17 MDS as a Psychological Model

MDS has been used not only as a tool for data analysis but also as a framework for modeling psychological phenomena. This is made clear by equating an MDS space with the notion of psychological space. A metric geometry is interpreted as a model that explains perceptions of similarity. Most attention has been devoted to investigations where the distance function was taken as a composition rule for generating similarity judgments from dimensional differences. Minkowski distances are one family of such composition rules. Guided by such modeling hypotheses, psychophysical studies on well-designed simple stimuli such as rectangles uncovered interesting regularities of human similarity judgments. This model also allows one to study how responses conditioned to particular stimuli are generalized to other stimuli.

17.1 Physical and Psychological Space

In most applications of MDS today, little attention is devoted to the Shepard diagram. It may therefore surprise the reader that ordinal MDS was originally invented to study the shape of the regression curve in this diagram, not the MDS configuration. This also makes clear how closely MDS used to be related to efforts for modeling psychological phenomena, where the MDS geometry served as a model of psychological space and the distance function as a model of mental arithmetic.

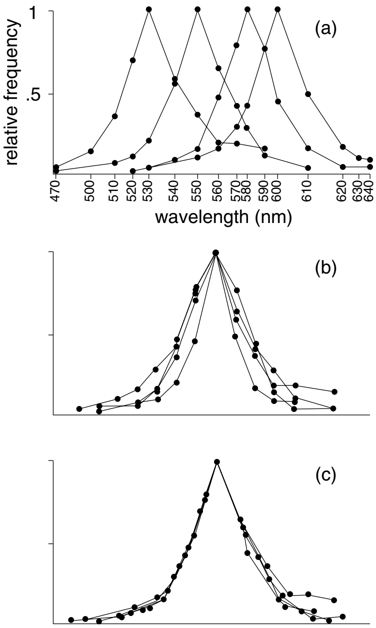

FIGURE 17.1. (a) Four generalization gradients over the electromagnetic spectrum, with intervals adjusted to make gradients similar; panel (b) shows superimposed gradients constructed over the nonadjusted scale; panel (c) shows superimposed gradients from panel (a).

The Shape of Generalization Gradients in Learning

The scientific context for the interest in Shepard diagrams becomes clear from the following experiment. Guttman and Kalish (1956) trained four groups of pigeons to peck at a translucent plastic key when illuminated from behind by monochromatic light with wavelengths 530, 550, 580, and 600 nm, respectively. After the learning sessions, they assessed the frequency with which the pigeons in each group pecked at the key when illuminated with different colors. Figure 17.1a shows that the probability of pecking at the key is highest for the original conditioned color and decreases monotonically as a function of the difference between the original and the test color.

One can ask whether such generalization gradients always have the same shape. Are they, say, always exponential decay functions over the stimulus dimensions? This is difficult to decide, because it is the *psychological*, not the *physical*, stimulus dimensions that are relevant. In a simple case such as the Guttman–Kalish experiment, the stimuli vary on just one dimension. The physical (here: wavelength) and the psychological (here: hue) dimensions are related to each other by a psychophysical mapping. The shape of generalization gradients depends on this mapping. This can be seen from

Figure 17.1, taken from Shepard (1965). The X-axis in panel (a) shows the physical wavelengths of the stimuli, but its units have been somewhat compressed and stretched locally to make the four gradients as similar as possible. Without these adjustments, that is, over the physical wavelength scale, the gradients are less similar (Figure 17.1b). After the adjustment, the gradients are almost equal in shape (Figure 17.1c).

Thus, knowledge of the psychological space of the stimuli, or at least the "psychological" distances between any two stimuli, S_i and S_k, is necessary for meaningful statements on the shape of generalization gradients. Older approaches often tried to arrive at psychological distances directly by summing just noticeable differences (JNDs) between S_i and S_k. The idea that this sum explains the subjective dissimilarity of S_i and S_k goes back to Fechner (1860). There are many problems associated with this model (Krantz, 1972), but one is particularly important for MDS: "Unfortunately, in order to sum JNDs between two stimuli, this summation must be carried out along some path between these stimuli. But the resulting sum will be invariant ... only if this path is a least path, that is, yields a shortest distance (in psychological space) between the two stimuli. We cannot presume, in arbitrarily holding certain physical parameters constant ..., that the summation is constrained thereby to a shortest path ... in psychological space, even though it is, of course, confined to a shortest path ... in physical space. ... These considerations lead us to look for some way of estimating the psychological distance between two stimuli without depending either upon physical scales or upon any arbitrary path of integration" (Shepard, 1957, p. 334).

Relating Physical Space to Psychological Space

An *external* approach for the problem of estimating psychological distances first assumes a particular correspondence of physical space to psychological space and then explains how the response probabilities are distributed over this space. An *internal* approach, in contrast, builds directly and exclusively on the response probabilities and formulates how these arise as a function of unknown psychological distances. Let us consider Shepard's original derivations (Shepard, 1957). Let p_{ik} be the probability of giving the S_i response to stimulus S_k. If $i = k$, then p_{ik} is the probability of giving the correct response. It is postulated that there exists a function f such that p_{ik} is proportional to $f(d_{ik})$, where d_{ik} is the psychological distance between S_i and S_k,

$$p_{ik} = c_i \cdot f(d_{ik}), \tag{17.1}$$

with c_i a proportionality constant associated with S_i. Summing over all k, we obtain $\sum_k p_{ik} = 1$ and $c_i \cdot \sum_k f(d_{ik})$ for the two sides of (17.1), so that

$c_i = 1/\sum_k f(d_{ik})$. Inserting this term for c_i in (17.1) yields

$$p_{ik} = f(d_{ik})/\sum_j f(d_{ij}). \tag{17.2}$$

With the p_{ik}-values given as data, we now search for a function f that satisfies (17.2). The important point here is that the d-values on the right-hand side are not just any values that satisfy (at least approximately) all equations of type (17.2), but they must also possess the properties of distances and even of Euclidean distances in a space of given dimensionality. Moreover, we would not accept any function f, but only those that are *smooth* (continuous) and monotone increasing or decreasing. Then f is invertible, so that response probabilities can in turn be derived from the psychological distances. If we assume that the psychological space is related to the physical space by a smooth transformation, then straight lines in physical space are transformed into lines in psychological space that may not be straight but smoothly curved. Hence, given any three stimuli on a straight line in physical space, their psychological images should also be approximately on a straight line if the stimuli are physically similar. From this assumption and some additional simple postulates on decay and diffusion of memory traces, Shepard (1958a) derives that f is a negative exponential function. Elsewhere, without any assumptions, Shepard (1957) simply defines f to be a negative exponential function. This function turns (17.2) into

$$p_{ik} = \exp(-d_{ik})/\sum_j \exp(-d_{ij}). \tag{17.3}$$

Because $d_{ii} = 0$, $\exp(-d_{ii}) = \exp(0) = 1$ and so

$$p_{ik}/p_{ii} = \exp(-d_{ik}). \tag{17.4}$$

Dividing p_{ik} by p_{ii} means that the probability of giving the i response to stimulus k is expressed relative to the probability of responding properly to S_i. Thus, norming all response probabilities in this way, and specifying that d_{ik} is a Euclidean distance in a space with dimensionality m, we end up with a metric MDS problem that requires finding a point space such that its distances satisfy (17.4) as closely as possible. A reasonable choice for m should be the dimensionality of the physical space.

Determining the Shape of Generalization Gradients via MDS

The discussion above led to a confirmatory MDS problem: the data (i.e., the ratios p_{ik}/p_{ii}) are to be optimally mapped into a particular model. The fit of the model to the data is then evaluated. Shepard (1958b) concluded that the negative exponential function allows one to explain the data sufficiently well, but other functions, such as a simple linear one, may also be in good or even better agreement with the data. Shepard tried to solve this problem and allow the data to "reveal themselves" by requiring only that f in (17.1) be monotonically decreasing rather than some specific parametric function. In other words, expressed in terms of the generalization gradients, he required that they should decrease from the correct stimulus S_r monotonically into all directions of the stimulus space.

To see how a psychological scale (e.g., the X-axis in Figure 17.1a) is derived, fold Figure 17.1b at the points where the gradients peak. What will then be obtained is nothing other than a Shepard diagram, where the data appear on the Y-axis and the "psychological" distances on the X-axis. Hence, finding the psychological scale amounts to using ordinal MDS with $m = 1$ in the present case. Of course, the Shepard diagram will show a scatter of points only, and the various gradients have to be found by unfolding the Shepard diagram and connecting the respective points. The unfolding is done simply by arraying the points in the order of their physical stimulus coordinates (here: wavelengths) and with distances among them as computed by the MDS procedure.

17.2 Minkowski Distances

Over a 2D stimulus space, the generalization gradients are surfaces such as the cones and pyramids shown schematically in Figure 17.2. Assume that the directions labeled as D1 and D2 are psychologically meaningful dimensions such as hue and saturation for color stimuli. Assume further that the correct stimulus S_r corresponds to the point where D1 and D2 intersect. Cross (1965a) then distinguishes the following three models: (1) the *excitation model*, which assumes that the generalization gradient decreases evenly around S_r into all directions of the psychological space; (2) the *discrimination model*, which says that the strength of reacting to a stimulus different from S_r on both dimensions corresponds to the sum of the generalization of S_r on both dimensions; and (3) the *dominance model*, where the strength of reacting to $S_i \neq S_r$ is determined by only that dimension on which S_i and S_r differ most. These models are illustrated in Figure 17.2. The gradients are shown as linear functions to simplify the pictures. Note that the gradients for the discrimination model and the dominance model have the same shape (for a two-dimensional psychological space) but differ in their orientation relative to the dimensions.

The Family of Minkowski Distances

The generalization models in Figure 17.2 illustrate three special cases of the *Minkowski metric* or, equivalently, the *Minkowski distance*. The general

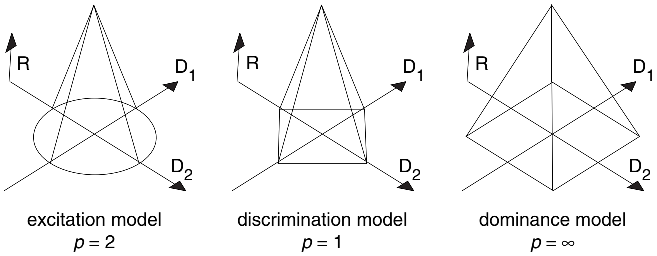

FIGURE 17.2. Three models of generalization over a 2D stimulus continuum; S_r corresponds to intersection of D1 and D2 (after Cross, 1965b).

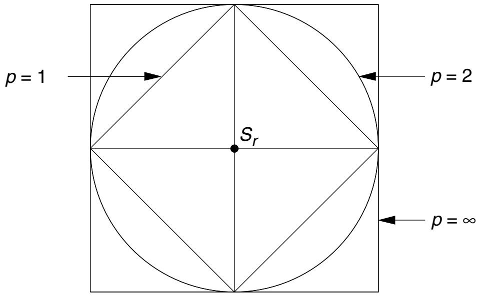

FIGURE 17.3. Three circles with same radius around S_r in 2D for different p-values in the Minkowski distance formula.

formula for this metric is

$$d_{ij}(\mathbf{X}) = \left(\sum_{a=1}^{m} |x_{ia} - x_{ja}|^p \right)^{1/p}, \ p \geq 1. \tag{17.5}$$

For $p = 2$, equation (17.5) yields the usual *Euclidean* distance formula. For $p = 1$, we obtain the *city-block* metric, and for $p \to \infty$, the *dominance* metric.

The implications of choosing different p-values can be seen from the following. If we look, from above, at the three gradient models in Figure 17.2, a circle, a diamond, and a square, respectively, appear in the psychological space. Superimposing these three figures leads to the diagram in Figure 17.3. Assume, for simplicity, that the point S_r has coordinates $(0, 0)$. Then, (17.5) reduces to

$$d_{rj} = (|x_{j1}|^p + |x_{j2}|^p)^{1/p}. \tag{17.6}$$

For $p = 1$ we obtain d_{rj} as just the sum of the absolute coordinates of S_j. Thus, all stimuli located on the diamond in Figure 17.3 have the same city-block distance to S_r. The diamond is therefore called the *isosimilarity curve* of the city-block metric. It is the set of all points with the same distance to

S_r. But this is just the definition of a circle in analytical geometry, so the diamond is nothing but *a circle in the city-block plane*, even though it does *not look* like a circle at all. Our Euclidean notion of a circle corresponds exactly to the isosimilarity curve for $p = 2$. Finally, the circle for $p \to \infty$ looks like a square with sides parallel to the dimensions.

It is important to realize that the distance d_{rj} for two given points S_r and S_j remains the same under rotations of the coordinate system only if $p = 2$. For $p = 1$, d_{rj} is smallest when both stimulus points lie on one of the coordinate axes. If the coordinate system is rotated about S_r, then d_{rj} grows (even though the points remain fixed), reaches its maximum at a 45° rotation, and then shrinks again to the original value at 90°. This behavior of a distance function may appear strange at first, but "... under a good many situations, [this distance] describes a reasonable state of affairs. For example, suppose one were in a city which is laid out in square blocks. A point three blocks away in one direction and four blocks away in the other would quite reasonably be described as seven blocks away. Few people, if asked, would describe the point as five blocks distant. Further, if new streets were put in at an angle to the old, the 'distance' between the two points would change" (Torgerson, 1958, p. 254).

Minkowski Distances and Intradimensional Differences

Further properties of different Minkowski distances follow directly from (17.5). Cross (1965b, 1965a) rearranges its terms in a way that we show here for the special case of (17.6):

$$\begin{aligned} d_{rj}^p &= |x_{j1}|^p + |x_{j2}|^p, \\ d_{rj} d_{rj}^{p-1} &= |x_{j1}|^{p-1} \cdot |x_{j1}| + |x_{j2}|^{p-1} \cdot |x_{j2}|, \\ d_{rj} &= \underbrace{(|x_{j1}|^{p-1}/d_{rj}^{p-1})}_{w_1} \cdot |x_{j1}| + \underbrace{(|x_{j2}|^{p-1}/d_{rj}^{p-1})}_{w_2} \cdot |x_{j2}|, \\ d_{rj} &= w_1 \cdot |x_{j1}| + w_2 \cdot |x_{j2}|. \end{aligned} \tag{17.7}$$

It follows that for $p = 1$, d_{rj} is just the sum of the coordinate values of stimulus S_j, because $w_1 = w_2 = 1$. If $p > 1$, then the coordinates are weighted by w_1 and w_2 in proportion to their size. If $p \to \infty$, d_{rj} approximates its largest coordinate value. This can be seen most easily from a numerical example. Table 17.1 shows such an example for $S_r =$ (0,0) and $S_j =$ (1,2), for which $|x_{j1}| = 1$ and $|x_{j2}| = 2$. For $p = 1$, we obtain $d_{rj} = (1/3)^0 \cdot 1 + (2/3)^2 \cdot 2 = 1 \cdot 1 + 1 \cdot 2 = 3$. For $p = 2$, we get $d_{rj} = (1/\sqrt{5})^1 \cdot 1 + (2/\sqrt{5})^1 \cdot 2 = 0.44721360 + 1.78885438 = 2.23606798$.

Generally, if $p \to \infty$, then $d_{rj} \to 2$; that is, as p grows, the larger of the two coordinates of S_j (i.e., the larger of the two-dimensional differences between S_r and S_j) tends to *dominate* the global distance value. Indeed, d_{rj} approximates the limiting value 2 quite rapidly as p grows: for $p = 20$, d_{rj} differs from 2 only in the seventh position after the decimal point.

TABLE 17.1. Demonstration of how dimensional differences (x_{ja}) enter the distance of two points r and j under different Minkowski p parameters, with $x_{r1} = 0, x_{r2} = 0, x_{j1} = 1, x_{j2} = 2$.

p	$w_1 \cdot x_{j1}$	$w_2 \cdot x_{j2}$	w_2/w_1	d_{rj}
1.0	1.00000000	2.00000000	1.00	3.00000000
1.5	0.63923401	1.80802681	1.41	2.44726081
2.0	0.44721360	1.78885438	2.00	2.23606798
3.0	0.23112042	1.84896340	4.00	2.08008382
4.0	0.11944372	1.91109947	8.00	2.03054318
5.0	0.06098020	1.95136642	16.00	2.01234662
10.0	0.00195141	1.99824382	512.00	2.00019523
20.0	0.00000191	1.99999819	524288.00	2.00000010

In terms of Figure 17.3, increasing p from 1 to 2 means that the diamond bulges outwards and approximates the Euclidean circle. For Minkowski parameters greater than 2, the circle then moves towards the square for $p \to \infty$. Hence, the three generalization models in Figure 17.2 correspond to different ways of *composing* a distance from given *intradimensional* differences between pairs of stimuli. For example, given two tones that differ in frequency and sound pressure, one possible composition rule yielding their subjective global dissimilarity would be simply to add their frequency and pressure differences in the corresponding psychological space, that is, add their differences in pitch and loudness. This corresponds to computing a city-block distance. The Euclidean distance formula, on the other hand, implies a composition rule that is much harder to understand. What is clear, though, is that, for all $p > 1$, the differences first are weighted and then added, with the larger differences receiving a larger weight. In the extreme case ($p \to \infty$), the largest difference completely dominates the dissimilarity judgment.[1]

Torgerson (1958), Garner (1962), and others argue that if the stimuli are such that their dimensions are obvious and natural (*analyzable stimuli*), then the city-block distance should be the best model to explain dissimilarity judgments. If, on the other hand, the stimuli are *integral*, then the Euclidean metric should be more appropriate.[2]

[1] Interpreting the Minkowski distance as a composition rule is just one possibility. Micko and Fischer (1970) and Fischer and Micko (1972), for example, present an alternative conceptualization in which the composition rule is not a summation of intradimensional differences. Rather, an attention distribution is postulated to exist over all directions in space, so that the effect of an increment in p in the Minkowski model corresponds to a concentration of attention in certain spatial directions.

[2] An example of an analyzable stimulus is the one-spoked wheel shown in Figure 1.6. Its “obvious and compelling dimensions” (Torgerson, 1958, p. 254) are its size and the inclination angle of its spoke. A color patch, on the other hand, is an integral stimulus whose dimensions hue, saturation, and brightness can be extracted only with effort.

Wender (1971) and Ahrens (1972) propose that as similarity judgments become more difficult—because of, say, time constraints or increasing complexity of the stimuli—subjects tend to simplify by concentrating on the largest stimulus differences only. Hence, we should expect that such similarity data could be explained best with large Minkowski p parameters.

Maximum Dimensionality for Minkowski Distances

Suppose that $\mathbf{D}$ is a matrix of Minkowski distances. If $\mathbf{D}$ is Euclidean, then there are at most $m = n - 1$ dimensions. But what about other cases of Minkowski distances? Fichet (1994) shows that for city-block distances the dimensionality can be at most $[n(n-1)/2]-1$. For the dominance distance, the maximum dimensionality is $n - 1$ (Critchley & Fichet, 1994), a result that goes back to Fréchet (1910). Note though that these theoretical results are not based on analyses that would allow us to identify the dimensionality of the underlying configuration $\mathbf{X}$ of a given $\mathbf{D}$, except for Euclidean distances (see Section 19.3).

In addition, Critchley and Fichet (1994) show that certain Minkowski distance matrices are *exchangeable*. To be more precise, for every Euclidean distance matrix, there exists a city-block and dominance distance matrix having the same values (most likely in a different dimensionality and with a different configuration). Also, for every city-block distance matrix there exists a dominance distance matrix having the same values. And, of course, all unidimensional Minkowski distance matrices are equal irrespective of the Minkowski parameter p. These results imply that a solution found by MDS using the Euclidean distance can be exchanged by a solution using the city-block distance (or the dominance distance) *without* changing the Stress value, although the dimensionality of the three solutions is most likely not the same.

17.3 Identifying the True Minkowski Distance

How can the true Minkowski distance be identified? There are two approaches, one based on scaling proximities in MDS with different metrics, and one based on analyzing the proximities and assuming certain properties of the psychological space.

Take two points in psychological space. The Euclidean distance between these points is not affected by rotations of the dimension system. The city-

---

Indeed, "if dimensions are integral, they are not really perceived as dimensions at all. Dimensions exist for the experimenter... But these are constructs... and do not reflect the immediate perceptual experience of the subject in such experiments..."(Garner, 1974, p. 119).

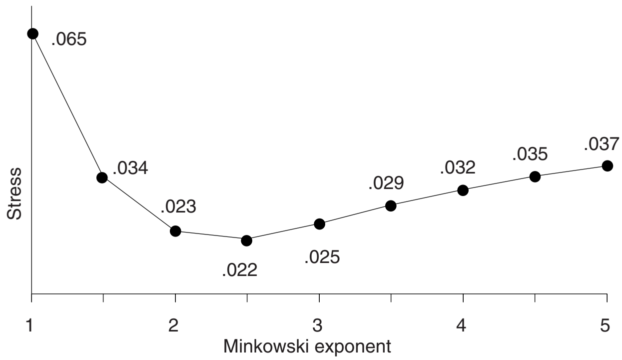

FIGURE 17.4. Stress values for representing data in Table 4.1 in 2D MDS spaces using Minkowski distances with different exponents (after Kruskal, 1964a).

block distance, however, is smallest if these points lie on a line parallel to a dimension and greatest if this line forms an angle of 45° to the dimensions.

This suggests that one should test whether two points with a given Euclidean distance are perceived as more dissimilar if they differ on just one dimension rather than on several dimensions. If this matters, then the Euclidean distance cannot be the true metric. Shepard (1964) attempted to check this condition by constructing one-spoked wheels with dimensions "size" and "angle of spoke" (as in Figure 1.6) and assuming that the psychological space is essentially equivalent to this 2D physical space. He observed that stimuli that differed on one dimension only were perceived as relatively similar as compared to those that differed on two dimensions, although their Euclidean distances in physical space were equal. He took this finding as supporting evidence for the city-block metric, which was predicted to be appropriate for such analyzable stimuli.

Determining the True Minkowski Distance by MDS

A second approach for determining the true Minkowski distance is to test how well given proximities can be represented in a space with a given metric. Such scaling tests are easy to compute but difficult to evaluate. If the dimensionality question can be settled beforehand in some way, Kruskal (1964a) suggests computing MDS representations for a large range of different p-values and then selecting as the true metric the one that leads to the lowest Stress. This is shown in Figure 17.4 for Ekman's color data from Table 4.1. The lowest Stress (.0215) occurs at $p = 2.5$. Kruskal (1964a) comments on this finding: "We do not feel that this demonstrates any significant fact about color vision, though there is the hint that subjective distance between colors may be slightly non-Euclidean" (p. 24).[3]

Ahrens (1974) proposes varying both p and m. In this way, a curve like the one in Figure 17.4 is obtained for each m. If these curves all dip at the same p-value, then we can decide the metric question independently of the dimensionality question.

Yet, proposals for deciding on the true metric empirically and not by theoretical considerations assume that the Stress values arrived at under different specifications for p and m are comparable. This requires that all solutions must be global minima, because otherwise it would not make sense to conclude that $p = 1$, say, yields a better solution than $p = 2$. The global minimum condition can be checked by using many—Hubert, Arabie, and Hesson-McInnis (1992) used 100!—different starting configurations for each fixed pair of p and m.[4]

We must, moreover, decide whether any small difference between two Stress values is significant. In Figure 17.4, the Stress values around p = 2.5 are quite similar. Should we really conclude that the subjects use p = 2.5 and not, say, $p = 2$, because the Stress is slightly smaller for the p parameter than for the latter? Probably not. It seems more reasonable to decide that the subjects used a p parameter close[5] to 2.

Distinguishing among MDS Solutions with Different Minkowski Distances

There are p-values that lead to the same Stress for a given 2D configuration, for example, the extreme cases $p = 1$ and $p \to \infty$. Figure 17.3 shows why this is so. If the dimension system is rotated by 45°, the isosimilarity contour for $p = 1$ is transformed into the isosimilarity contour for $p \to \infty$, except for its overall size. This means that city-block distances computed from a given MDS configuration and a given coordinate system are, except

[3] There are several ways to minimize Stress for Minkowski distances. A general gradient approach is taken in KYST, SYSTAT, and MINISSA. Groenen et al. (1995) and Groenen et al. (1999) give a majorization algorithm of which the SMACOF algorithm of Section 8.6 is a special case. The majorizing algorithm turns out to have a quadratic majorizing function for $1 \leq p \leq \infty$, so that each update can be found in one step. For p outside this range, the update has to be found by an iterative procedure.

[4] For the special (but important) case of city-block distances, Groenen and Heiser (1996) found many local minima. To find the global minimum, they applied the tunneling method (see Section 13.7). Different approaches were pursued by Heiser (1989b) and Hubert et al. (1992), who used combinatorial strategies, and Pliner (1996), who proposed to apply the smoothing strategy (see Section 13.5).

[5] Indeed, by scaling the data with more modern MDS programs, one finds that the minimum Stress is at $p = 2$. Arabie (1991) conjectured, moreover, that "to the extent that our theory predicts a circle ..., the curve in Figure [17.4] should be flat unless disturbed by either (a) numerical artifacts in computation or (b) noise in the data."

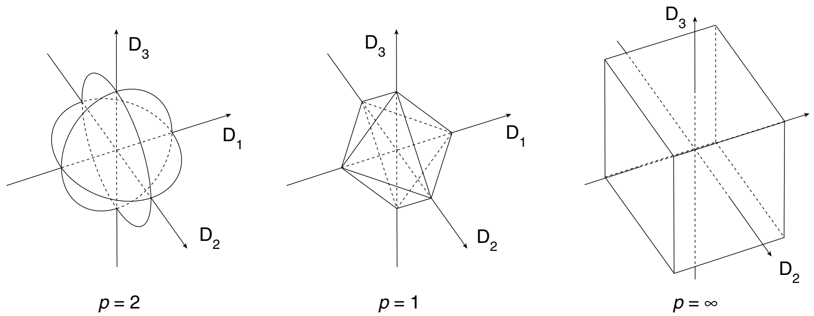

FIGURE 17.5. Unit balls in 3D for the Euclidean, the city-block, and the dominance metric, respectively.

for an overall multiplicative constant, identical to dominance distances, provided the dimension system is rotated by 45°. The converse is also true. Hence, given some MDS configuration that is perfect for $p \to \infty$, it must also be perfect for $p = 1$, and vice versa, because the Stress is the same for two sets of distances that differ by a multiplicative constant only.

The close relationship between city-block distances and dominance distances holds, however, only for 2D. In 3D, the unit circles become unit balls, and Figure 17.5 shows that these balls look quite different for $p = 1$ and $p = \infty$. The city-block ball has, for example, six corners, and the dominance ball has eight corners. The two types of distances therefore cannot be related to each other by a simple transformation and a stretch, as is true for the 2D case.

For given 2D configurations, Stress is, moreover, almost equal for distances with p-exponents of p_1 and $p_2 = p_1/(p_1 - 1)$ (Wender, 1969; Bortz, 1974). For example, for $p = 1.5$ and $p = (1.5)/(1.5-1) = 3$, the Stress values should be nearly equal. The geometrical reasons for this *quasi-equivalency* have been studied in detail by Wolfrum (1976a).

Furthermore, Stress may also be somewhat misleading. Consider the following case (Borg & Staufenbiel, 1984). For a given configuration, the distances are greatest for $p = 1$. When p grows, all distances that relate to line segments not parallel to one of the dimensions drop sharply in size. They continue to drop monotonically, but reach asymptotic values for larger ps ($p > 10$, say). As long as these size functions over p do not intersect, one obtains intervals of rank-equivalent distances over p (Wolfrum, 1976b). Yet, one should not expect that Stress (for nonperfect solutions) is equal for each p within such an interval, because the variance of the distances generally shrinks substantially if p grows. This makes it easier to fit a monotone regression function, and, hence, Stress tends to become smaller with greater p. Nevertheless, the existence of rank-equivalent intervals means that there is no unique optimal p-value but rather intervals of ps that are all equally

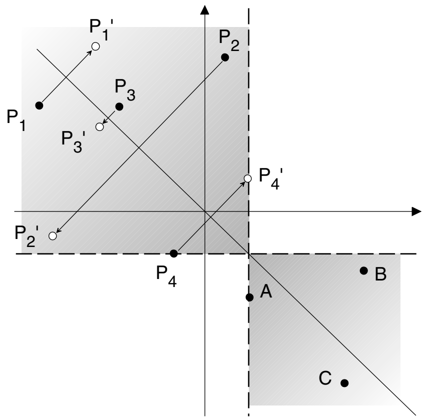

FIGURE 17.6. Demonstration of an indeterminacy in a city-block plane (after Bortz, 1974).

good, even though Stress would, in general, not allow one to diagnose this situation correctly.

On the other hand, there is also an opposite trend that makes low-Stress MDS solutions more likely when $p = 1$ than when $p = 2$, for example. To see this, consider the four corner points of the diamond curve in Figure 17.3. One can readily verify that the city-block distances between these points are all equal, whereas the Euclidean distances form two different classes. Thus, if $p = 1$, four points can be represented in a plane so that all possible distances among them are equal; but if $p = 2$, this is only true for the three corners of an equilateral triangle. Because making distances equal would reduce Stress, such solutions are systematically approximated over the iterations (Shepard, 1974). These effects become more and more pronounced as p approaches the extremes 1 and ∞. Shepard (1974, p. 404) concludes, therefore, that "while finding that the lowest Stress is attainable for $p = 2$ may be evidence that the underlying metric is Euclidean, the finding that a lower Stress is attainable for a value of p that is much smaller or larger may be artifactual."

Interpreting Non-Euclidean MDS Spaces

It has been suggested that the problem of finding the true p-value empirically is easier to solve if other criteria, especially the solution's interpretability, are also taken into account. However, interpreting non-Euclidean Minkowski spaces requires much care. Things are not always what they seem to be, for example, a circle in a city-block space looks like a square. In addition, it can happen that for $p = 1$ and $p \to \infty$ the configurations are indeterminate in peculiar ways. Bortz (1974) reports some examples of *partial isometries*, that is, transformations that preserve the distances within a point configuration while substantially changing the configuration itself. Consider Figure 17.6. If we reflect all points labeled by capital Ps on the diagonal line, we find that the city-block distances of their images (primed Ps) to any point in the shaded region in the lower right-hand corner are exactly the same as before. Hence, either configuration is an equally good data representation, although they may suggest different substantive interpretations. For $p = 2$, no such partial isometries exist in general.

Robustness of the Euclidean Metric

Is the Euclidean metric robust if incorrect? That is, is it likely that MDS closely approximates a true configuration defined by non-Euclidean distances if the scaling is done with $p = 2$? Shepard (1969) concluded from simulation studies using as proximities non-Euclidean distances and even *semi-metrics* (measures that satisfy only nonnegativity and symmetry, but not the triangle inequality) that the true underlying configuration could be recovered almost perfectly with $p = 2$.

This successful recovery of the original configuration using $p = 2$, however, may be partially attributed to the large number (=50) of points in 2D so that the points' locations were highly restricted. The circular isosimilarity contour of the Euclidean distance then is a good approximation to the isosimilarity contours of other Minkowski metrics (see Figure 17.3).

There are no systematic studies that allow one to predict under what conditions the Euclidean metric is robust and when it is not. However, using the Euclidean metric if, say, the city-block metric is true may lead to erroneous conclusions. Consider the following case. Lüer and Fillbrandt (1970), Lüer, Osterloh, and Ruge (1970), and Torgerson (1965) report empirical evidence that similarity judgments for simple two-dimensional stimuli (such as one-spoked wheels) seem to be perceived in an "over-determined" (3D) psychological space. That is, the psychological space seemed to contain additional and redundant dimensions. However, when scaling the data with $p = 1$ rather than with $p = 2$, the underlying physical space is clearly recovered (Borg, Schönemann, & Leutner, 1982). Taking a closer look reveals that using $p = 2$ warps the city-block plane by pulling two of its "corners" upwards and pushing the two other corners downwards along the third dimension.

17.4 The Psychology of Rectangles

We now consider a classic case using MDS as a model of judgmental behavior. In this model, the Minkowski distance formula is taken as a theory of how a dissimilarity judgment on two stimuli is generated. The choice of the

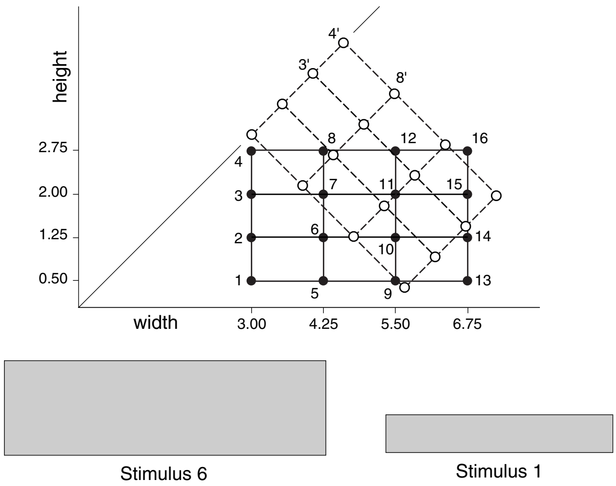

FIGURE 17.7. Design for two sets of rectangles by varying width and height (upper panel). As an example, stimuli 6 and 1 are shown (lower panel).

particular p-values is decided a priori on theoretical grounds. Two different dimension systems appear natural, so that we have to decide empirically which one is the more appropriate.

Two-Dimensional Models for Rectangle Perception

The stimuli here are rectangles. A particular design for rectangles is given in Figure 17.7. It defines two sets of rectangles, characterized by the grid of 16 solid points connected by solid lines and the rotated set of 16 open points connected by dashed lines. The first set is called the width $\times$ height (WH) design, because it is orthogonal to the width and height dimensions. In other words, for each level of width, there are rectangles of all height levels. Note that for all rectangles it holds that their width exceeds their heights.

The dashed grid is orthogonal to the WH system rotated by 45°. The point coordinates on this system can be computed from the width $\times$ height system as width + height and width − height (multiplied by a constant). Psychologically, these dimensions represent something like size and shape (SS). (If width and height are rescaled logarithmically, then size becomes area.) The SS system represents an alternative model for the perception of rectangles.

Borg and Leutner (1983) randomly assigned 42 subjects to two groups of 21 persons each, one group judging the SS rectangles and the other the WH stimuli. Each subject rated all possible 120 stimulus pairs twice on a scale with end categories 0=equal, identical, and 9=extremely different.

TABLE 17.2. Dissimilarity ratings for rectangle pairs; row and column numbers correspond to rectangle numbers in Figure 17.7; ratings averaged over all subjects and replications in WH group (lower half) and in SS group (upper half).

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1		2.05	2.64	3.31	4.93	4.31	4.60	5.79	6.50	6.55	6.19	5.52	8.00	6.98	6.79	7.14
2	4.33		2.12	2.71	4.71	4.69	4.43	4.98	6.40	5.98	5.81	5.71	8.14	6.95	6.76	6.79
3	6.12	4.07		1.79	5.40	5.07	4.36	4.24	6.93	6.29	5.98	5.71	8.17	7.40	6.76	6.71
4	7.21	5.62	3.24		6.36	5.83	4.88	4.31	7.14	6.52	5.71	5.79	8.67	7.69	7.17	6.40
5	2.38	5.76	7.12	7.57		3.17	4.19	4.57	3.52	3.79	3.69	4.95	6.33	5.67	5.29	4.69
6	4.52	2.52	5.48	6.86	4.10		3.43	3.93	4.12	3.57	3.74	3.60	6.62	5.76	5.31	4.90
7	6.00	4.52	3.38	5.21	6.10	4.31		3.43	5.64	4.07	3.48	2.98	7.26	5.83	5.64	5.26
8	7.76	6.21	4.40	3.12	6.83	5.45	4.00		5.55	4.45	3.71	3.64	6.95	5.98	5.24	5.00
9	3.36	6.14	7.14	8.10	2.00	4.71	6.52	7.71		2.86	4.45	5.79	4.14	3.02	3.00	4.57
10	5.93	4.24	6.07	6.93	5.00	2.81	5.43	5.67	4.38		2.86	4.17	4.50	3.48	3.05	3.17
11	6.71	5.60	4.29	5.90	6.86	4.50	2.64	5.21	6.26	3.60		3.31	5.52	3.83	3.40	2.50
12	7.88	6.31	5.48	5.00	7.83	5.55	4.43	2.69	7.21	5.83	3.60		5.95	5.17	3.88	3.55
13	3.69	6.98	7.98	8.45	2.60	5.95	7.69	7.86	1.60	4.31	6.95	7.43		2.38	4.29	5.43
14	5.86	4.55	6.64	7.17	4.86	2.88	5.40	6.50	4.14	1.19	3.79	5.88	4.17		2.64	3.81
15	7.36	5.88	4.55	6.79	6.93	4.50	3.50	5.55	5.95	3.95	1.48	4.60	6.07	4.02		2.74
16	8.36	7.02	5.86	5.40	7.57	5.86	4.52	3.50	6.86	5.17	3.71	1.62	7.07	5.26	3.45	

The resulting proximities, averaged over all 21 subjects in each group, are shown in Table 17.2.

An ordinal MDS representation of the WH data is given by the solid points in Figure 17.8. Because the city-block metric was used, the coordinate axes cannot be rotated without adversely affecting Stress. The MDS result thus suggests that the solid grid of the physical space (Figure 17.7) was transformed into the MDS representation by simple rescalings of the width and height dimensions. These rescalings are such that the physical units decrease more on each dimension the more one moves away from the origin. Thus, perceptually, physically constant increments of an attribute affect the overall impression of similarity increasingly less the more the rectangle already possesses this attribute. This suggests that the psychophysical rescalings might follow the Weber–Fechner law, which postulates a logarithmic correspondence of psychological and physical units. Indeed, the design configuration (grid of solid points in Figure 17.7) can be rescaled in this way to closely fit the MDS representation (grid of open squares in Figure 17.8). Thus, it seems that the subjects in the WH group judged the dissimilarities of the rectangles by first logarithmically rescaling the width and height dimensions, and then simply adding intradimensional differences over the dimensions. But if this were so, what should be expected for the MDS configuration of the SS data?

If width and height are the dimensions that the subjects attend to, and not size and shape, then the SS design grid in Figure 17.7 should be psychophysically rescaled along the width and height axes. A nonlinear rescaling such as the logarithm would lead to some bending of the design lattice, destroying all right angles. The solid points in Figure 17.9 show the MDS representation for the SS data, together with the logarithmically rescaled SS

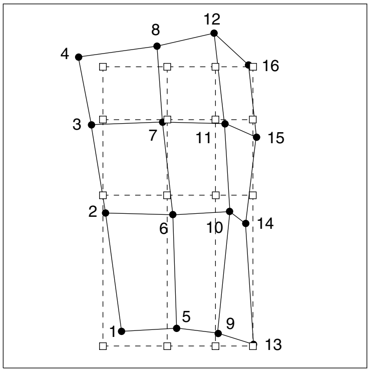

FIGURE 17.8. City-block MDS configuration (solid points) of data of WH group in Table 17.2 with fitted physical WH configuration (squares).

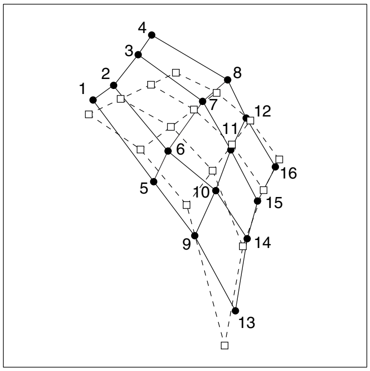

FIGURE 17.9. City-block MDS configuration (solid points) of data of SS group in Table 17.2 with fitted physical SS configuration (squares).

design grid (open squares). One notes that the predictions are not strictly satisfied. In particular, the rectangles in the upper left-hand corner (which look like squares!) seem to involve some further effects. The result suggests, however, that explaining similarity judgments for rectangles seems impossible with size and shape dimensions, because there is no way to explain the bending effects by any rescalings of these dimensions. Rather, a size–shape theory requires additional components such as "dimensional interaction" [for such a theory, see Krantz and Tversky (1975)].

Some Open Questions on Rectangle Perception

These findings are, unfortunately, less simple to interpret than it may appear at first sight. In the following, we give a brief listing of some problem points.

(1) Hubert et al. (1992) reanalyzed the above rectangle data to test the performance of their combinatorial algorithm for MDS. Using 100 random configurations as initial configurations, as well as the physical stimulus space, they found that most solutions for the WH data were similar to Figure 17.8. However, for the SS data, there exist a number of quite different solutions with almost the same Stress. Indeed, based on results from Systat's MDS module, Hubert et al. (1992) concluded generally for gradient-based algorithms that "where one begins is close to where one ends" (p. 234). In particular, it turns out that the SS data can also be explained in a grid that is roughly similar to the SS design configuration.

(2) Staufenbiel and Borg (1987) found similar results for ellipses constructed in WH and SS designs. Using the city-block metric and the Kyst program with an ordinal as well as an interval MDS model, the proximities of both types of ellipses could be explained with low Stress by different

configurations. These configurations were related to the WH or the SS design configuration by monotonic adjustments of the width and height dimension or by the size and shape dimension, respectively. The particular configuration computed by the KYST program was a function of the initial configuration, as observed above. Solutions that were not either roughly WH or SS consistent, however, did not result when random starts were chosen. Using confirmatory MDS to enforce solutions that were perfectly consistent with either a WH or an SS model confirmed that either dimension system allows one to explain the data well.

(3) Schönemann, Dorcey, and Kienapple (1985), Schönemann and Lazarte (1987), and Lazarte and Schönemann (1991) studied whether it makes sense to aggregate individual proximities in the first place. They concluded that such aggregation "was unjustified because distinct strategy groups were found. Some subjects used mainly height and width, others mainly area and shape, and still others mainly shape alone to form their dissimilarity ratings" (Schönemann, 1994, p. 156).

(4) A closer look at the raw data at the subject level also showed that most ratings were *subadditive* in the sense that $\delta(x,y) + \delta(y,z) > \delta(x,z)$. This relation is interesting if x, y, and z differ on one dimension only, because for triples where y lies between x and z one should expect that $\delta(x,y) + \delta(y,z) \approx \delta(x,z)$, provided one takes the data seriously as they come and does not allow for transformations such as adding some constant.[6] Subadditivity is also evident in Table 17.2 for the unidimensional triple $(1,5,9)$, for example, where one finds $\delta(1,5) + \delta(5,9) = 2.38 + 2.00 > \delta(1,9) = 3.36$. This inequality suggests "a ceiling effect. Once the ceiling was removed (by transforming the data with Fisher's z-transformation), most distortions, such as curvature and non-parallelism of lines, markedly diminished" (Schönemann, 1994, p. 156).

(5) δs that satisfy the triangle inequality "can always be modeled as distances. However, because the observed direct dissimilarities are consistently segmentally subadditive along any possible judgment dimension [of the hypothesized systems; our addition], they cannot be modeled as Minkowski metrics because these metrics assume intradimensional additivity" (Lazarte & Schönemann, 1991, p. 144). (This is shown in the section below.)

(6) One may even question the whole notion of a psychological space—in the sense of a metric geometrical space—where all stimuli are represented at the same time and whose distances, after a possible additional transforma-

---

[6]If one admits an arbitrary additive constant (interval scale), then subadditivity becomes less meaningful, because one can at least reduce systematic subadditivity for one-dimensional triples by subtracting a sufficiently large constant from all δs (Attneave, 1950). On an interval scale, all such constants are considered admissible and substantively meaningless. One may question, however, whether it is scientifically wise to eliminate an apparent empirical lawfulness—subadditivity of one-dimensional triples—by such transformations.

tion, define the observed dissimilarities. Lazarte and Schönemann (1991), for example, used simple linear models ("psychophysical maps") to relate the observed dissimilarity to physical dimensions of the observed stimulus pair and describe a strategy that is a function of pair-by-pair comparisons. Restle (1959) and Tversky and Gati (1982), among others, proposed alternative (set-theoretical) models that explain similarity judgments on the basis of the common and the distinctive features of the stimuli.

In summary, one notes that building psychological models via MDS is a difficult and complex undertaking. Early MDS applications tended to be over-optimistic, relying almost exclusively on the global loss, Stress, for answering a whole series of questions—such as the appropriateness of a particular mapping of the data into distances, the dimensionality of the psychological space, the true metric of this space, or the validity of the metric space model as such—all at the same time. This clearly was asking too much from one measure.

17.5 Axiomatic Foundations of Minkowski Spaces

Under certain circumstances, one can study the appropriateness of a multidimensional scaling representation in a way that does not rely on computing this representation and therefore does not depend on minimizing a loss function such as Stress. The approach requires that a theory be given that explains the observed proximities as resulting from an additive combination of dimensional differences. For example, one may hypothesize that similarity judgments on pairs of rectangles can be explained by city-block distances of these rectangles with respect to the physical dimensions width and height. A somewhat less demanding theory might allow for a reasonable psychophysical scaling of the width and height dimensions and for a monotonic function that relates the computed distances to dissimilarity ratings ("response function").

Outside psychophysics, such theories may appear too difficult to formulate. Yet, it is nevertheless worthwhile to study what they imply for MDS, because they provide interesting insights into some of the mathematical properties of MDS representations that are not revealed by mere data fitting. Moreover, to view distances as the image of some underlying composition rule for the basic dimensions of the objects corresponds to a common way of interpreting MDS spaces.

Asking for the conditions that must be satisfied by a set of observations (such as dissimilarity judgments) so that they can be mapped (by an ordinal transformation, say) onto some elements of a particular mathematical system (such as distances of a Euclidean space) is the domain of measurement theory (see, e.g., Krantz, Luce, Suppes, & Tversky, 1971; Schönemann & Borg, 1983). Measurement theorists attempt to specify, first of all, condi-

p	ap	bp	cp	dp
q	aq	bq	cq	dq
r	ar	br	cr	dr
s	as	bs	cs	ds
	a	b	c	d

FIGURE 17.10. An $A \times P$ array.

tions (*axioms*) that must be satisfied by the observations or else the desired representation does not exist (with Loss = 0). Such *necessary* conditions may not be sufficient, that is, they may not guarantee the existence of the model representation, and so one typically asks for conditions that are not only necessary but also *sufficient*. The art of measurement theory is to formulate conditions that are not only necessary and sufficient, but that also can be tested on a *finite* set of data assumed to have a relatively *weak* scale level (such as an ordinal one). It is generally easier to axiomatize an assumed infinite set of data for which no transformation is allowed.

The way one sets up such axiomatic systems is to start with the desired representation and check what properties it implies for observations that can be mapped into this model. So, what are the properties that Minkowski spaces imply for its data? For simplicity, we consider the 2D case only. It represents the most interesting case for psychological modeling and can be easily generalized to higher dimensionality.

Let $A = \{a, b, c, \ldots\}$ and $P = \{p, q, r, \ldots\}$ denote the levels of two design factors, A and P, and let $A \times P$ be the set of all combinations $ap, bp, bq, \ldots$ in the factorial design (Figure 17.10). Assume that dissimilarities are collected for pairs of objects characterized by the cells of this design structure. Under what conditions can such dissimilarities (δs) be interpreted as Minkowski distances computed on dimensions that are some monotonic functions of A and P? This is possible only if the δs possess some general properties.

If the δs are ordinal measures, then any monotone transformation is admissible. Yet, even under such transformations, some properties must hold. For example, distances are always *symmetric* and, thus, δs must be symmetric, because there is no admissible transformation (on any scale level) that would turn nonsymmetric δs into symmetric values. Furthermore, the distance of any point to itself is always 0, and any distance between two different points in greater than zero (*minimality*). For ordinal dissimilarities, symmetry and minimality require that $\delta(x, y) = \delta(y, x) > \delta(x, x) = \delta(y, y)$, for all objects x and y. If the δs do not satisfy this condition, they cannot be represented by Minkowski distances or, indeed, by any other distance.

In the following, we discuss further *qualitative* requirements (i.e., conditions involving only notions of order and equality on the δs) and also some properties that can only be partially tested with ordinal data.

Dimensional Axioms

According to Gati and Tversky (1982), a two-way proximity structure is called *monotone* if the following three conditions are satisfied.

The first condition is called *dominance*:

$$\delta(ap, bq) > \delta(ap, aq), \delta(aq, bq); \tag{17.8}$$

that is, any two-way difference always exceeds its one-way components. The second condition is called *consistency*:

$$\begin{gathered} \delta(ap, bp) > \delta(cp, dp) \quad \text{if and only if} \quad \delta(aq, bq) > \delta(cq, dq), \\ \text{and} \\ \delta(ap, aq) > \delta(ar, as) \quad \text{if and only if} \quad \delta(bp, bq) > \delta(br, bs); \end{gathered} \tag{17.9}$$

that is, the ordering of differences on one dimension is independent of the other dimension. The third condition is called *transitivity*:

$$\begin{array}{llcl} \text{if} & \delta(ap, cq) & > & \delta(ap, bp), \delta(bp, cp), \\ \text{and} & \delta(bp, dp) & > & \delta(bp, cp), \delta(cp, dp), \\ \text{then} & \delta(ap, dp) & > & \delta(ap, cp), \delta(bp, dp). \end{array} \tag{17.10}$$

Condition (17.10) is required to hold also for the second dimension. Transitivity on the δs is equivalent to transitivity of betweenness for the points: $a|b|c$ and $b|c|d$ imply $a|b|d$ and $a|c|d$, where $a|b|c$ means that b lies between a and c (Gati & Tversky, 1982).

The conditions of dominance (17.8), consistency (17.9), and transitivity (17.10) are called *monotonicity* axioms (for a two-way monotone proximity structure) because they specify requirements on the order among the δs.

A more particular property of Minkowski distances is *decomposability*:

$$\delta(ap, bq) = F[g(a, b), h(p, q)], \tag{17.11}$$

where F is a strictly increasing function in two arguments, and g and h are real-valued functions defined on $A \times A$ and $P \times P$, respectively. The arguments g and h are the contributions of the two dimensions to the dissimilarity. If δ is symmetric, g and h satisfy $g(a, b) = g(b, a)$ and $h(p, q) = h(q, p)$. If δ is also minimal, one can set $g(a, a) = 0$ and $h(p, p) = 0$, for all a and p. If g and h can be assumed to be absolute-value functions, then (17.11) can be expressed as *intradimensional subtractivity*:

$$\delta(ap, bq) = F(|X_a - X_b|, |Y_p - Y_q|), \tag{17.12}$$

where X_a and Y_p represent the coordinates of a and p on dimensions X and Y, respectively.

If one assumes that the two dimensions contribute additively to δ, then (17.11) becomes

$$\delta(ap, bq) = F[g(a, b) + h(p, q)], \tag{17.13}$$

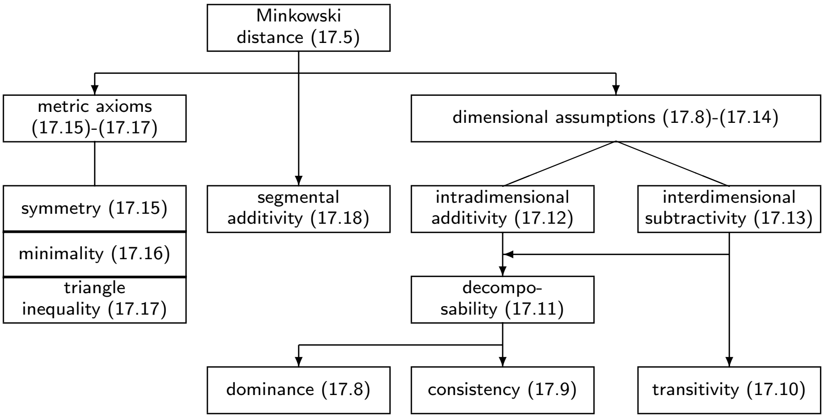

FIGURE 17.11. A hierarchy of conditions necessary for Minkowski distances; entailment is denoted by an arrow.

which is called *interdimensional additivity*. If (17.13) holds, then the actual dissimilarities, not merely their order, are independent of the second dimension, so that:

$$\delta(ap, bp) = \delta(aq, bq) \quad \text{and} \quad \delta(ap, aq) = \delta(bp, bq). \tag{17.14}$$

The conditions (17.8)–(17.14) (sometimes collectively called *dimensional assumptions*) are organized in a hierarchy (Figure 17.11). The diagram shows, for example, that (17.11) implies both (17.8) and (17.9). All Minkowski metrics imply (17.13) and (17.12).

If axiom (17.8), say, is not satisfied by δs that are, by construction or by hypothesis, related to an $A \times P$ design, then these δs cannot be modeled by any Minkowski metric that operates on A and P. This does *not* rule out that the δs can be represented by a Minkowski metric computed on dimensions other than A and P in the same space and/or in a higher-dimensional space. (Indeed, *any* $\delta(i, j)$s, $i < j$, can be represented by Euclidean distances in $n - 2$ dimensions, where n is the number of objects. See Chapter 19.)

Distance Axioms

In addition to the dimensional assumptions, it must also be possible to map the δs into distances d. Distances satisfy three conditions, the *metric axioms*. Two of them, symmetry and minimality, were already discussed above, but are repeated here for completeness. For any points x, y, and z,

$$d(x, y) = d(y, x) \quad (\textit{symmetry}), \tag{17.15}$$

$$d(x, y) > d(x, x) = d(y, y) = 0 \quad (\textit{minimality}), \tag{17.16}$$

$$d(x, z) \geq d(x, y) + d(y, z) \quad (\textit{triangle inequality}). \tag{17.17}$$

Axioms (17.15)–(17.16), in practice, are almost never testable, simply because one rarely collects a complete matrix of δs. Axiom (17.17) can be trivially satisfied in all MDS models that allow at least an interval transformation of the data: one simply determines the triangle inequality that is violated most and then finds a constant c that, when added to every δ in this inequality, turns the inequality into an equality; the same c is then added to every δ, an admissible transformation for interval-scaled δs.

Segmental Additivity Axiom

Minkowski distances assume a dimensional structure that restricts the choice of such additive constants c, because the triangle inequality becomes an equality for points that lie on a straight line in psychological space. That is, for any three points x, y, and z that are ordered as $x|y|z$ on a straight line (such as a dimension), *segmental additivity* is satisfied:

$$d(x, z) = d(x, y) + d(y, z). \tag{17.18}$$

Minkowski Space Axioms in Practice

Tversky and Krantz (1970) have shown that segmental additivity in conjunction with the dimensional assumptions and the metric axioms imply the Minkowski distance. If one wants to test the dimensional conditions (17.8)–(17.14) on real (2D) data, one has to specify the $A \times P$ structure that supposedly underlies the δs (see, e.g., Krantz & Tversky, 1975; Tversky & Gati, 1982; Schönemann & Borg, 1981b).

Staufenbiel and Borg (1987) tested some of these conditions for ellipses constructed in designs analogous to the above WH and SS designs for rectangles. Their data are interesting because they also collected similarity judgments on pairs of identical stimuli, which allow one to test the minimality requirement. It was found that minimality was satisfied for data aggregated over subjects in the sense that $\delta(i, i) < \delta(i, j)$, for all $i \neq j$. Tests of the triangle inequality showed marked subadditivity. Subadditivity correlated highly with violations of minimality on the subject level: these subjects seemed to avoid using the category "0 = equal, identical" on the rating scale, thus, in effect, always adding a positive constant to each δ. Tests of the equality requirements (17.14) showed that they were satisfied in only 20% of the cases. However, the violations revealed no particular systematic pattern and, thus, could be explained as largely due to error.

17.6 Subadditivity and the MBR Metric

Subadditivity of dissimilarities is a frequently observed phenomenon. If the δs are judgments on a rating scale, there are various ways to explain why $\delta(x,y)+\delta(y,z)>\delta(x,z)$ might occur even for triples (x,y,z) that differ on one dimension only. One possibility was offered by Staufenbiel and Borg (1987), who argue that respondents tend to stay away from the lower bound of the scale, thus in effect adding a positive constant to all distance estimates (see item (2) in Section 17.4). Another, or possibly additional, explanation concentrates more on the upper bound, which makes it impossible for the respondent to generate huge dissimilarities. Thus, if $\delta(x,y)$ is rated as quite different, and $\delta(y,z)$ is also rated as quite different, then the respondent tends to run out of possibilities to properly express the extent of the difference of x and z. Because of upper response bounds, "the subject therefore has to contract his response in a continuous fashion, more so for larger than for smaller arguments" (Schönemann, 1982, p. 318). Even with unbounded response scales, subjects typically underestimate large differences (Borg & Tremmel, 1988). The MBR metric (*metric for monotone-bounded response scales*) proposes a hypothesis on how numerical dissimilarities—not just some monotone transformation of them—might be generated under such upper-bound conditions. Let us consider the 2D case and assume that u is the upper bound. The MBR metric of Schönemann (1982) predicts that, given two stimuli, x and y, and given their differences on the *physical* dimensions, Δ_1^* and Δ_2^* (measured in the metric of the observations), it holds that

$$\delta(x,y)=d_M^*(x,y)=\frac{\Delta_1^*+\Delta_2^*}{1+\Delta_1^*\Delta_2^*/u^2},\quad 0\le d_M^*\le u. \tag{17.19}$$

The numerator of the composition rule on the right-hand side of this formula is the city-block metric. The denominator is a contraction factor that ensures that the distance of x and y does not exceed the upper bound u when either Δ_1^* or Δ_2^*, or both, are close to it. This upper bound may be experimenter-imposed ("Please tell me the dissimilarity on a scale from 0 to 9."), but it may also be self-imposed by the subjects (e.g., as a consequence of their laziness to generate best-possible answers) or imposed by nature (e.g., in form of limitations of the subjects' cognitive capacities). The proper value for u is therefore open to some experimentation. Simple specifications for u in practice are to set it equal to the greatest category of the response scale or to the greatest observed dissimilarity. However, Lazarte and Schönemann (1991) found that "within subjects, the MBR with a slightly reduced upper bound was optimal in restoring additivity among collinear points" (p. 144). Formally, one notes that "permitting [u] to vary across subjects, one obtains a one-parameter family of subject-specific MBR's" (Schönemann et al., 1985, p. 6).

To apply the MBR in practice, one first expresses all observations relative the upper bound u (which need not be the same for all subjects). Dividing the dissimilarities by u, formula (17.19) simplifies to a standardized version,

$$\delta(x,y) \quad = \quad d_M(x,y)=\frac{\Delta_1+\Delta_2}{1+\Delta_1\Delta_2},\quad 0\le a,b,d_M\le 1. \tag{17.20}$$

This formula can be further simplified by applying the hyperbolic tangent transformation (Schönemann, 1983). This yields

$$\begin{aligned} d_M &= (\Delta_1 + \Delta_2)/(1 + \Delta_1\Delta_2) \\ &= [\tanh(u) + \tanh(v)]/[1 + \tanh(u)\tanh(v)] \\ &= \tanh(u + v), \end{aligned} \tag{17.21}$$

where $\Delta_1 = \tanh(u)$ and $\Delta_2 = \tanh(v)$. Hence,

$$\tanh^{-1}(d_M) = u + v. \tag{17.22}$$

This offers a way for testing the model: preprocessing the given dissimilarities by applying the inverse hyperbolic tangent should "linearize" the data (expressed as proportions to some upper bound such such as the greatest category on the response scale[7]) so that they can be explained by a simple city-block distance

The MBR metric may strike one as a rather odd composition rule. Should one understand it as a model for how dissimilarity judgments are actually generated? Schönemann (1990) suggests that subjects first compute a city-block metric and then do some contraction to fit it into the bounded rating scale. However, he adds: "We do not expect subjects to do this literally, but we know they must make some contracting adjustment if they want to use the simple city-block addition rule" (p. 154).

One may want to think of alternatives to the MBR metric that seem more plausible as composition rules. One example is the rule

$$f(x, y) = \Delta_1 + \Delta_2 - \Delta_1\Delta_2, \quad 0 \le \Delta_1, \Delta_2, f(x, y) \le 1. \tag{17.23}$$

This function yields values that are very similar to MBR distances but always somewhat smaller. But what are the formal properties of these composition rules? One property that can be proved is that

$$\max(\Delta_1, \Delta_2) \le f(x, y) \le d_M(x, y) \le \Delta_1 + \Delta_2, \tag{17.24}$$

so that the two composition rules lead to values that lie between the two extreme metrics of the Minkowski family, the dominance distance and the city-block distance.

Formally, though, the MBR distance has some nice additional properties. Circles in the MBR plane have a peculiar resemblance to circles in different Minkowski planes. Namely, circles with small radius closely resemble city-block circles (see Figure 17.3), and the larger the radius, the more they

[7]Note that the value for the upper bound b must be chosen such that all distance estimates fall into the half-open interval $[0, 1)$. This is required to make sure that the tangent function exists everywhere. Hence, one proper choice for b is $\max(\delta) + \varepsilon$, where ε is "a small constant" (Schönemann et al., 1985).

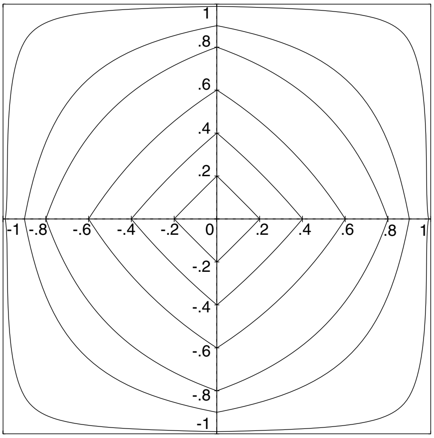

FIGURE 17.12. Circles in the MBR plane, with radii 0.20, ..., 0.90, 0.99.

approximate Euclidean circles, before they asymptotically tend towards dominance circles. This is shown in Figure 17.12. Thus, the shape of a circle in the MBR plane depends on its radius. MBR distances, in other words, emulate the various Minkowski distances depending on the size of the distance: relatively small MBR distances are like city-block distances, large MBR distances are like dominance distances, and intermediate MBR distances are like Euclidean distances.

But does the MBR distance solve the subadditivity problem? Tversky and Gati (1982) report evidence that shows that subadditivity may not affect all triples in a set of objects to the same extent. Dissimilarities were collected for a series of simple 2D stimuli, using different methods of assessment. Three types of stimulus triples were distinguished: corner triples [such as (ap, aq, bq) in Figure 17.10], unidimensional triples [such as (ap, aq, ac)], and two-dimensional triples [such as (ap, aq, ar)]. In uni- and two-dimensional triples, all stimuli differ on the same number of dimensions. Geometrically, such triples lie on a straight line in the design space (collinear points). In corner triples, two pairs differ on one dimension only, and one pair on both dimensions. City-block distances are additive in all cases, but Euclidean distances are subadditive for corner triples and additive for collinear triples. (Under nonlinear transformations of the dimensions, unidimensional triples, in any case, remain collinear.) The observed dissimilarities, then, were almost additive for corner triples, but clearly subadditive for both uni- and two-dimensional triples.

It could be argued that the data are only interval-scaled so that they can be admissibly transformed by adding a constant, for example. Indeed, by subtracting an appropriate constant from all dissimilarities, one could produce values that are, more or less, segmentally additive for the collinear triples. This transformation would, however, make the corner triples *super-*

additive, so that the direct dissimilarity between two points such as *ap* and *bq* in Figure 17.10 becomes *larger* than the sum of its intradimensional differences, which is impossible to model by any distance function. MDS analyses with the program KYST thus showed the least Stress for $p < 1$ for all data sets (except for "color"). With $p < 1$, however, the Minkowski formula (17.5) does not yield a distance. Tversky and Gati (1982) took this finding as supportive for a nongeometric ("feature-matching") model of psychological similarity.

17.7 Minkowski Spaces, Metric Spaces, and Psychological Models

In summary, one may question the ultimate validity of Minkowski spaces for modeling psychological similarity. Indeed, even the much wider class of metric spaces (i.e., sets with distance functions that relate their elements) may be inappropriate, because dissimilarities may systematically violate the symmetry requirement, for example. In this situation, one has four alternatives: (a) give up distance models altogether, as Tversky and Gati (1982) and Gati and Tversky (1982) recommend; (b) modify the distance models by additional notions to make them more flexible (see, e.g., Krumhansl, 1978); (c) possibly drop the restriction to Minkowski spaces and also consider other geometries such as curved spaces (see, e.g., Lindman & Caelli, 1978; Drösler, 1979); and (d) study the conditions under which Minkowski models are likely to be bad or good models of similarity.

The last route is, in fact, necessary for any modeling attempts, because no model is valid without bounds. In this sense, research by Tversky (1977) is relevant. He reports some examples, conditions, and set-theoretical models that allow one to predict when the general distance axioms can be expected to be violated in dissimilarity judgments. For example, symmetry should not hold if one object is a prototype and the other one a variant of this prototype, just as an ellipse is an "imperfect" circle. In that case, the variant should be judged as more similar to the prototype than vice versa. The triangle inequality should be violated if the similarity judgments are based on different criteria. For example, although Jamaica may be judged similar to Cuba, and Cuba is seen as similar to Russia, Jamaica is not seen as similar to Russia at all. (The criteria of similarity in this example could be geographic closeness in the first case and political alignment in the second.) In spite of such counterexamples, the distance axioms are often satisfied in practice. The counterexamples suggest conditions when this should not be the case.

More generally, such fine-grained studies into the foundations of MDS as a psychological model show how one could proceed in cumulative theory building, beginning with exploratory studies on convenient stimuli such as nations (see Chapter 1), proceeding to efforts where one explicitly models judgments for well-designed stimuli such as rectangles, and finally turning to the axiomatic foundations of a particular model.

Studying well-designed stimuli does not have to limit itself to simple contrived stimuli such as rectangles, for example. Steyvers and Busey (2000) study similarity ratings on extremely complex stimuli, namely faces. They comment on the method to collect global ratings of similarity on pairs of faces and then analyzing these data as follows: "The resulting MDS solutions ... can give valuable insights about the way faces are perceived, and sometimes form a useful basis for modeling performance in recognition and/or categorization tasks" (p. 116). However, "this approach explicitly ignores the physical representation of the features comprising the faces. In this purely top-down approach, the multidimensional representations are sometime difficult to relate back to the physical stimulus" (p.116). To remedy this problem, they suggest a complementary *bottom-up approach*, which offers a way to predict the usual similarity ratings for faces on the basis of studying, via MDS, the structure of proximities derived from a large number of physical measurements on these faces (e.g., eye width, eye separation, or nose length), possibly even the vectors containing the light intensities of all the pixels of an image of each face. Using this methodology, they conclude, for example, that facial adiposity (from narrow and skinny to wide and pudgy) and age (from young to old) are major dimensions of the perceived similarity of faces.

17.8 Exercises

Exercise 17.1 Consider the data in Table 1.4 on p. 12.

(a) Repeat the two-dimensional MDS analysis that led to Figure 1.7 using an ordinal MDS approach and city-block distances.

(b) Repeat the MDS analysis using an explicit starting configuration with coordinates as shown in Figure 1.6. Compare the solutions with and without an external starting configuration. Discuss using such an external starting configuration. Is it justified?

(c) Repeat the MDS analysis with $p = 2$. Compare the $p = 2$ solution to the one computed with the city-block metric both in terms of the configuration and in terms of the Stress value.

(d) Specify the set of admissible transformations for the city-block and the Euclidean solutions.

Exercise 17.2 Consider Table 17.2 on p. 374.

(a) Check the dissimilarity ratings in the lower-half matrix for subadditivity and find the intradimensional triple and the corner triple that violate subadditivity most.

(b) Apply the MBR theory to these data. For this you first have to transform the data so that they lie in the half-open interval [0,1). One reasonable way of doing this in this particular case is to divide all values by the maximal value of the rating scale (i.e., by 9). Then, use the inverse hyperbolic tangent function. Finally, check whether the transformed data can be represented in a 2D city-block plane with lower Stress than without this transformation, using linear MDS in both cases.

(c) Plot the original dissimilarity ratings from Table 17.2 against the transformed data. Describe the effect of the hyperbolic tangent transformation on the values.

(d) Discuss the transformation that maps the dissimilarities into the half-open interval [0,1). This mapping expresses the original dissimilarities as proportions relative to an upper bound b. Dividing the dissimilarities by the greatest observed dissimilarity value does not strictly achieve a mapping into the half-open interval $[0, 1)$. The upper bound value b must at least be "slightly" greater than the greatest dissimilarity. Why? (Hint: Note the "open" in half-open!)

(e) Discuss the consequences of choosing a relatively small upper-bound value b or a huge value for b, where "small" and "huge" means "relative to the size of the dissimilarities." How do such choices of b affect the following hyperbolic tangent transformation?

(f) Experiment with a few different choices for upper bounds b that are slightly greater (say, 0.1 to 0.000001) than the greatest observed dissimilarity. Test out how such different choices of b affect the MDS solutions of the rescaled data (see Borg & Staufenbiel, 1986).

(g) Check whether the dissimilarities in Table 17.2 provide evidence that subadditivity affects corner triples, unidimensional triples, and two-dimensional triples in the sense of Tversky & Gati to a different extent.

Exercise 17.3 Consider the data matrix below (Schönemann et al., 1985). It shows relative dissimilarity ratings (averaged over 20 subjects) for nine different rectangles. The physical width–height design characteristics (in cm) of the rectangles are shown in the first two columns.

Width	Height	No.	1	2	3	4	5	6	7	8	9
2.7	3.1	1	0	0.388	0.491	0.405	0.613	0.771	0.649	0.769	0.865
5.4	3.1	2	0.388	0	0.305	0.660	0.466	0.527	0.749	0.630	0.752
8.1	3.1	3	0.491	0.305	0	0.802	0.655	0.369	0.849	0.777	0.585
2.7	5.4	4	0.405	0.660	0.802	0	0.508	0.669	0.358	0.583	0.757
5.4	5.4	5	0.613	0.466	0.655	0.508	0	0.397	0.594	0.447	0.530
8.1	5.4	6	0.771	0.527	0.369	0.669	0.397	0	0.777	0.608	0.369
2.7	8.1	7	0.649	0.749	0.849	0.358	0.594	0.777	0	0.474	0.660
5.4	8.1	8	0.769	0.630	0.777	0.583	0.447	0.608	0.474	0	0.377
8.1	8.1	9	0.865	0.752	0.585	0.757	0.530	0.369	0.660	0.377	0

(a) Plot the design space of the rectangles. Sketch the nine rectangles.

(b) Scale the dissimilarities with and without a rational starting configuration. What evidence do you find that the respondents generated their dissimilarities from a width–height dimension system?

(c) Check the dissimilarities for subadditivities.

(d) Preprocess the data by the MBR logic and then repeat the MDS scalings. Do you find theoretically interesting differences?

Exercise 17.4 Consider the data in Table 1.4 on p. 12. Theoretical considerations suggest that they were generated by city-block composition of two intradimensional differences. Observe what happens when you scale these data with the "incorrect" Euclidean distance in 2D and in 3D, using the design configuration in Figure 1.6 as a starting configuration.

Exercise 17.5 Construct a grid of points in the plane (as in Figure 19.3, e.g.) and measure their city-block distances. Then scale these distances in Euclidean 3D space, using

(a) ordinal MDS,

(b) interval MDS, and

(c) classical scaling.

Carefully study the resulting configurations in the planes spanned by the principal components. Discuss the effects of using the improper Euclidean distance function with MDS models that allow for arbitrary monotone transformations, linear transformations, and ratio transformations of the data, respectively.