDS analyses all share a key purpose: They reveal the relational structures among the rated items. The technique has broad applicability, and is of the utmost value to researchers in the cognitive sciences and beyond.

REFERENCES

1. Shepard RN. Stimulus and response generalization: a stochastic model relating generalization to distance in psychological space. *Psychometrika* 1957, 22:325–345. doi:10.1007/BF02288967.

2. Shepard RN. Toward a universal law of generalization for psychological science. *Science* 1987, 237:1317–1323. doi:10.1126/science.3629243.

3. Hintzman DL. "Schema abstraction" in a multiple-trace memory model. *Psychol Rev* 1986, 93:411–428. doi:10.1037/0033-295X.93.4.411.

4. Hintzman DL. Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychol Rev* 1988, 95:528–551. doi:10.1037/0033-295X.95.4.528.

5. Goldinger SD. Echoes of echoes? An episodic theory of lexical access. *Psychol Rev* 1998, 105:251–279. doi:10.1037/0033-295X.105.2.251.

6. Goldinger SD, Azuma T. Episodic memory reflected in printed word naming. *Psychonomic Bull Rev* 2004, 11:716–722. doi:10.3758/BF03196625.

7. Goldstone RL. The role of similarity in categorization: Providing a groundwork. *Cognition* 1994, 52:125–157. doi:10.1016/0010-0277(94)90065-5.

8. Nosofsky RM. Attention, similarity, and the identification-categorization relationship. *J Exp Psychol: General* 1986, 115:39–57. doi:10.1037/0096-3445.115.1.39.

9. Nosofsky RM. Similarity scaling and cognitive process models. *Annu Rev Psychol* 1992, 43:25–53. doi:10.1146/annurev.ps.43.020192.000325.

10. Papesh MH, Goldinger SD. A multidimensional scaling analysis of own- and cross-race face spaces. *Cognition* 2010, 116:283–288. doi:10.1016/j.cognition.2010.05.001.

11. Bergmann Tiest WM, Kappers AML. Analysis of haptic perception of materials by multidimensional scaling and physical measurements of roughness and compressibility. *Acta Psychologica* 2006, 121:1–20. doi:10.1016/j.actpsy.2005.04.005.

12. Goldstone RL, Medin DL, Gentner D. Relational similarity and the nonindependence of features in similarity judgments. *Cogn Psychol* 1991, 23:222–262. doi:10.1016/0010-0285(91)90010-L.

13. Goldstone RL, Medin DL, Halberstadt J. Similarity in context. *Mem Cogn* 1997, 25:237–255. doi:10.3758/BF03201115.

14. Attneave F. Dimensions of similarity. *Am J Psychol* 1950, 63:516–556. doi:10.2307/1418869.

15. Richardson MW. Multidimensional psychophysics. *Psychol Bull* 1938, 35:659–660.

16. Kruskal JB. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 1964, 29:1–27. doi:10.1007/BF02289565.

17. Kruskal JB. Nonmetric multidimensional scaling: A numerical method. *Psychometrika* 1964, 29:115–129. doi:10.1007/BF02289694.

18. Shepard RN. The analysis of proximities: multidimensional scaling with an unknown distance function. Part I. *Psychometrika* 1962, 27:125–140. doi:10.1007/BF02289630.

19. Shepard RN. The analysis of proximities: multidimensional scaling with an unknown distance function. Part II. *Psychometrika* 1962, 27:219–246. doi:10.1007/BF02289621.

20. Torgerson WS. *Theory and Methods of Scaling*. New York: Wiley; 1958.

21. Torgerson WS. Multidimensional scaling of similarity. *Psychometrika* 1965, 30:379–393. doi:10.1007/BF02289530.

22. Ding CS. Multidimensional scaling modeling approach to latent profile analyses in psychological research. *Int J Psychol* 2006, 41:226–238. doi:10.1080/00207590500412219.

23. Newton I. *Opticks*. London: Smith and Walford; 1704.

24. Ashby GF, Perrin NA. Toward a unified theory of similarity and recognition. *Psychol Rev* 1988, 95:124–150. doi:10.1037/0033-295X.95.1.124.

25. Russel JA, Bullock M. Multidimensional scaling of emotional facial expressions: similarity from pre-schoolers to adults. *J Pers Soc Psychol* 1985, 48:1290–1298. doi:10.1037/0022-3514.48.5.1290.

26. Pedelty L, Cohen Levine S, Shevall SK. Developmental changes in face processing: results from multidimensional scaling. *J Exp Child Psychol* 1985, 39: 421–436. doi:10.1016/0022-0965(85)90050-5.

27. Hornberger M, Bell B, Graham KS, Rogers TT. Are judgements of semantic relatedness systematically impaired in Alzheimer's disease? *Neuropsychologia* 2009, 47:3084–3094. doi:10.1016/j.neuropsychologia.2009.07.006.

28. Hollins M, Faldowski R, Rao S, Young F. Perceptual dimensions of tactile surface features: a multidimensional scaling analysis. *Percept Psychophys* 1993, 54:697–705. doi:10.3758/BF03211795.

29. Youngentob SL, Johnson BA, Leon M, Sheehe PR, Kent PF. Predicting odorant quality perceptions from multidimensional scaling of olfactory bulb glomerular activity patterns. *Behav Neurosci* 2006, 120: 1337–1345. doi:0.1037/0735-7044.120.6.1337.

30. Carroll JD, Green PE. Psychometric methods in marketing research: Part II, multidimensional scaling. *J Market Res* 1997, 34:193–204.

31. Rabinowitz GB. An introduction to nonmetric multidimensional scaling. *Am J Polit Science* 1975, 19:343–390.

32. Amato PR. Dimensions of the family environment as perceived by children: a multidimensional scaling analysis. *J Marriage Fam* 1990, 52:613–620. doi:10.2307/352928.

33. Kenkel NC, Orlóci L. Applying metric and nonmetric multidimensional scaling to ecological studies: some new results. *Ecology* 1986, 67:919–928.

34. Giguère G. Collecting and analyzing data in multidimensional scaling experiments: a guide for psychologists using SPSS. *Tutorials Quant Methods Psychol* 2006, 2:26–37.

35. SPSS Inc. *SPSS Base 20.0 for Windows User's Guide*. Chicago, IL: SPSS Inc.; 2011.

36. Jaworska N, Chupetlovska-Anastasova A. A review of multidimensional scaling (MDS) and its utility in various psychological domains. *Tutorials Quant Methods Psychol* 2009, 5:1–10.

37. Borg I, Groenen P. *Modern multidimensional scaling: theory and applications*. New York: Springer-Verlag; 1997.

38. Faye P, Brémaud D, Teillet E, Courcoux P, Giboreau A, Nicod H. An alternative to external preference mapping based on consumer perceptive mapping. *Food Qual Prefer* 2006, 17:604–614. doi:10.1016/j.foodqual.2006.05.006.

39. Gilmore GC, Hersh H, Caramazza A, Griffin J. Multidimensional letter similarity derived from recognition errors. *Percept Psychophys* 1979, 25:425–431. doi:10.3758/BF03199852.

40. Podgorny P, Garner WR. Reaction time as a measure of inter- and intraobject visual similarity: Letters of the alphabet. *Percept Psychophys* 1979, 26:37–52. doi:10.3758/BF03199860.

41. Konkle T, Brady TF, Alvarez GA, Oliva A. Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *J Exp Psychol: General* 2010, 139:558–578.

42. Goldstone RL. An efficient method for obtaining similarity data. *Behav Res Methods Instrum Comput* 1994, 26:381–386.

43. Hout MC.Goldinger SD, Ferguson RW. The versatility of SpAM: a fast, efficient spatial method of data collection for multidimensional scaling. *J Exp Psychol: General* 2012. doi:10.1037/a0028860.

44. Kriegeskorte N, Marieke M. Inverse MDS: inferring dissimilarity structure from multiple item arrangements. *Front Psychol* 2012, 3:1–12.

45. Busing FMTA, Commandeur JJF, Heiser WJ, Bandilla W, Faulbaum F. PROXSCAL: a multidimensional scaling program for individual differences scaling with constraints. *Adv Stat Softw* 1997, 6:67–73.

46. Young FW, Takane Y, Lewyckyj R. ALSCAL: a nonmetric multidimensional scaling program with several individual-differences options. *Behav Res Methods* 1978, 10:451–453. doi:10.3758/BF03205177.

47. Carroll JD, Chang JJ. Analysis of individual differences in multidimensional scaling via an N-way generalization of Eckart-Young decomposition. *Psychometrika* 1970, 35:283–319. doi:10.1007/BF02310791.

48. Busing FMTA, Groenen PJK, Heiser WJ. Avoiding degeneracy in multidimensional unfolding by penalizing on the coefficient of variation. *Psychometrika* 2005, 70:71–98. doi:10.1007/s11336-001-0908-1.

49. Kruskal JB, Wish M. *Multidimensional Scaling*. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-011. Beverly Hills and London: Sage Publications; 1978.

50. Tversky A, Hutchinson JW. Nearest neighbor analysis of psychological spaces. *Psychol Rev* 1986, 93:3–22. doi:10.1037/0033-295X.93.1.3.

51. Torgerson WS. Multidimensional scaling: I. Theory and method. *Psychometrika* 1952, 17:401–419. doi:10.1007/BF02288916.

52. Janal MN, Clark WC, Carroll JD. Multidimensional scaling of painful and innocuous electrocutaneous stimuli: reliability and individual differences. *Percept Psychophys* 1991, 2:108–116. doi:10.3758/BF03212212.

53. Roskam E. *Metric Analysis of Ordinal Data in Psychology*. Doctoral dissertation, University of Leiden.; 1968.

54. Lee MD. Determining the dimensionality of multidimensional scaling representations for cognitive modeling. *J Math Psychol* 2001, 45:149–166. doi:10.1006/jmps.1999.1300.

55. Oh M-S. A simple and efficient Bayesian procedure for selecting dimensionality in multidimensional scaling. *J Multivar Anal* 2011, 107:200–209. doi:10.1016/j.jmva.2012.01.012.

56. Takane Y, Young FW, de Leeuw J. Nonmetric individual differences multidimensional scaling: an alternating least squares method with optimal scaling features. *Psychometrika* 1977, 42:7–67. doi:10.1007/BF02293745.

57. Shepard RN. Multidimensional scaling, tree-fitting, and clustering. *Science* 1980, 210:390–398. doi:10.1126/science.210.4468.390.

58. Cunningham JP. Free trees and bidirectional trees as representations of psychological distance. *J Math Psychol* 1978, 17:165–188. doi: 10.1016/0022-2496(78)90029-9.

59. Carroll JD. Spatial, non-spatial and hybrid models for scaling. *Psychometrika* 1976, 41:439–463. doi:10.1007/BF02296969.

60. Sattath S, Tversky A. Additive similarity trees. *Psychometrika* 1977, 42:319–345. doi:10.1007/BF02293654.

61. Henley NM. A psychological study of the semantics of animal terms. *J Verb Learn Verb Behav* 1969, 8:176–184. doi:10.1016/S0022-5371(69)80058-7.

62. Shepard RN, Arabie P. Additive clustering: representations of similarities as combinations of discrete overlapping properties. *Psychol Rev* 1979, 86:87–123. doi:10.1037/0033-295X.86.2.87.

63. Green PE, Camone FJ Jr, Smith SM. *Multidimensional Scaling: Concepts and Applications*. Needham Heights, MA: Allyn and Bacon; 1989.

FURTHER READING

Schiffman SS, Reynolds ML, Young FW. *Introduction to Multidimensional Scaling: Theory, Methods and Applications*. New York: Academic Press; 1981.

Shepard RN. How a cognitive psychologist came to seek universal laws. *Psychonomic Bull Rev* 2004, 11:1–23. doi: 10.3758/BF03206455.

Tversky A. Features of similarity. *Psychol Rev* 1977, 84:327–352. doi: 10.1037/0033-295X.84.4.327.