ON THE DANGERS OF AVERAGING ACROSS SUBJECTS WHEN USING MULTIDIMENSIONAL SCALING OR THE SIMILARITY-CHOICE MODEL

F. Gregory Ashby, W. Todd Maddox, and W. William Lee

University of California at Santa Barbara and Arizona State University

Abstract—When ratings of judged similarity or frequencies of stimulus identification are averaged across subjects, the psychological structure of the data is fundamentally changed. Regardless of the structure of the individual-subject data, the averaged similarity data will likely be well fit by a standard multidimensional scaling model, and the averaged identification data will likely be well fit by the similarity-choice model. In fact, both models often provide excellent fits to averaged data, even if they fail to fit the data of each individual subject. Thus, a good fit of either model to averaged data cannot be taken as evidence that the model describes the psychological structure that characterizes individual subjects. We hypothesize that these effects are due to the increased symmetry that is a mathematical consequence of the averaging operation.

MULTIDIMENSIONAL SCALING

In MDS, the dissimilarity between a pair of stimuli is assumed to increase with the distance between point representations of the stimuli in some psychological space. Typically (i.e., in nonmetric MDS), the observable ratings of judged dissimilarity are assumed to be related only ordinally to the underlying and unobservable perceived dissimilarities. The perceived dissimilarities are assumed to be proportional to psychological distance.

Let $d_s$ denote the perceived dissimilarity between a pair of stimuli $i$ and $j$, and suppose there are only two relevant psychological dimensions. Let $(x_i, y_i)$ denote the coordinates of stimulus $i$ in the psychological space. Many different MDS models can be constructed depending on how perceived dissimilarity is computed from the coordinate values of the relevant stimuli. The most widely used class of MDS models assumes dissimilarity is computed from the Minkowski metric:

$$d_y = [(x_i - x_j)^2 + (y_i - y_j)^2]^{1/2}.$$ (1)

The well-known euclidean metric results when $r = 2$, and the city-block metric when $r = 1$.

It is well known that a set of perceived dissimilarities can be represented by some MDS model if and only if the following four distance axioms are true.

1. Other kinds of averaging may have fundamental effects also. For example, Ennis (1888) illustrated the dangers of averaging across trials when confusable stimuli are used in identification experiments. This kind of averaging, which is implicitly assumed by the SCM, loses information about trial-by-trial variability in the percept and can lead to different decisions about the correct form of the function relating similarity and psychological distance.

2. Many models that elaborate on this idea are sometimes classified as MDS models. These include models that assume perceived dissimilarity is computed in a way that violates the basic properties of distance (Krumhansl’s, 1978, distance-density model) or that assume the stimulus representation varies probabilistically from trial to trial (Ennis, Palen, & Mullen’s, 1988, probabilistic MDS model). However, this article assumes the more narrow definition of MDS models described in the text.
AXIOM 1. Equal self-dissimilarities: \( d_{ii} = 0 \), for any stimulus \( i \).

AXIOM 2. Minimality: \( d_{ij} \geq 0 \), for any two stimuli \( i \) and \( j \).

AXIOM 3. Symmetry: \( d_{ij} = d_{ji} \), for every pair of stimuli \( i \) and \( j \).

AXIOM 4. Triangle inequality: \( d_{ij} + d_{jk} \geq d_{ik} \), for any three stimuli \( i, j, \) and \( k \).

If any of these axioms is false, then there is no MDS model that can account for the data. For the Minkowski metric, the distance axioms are satisfied only when \( r = 1 \). Thus, a set of perceived dissimilarities generated from Equation 1 with \( r < 1 \) cannot represent distances between points in a psychological space.

Unfortunately for MDS, Tversky and other researchers have argued persuasively that most, if not all, of these axioms are not generally true (Ashby & Perrin, 1988; Krumhansl, 1978; Tversky, 1977; Tversky & Gati, 1982). Despite these criticisms, MDS models frequently provide impressive fits to dissimilarity ratings, particularly when the data are averaged across subjects. For example, Shin and Nosofsky (1992) collected dissimilarity ratings for pairs of random dot patterns (each with nine dots) from six groups of 20 subjects each and then performed an MDS analysis. An MDS model with six psychological dimensions accounted for about 93% of the variance in the data, despite the fact that the stimuli had no obvious dimensional structure, and that the subjects had very little experience in the task.

In contrast, when single-subject analyses are performed, the MDS model seems to have more difficulty. For example, Ashby and Lee (1991) collected similarity ratings for 2 subjects on semicircles that varied in size and in the orientation of a radial line. Each subject made approximately 44 similarity judgments of every stimulus pair, and both subjects had first participated in an extensive stimulus identification experiment with the same stimuli. Despite the obvious dimensional structure of the stimuli, the single-subject analyses, and the use of highly practiced subjects, the best MDS model accounted for less than 60% of the variance in the data for the first subject and less than 73% of the variance in the data for the second subject. Because averaging across subjects is common practice when performing an MDS analysis (Burns & Shepp, 1988; Burns, Shepp, McDonough, & Wiener-Ehrlich, 1978; Melara, 1989; Nosofsky, 1991a), it is important to understand the discrepancy between the studies of Shin and Nosofsky (1992) and Ashby and Lee (1991).

Furnas (1989) described an elegant method for understanding the effects of averaging across subjects on the validity of the MDS model. To begin, consider a simple experiment with three stimuli denoted by \( i, j, \) and \( k \); in particular, consider the three perceived dissimilarities \( d_{ij}, d_{ik}, \) and \( d_{jk} \). The perceived dissimilarities can be assumed to lie on a ratio scale, since multiplication of each dissimilarity by the same positive constant does not affect the validity of any of the distance axioms. Furnas chose to multiply each of \( d_{ij}, d_{ik}, \) and \( d_{jk} \) by 1 divided by their sum. After this transformation, \( d_{ij} + d_{jk} + d_{ik} = 1 \), or equivalently, \( d_{ik} = 1 - (d_{ij} + d_{jk}) \). The advantage of this transformation is that the three perceived dissimilarities can now be represented in the flat, two-dimensional diagram shown in Figure 1.

The abscissa in Figure 1 represents all possible values of \( d_{ij} \), and the ordinate represents all possible values of \( d_{ik} \). The point labeled "1" corresponds to \( d_{ij} = 0.2 \) and \( d_{ik} = 0.6 \). Given these two values, we know that \( d_{jk} = 1 - (0.2 + 0.6) = 0.2 \). Because of the constraint that the three dissimilarities sum to 1, and because we assume perceived dissimilarities are never negative (as is true of every ratio scale), both \( d_{ij} \) and \( d_{ik} \) are bounded above and below by 0 and 1. In addition, the sum \( d_{ij} + d_{jk} \) must be less than or equal to 1, so every pair \((d_{ij}, d_{jk})\) must lie within the large unshaded triangle illustrated in Figure 1.

Furnas (1989) showed how this diagram can be used to investigate a number of interesting issues relating to MDS. For the present purposes, his discussion of the triangle inequality is of special relevance. Not all points in Figure 1 satisfy the triangle inequality. For example, the point \((0, 0)\) corresponds to dissimilarities of \( d_{ij} = 0, d_{ik} = 0, \) and \( d_{jk} = 1 \), which violate the triangle inequality.

With three stimuli, the triangle inequality imposes a total of

![Fig. 1. Furnas (1989) plot of the perceived dissimilarities in an experiment with stimuli i, j, and k. All sets of data falling in the shaded region satisfy the triangle inequality. The points labeled "1" and "2" represent two hypothetical data sets; the point on the line segment connecting them represents their averaged data.](image-url)
Averaging Across Subjects

three constraints on the perceived dissimilarities. The first is that

\[ d_{ij} + d_{ik} \geq d_{jk}. \]

Because the three dissimilarities sum to 1, this inequality can be rewritten as

\[ d_{ij} + d_{jk} \geq 1 - (d_{ij} + d_{jk}), \]

or equivalently as

\[ d_{ij} \geq -d_{jk} + 0.5. \]  \hspace{1cm} (2)

The second constraint is that

\[ d_{ji} + d_{ik} \geq d_{jk}, \]

which implies

\[ d_{jk} \leq 0.5. \]  \hspace{1cm} (3)

and the third constraint is that

\[ d_{jk} + d_{ik} \geq d_{ij}, \]

which implies

\[ d_{ij} \leq 0.5. \]  \hspace{1cm} (4)

Inequalities 2, 3, and 4 are simultaneously satisfied only by points falling in the small shaded triangular region illustrated in Figure 1. Thus, any data corresponding to a point falling outside the shaded region is incompatible with all MDS models. If a data point falls inside the shaded region, then it is compatible with some MDS model, although it still might be incompatible with the most familiar versions of MDS, namely, those based on euclidean or city-block distance.

Consider next an experiment with 2 subjects. Suppose their data can be described by the two points labeled “1” and “2” in Figure 1. Furnas (1989) showed that their averaged data can be described by the midpoint of the line segment connecting their individual data points. This result leads to the following important consequences of averaging data on the validity of the triangle inequality. First, because the shaded region in Figure 1 is convex 2 (Royden, 1968), if the data of each individual subject satisfy the triangle inequality, then the averaged data must satisfy the triangle inequality. Second, if the data of each individual subject violate the triangle inequality, then the averaged data frequently will satisfy the triangle inequality, particularly when there are reasonable individual differences.

These results occur because averaging across subjects tends to reduce differences in the pair-wise dissimilarities. For example, if \( d_{ij} \) is much smaller than \( d_{jk} \) for Subject 1 but much larger than \( d_{ij} \) for Subject 2, then the two dissimilarities will be nearly equal in the averaged data. If \( d_{ij} = d_{jk} = d_{ik} \), then the triangle inequality is satisfied and the data can be described perfectly by the euclidean MDS model with the stimulus points arranged in an equilateral triangle.

These results suggest, but do not imply, that standard MDS models will fit averaged data much better than the data of individual subjects when moderate individual differences occur.

To test this hypothesis, we performed the following set of Monte Carlo simulations. Nine stimuli were assumed to lie in a two-dimensional psychological space with the coordinate values shown in Figure 2. Call this the group psychological space. To model individual differences, we assumed that the psychological space for each subject was a distorted version of the group space. Specifically, we assumed that if \( (x_i, y_i) \) are the coordinates of stimulus \( i \) in the group psychological space, then the coordinates of stimulus \( i \) in the psychological space of subject \( m \) are

\[ (x_i + \epsilon_{xim}, y_i + \epsilon_{yim}), \]

where the terms \( \epsilon_{xim} \) and \( \epsilon_{yim} \) represent the magnitude of individual differences. We assumed \( \epsilon_{xim} \) and \( \epsilon_{yim} \) are independent and that each has a normal distribution with a mean of 0 and a standard deviation of \( \sigma = 0.50 \) or \( \sigma = 0.67 \). This rule was used to construct a unique psychological space for each of 30 subjects. Next, for every subject, the perceived dissimilarity of each pair of stimuli was computed separately from Equation 1 with \( r = 0.5, r = 1, \) and \( r = 2 \).

The \( r = 1 \) and the \( r = 2 \) data were generated primarily as controls. We are most interested in the \( r = 0.5 \) data. In this case, it is known that the resulting perceived dissimilarities are

---

5. To determine if a region is convex, pick any two points in the region and connect them with a line segment. If all points on the segment fall inside the region, then the region is convex.

---

Fig. 2. Psychological structure of the stimuli used in the Monte Carlo simulations of a similarity rating experiment.


(Received 4/14/93; Revision accepted 11/23/93)