THE PROBLEM OF COMBINING INTEGRAL AND SEPARABLE DIMENSIONS

Mikael Johannesson

Lund University Cognitive Science
Kungshuset Lundagård
S-222 22, LUND, Sweden
and
Department of Computer Science
University College of Skövde, Box 408
S-541 28 SKÖVDE, Sweden

mikael.johannesson@ida.his.se

Abstract: For geometrical models of cognition, the notion of distance rules - or metrics - is fundamental. Within psychology, it is well established that pairs of dimensions that are processed holistically - *integral dimensions* - normally combine so as they are best described with a Euclidean metric, whereas pairs of dimensions that are processed analytically - *separable dimensions* - most often combine with a city-block metric. The experimental tradition studying information integration has typically been limited to two-dimensional stimuli. A next step is to study information integration when dealing with more complex stimuli. This step give rise to several interesting questions regarding information integration behaviour, especially when both integral and separable pairs are included. For example: How do we integrate information when both integral and separable pairs are involved? This paper extends earlier research regarding information integration in that it deals with stimuli with more than two dimensions, and with complex stimuli consisting of both dimensional pairs previously identified as holistic, and dimensional pairs previously identified as analytical. The general pattern identified is that information integration can be more accurately described with a rule taking aspects of stimuli into consideration compared to a traditional rule. For example, it appears that combinations of analytical and holistic stimuli, are better described by treating the different subspaces individually and then combining these with addition, compared to any single Minkowskian rule, and much better compared to any of the Minkowskian rules traditionally used (i.e. the city-block-, the Euclidean or the dominance-metrics). For stimuli that are subject to confusion (e.g. when stimuli are too similar) single Minkowskian rules appears to describe the data best - but with more substantial violations against the assumptions of correspondence and interdimensional additivity.

INTRODUCTION

When studying and modelling information integration behaviour and similarity, a spatial metaphor (e.g. Palmer, 1978) is often adopted as the underlying framework. In such a case, the representing space consists of a number of dimensions, each corresponding to some quality or property (Gärdenfors, 1992, 2000). Objects, or mental objects, are represented by coordinates, and the psychological similarity is reflected by distance relationships between them. In essence, the closer two objects are (i.e. the shorter the distance), the more similar they are.

The most commonly investigated combination rules, or metrics, for describing distances in a multi-dimensional space have been instances of the generalised Minkowski metric which is given by Eq. 1.

$$(1) \quad d(i,j) = \left\{ \sum_{k=1}^{n} | x_{ik} - x_{jk} |^r \right\}^{1/r} \quad ; \; r \geq 1$$

where *d(i,j)* is the distance between object *i* and *j*, x_{ik} refers to the position of object *i* on the *k*th axis and *n* is the number of constituting dimensions.

Since the coefficient *r* in Eq. 1 is not restricted to be discrete, the number of possible instances is infinite, and all of them have different properties. However, three extreme cases can be identified. These are when: r = 1; *the city-block metric* (or *Householder-Landahl metric*) - The distance between two objects simply is the sum of the absolute differences for each of the underlying dimensions; r = 2: *the Euclidean metric* - The distance corresponds to the square root of the sum of the squared differences for each of the underlying dimensions. Compared to the city-block, the Euclidean metric puts less emphasis on increasing number of dimensions where the objects differ; and r = ∞[1] : *the dominance metric* - The distance between two objects is a function of the dimension which have the largest difference (for the particular object pair under consideration). Therefore, as for the Euclidean, but contrary to the city-block metric, the emphasis on the number of differing dimensions is small compared to the magnitude(s).

[1] All calculations with "the dominance metric" presented in this paper are really calculations with Minkowski-r = 50.0. For the purposes of this paper, this sufficiently well contrasts with calculations using the Euclidean- or city-block-metrics.

The three metrics mentioned are generally the most well-known instances of the Minkowski-metric. When it comes to cognitive modelling, the city-block and the Euclidean metrics are the most common, and this is especially true for the Euclidean metric. For example, in AI, the Euclidean metric is often taken for granted. Within the field of cognitive psychology, however, it is well established that some pairs of dimensions combine, with Garner's (1974) terminology, to form *integral dimensions* and others to form *separable dimensions* (see e.g. Garner, 1974, 1977; Gottwald & Garner, 1975). The terms integral and separable refer to a variety of properties of a pair of dimensions. A typical description of an integral pair is that it is processed as holistic, unanalysable, directly and effortlessly by subjects and that the constituent dimensions combine so as to conform to a Euclidean metric; pairs of hue, saturation or brightness of colour (see e.g. Gottwald & Garner, 1975; Hyman & Well, 1967; Kemler Nelson, 1993; Ruskin & Kaye, 1990) and the auditory dimensions of pitch and loudness (Kemler Nelson, 1993) typically do this. The corresponding description for a separable pair is that the constituent dimensions are processed independently by subjects and that they combine so as to conform to a city-block metric, e.g. size and reflectance of squares (Attneave, 1950). The difference regarding the best applying combination rule is well motivated with reference to the properties of the two metrics; information integration according to the city-block rule could be characterised as differentiated whereas information integration according to a Euclidean rule could be characterised as holistic.

Now, the combination rules for integral and separable dimensions are well investigated for dimensional *pairs*. But, what about dimensional triples? Quadruples? Real world objects? How do we integrate information when both integral and separable pairs are involved? The focus here is upon how to describe and predict information integration behaviour. Such knowledge is not only important from a theoretical perspective (e.g. for basic research within the area of similarity based representations and concept formation), but also from a more practical and pragmatic machine learning perspective (e.g. for tuning machine learning algorithms).

Simple parallelograms varying in saturation, brightness, height and tilt could serve as an example. Pairs of the dimensions of colour, i.e. of hue, brightness and saturation, are often used as prototypical examples of integral dimensions (see e.g. Gottwald & Garner, 1975; Hyman & Well, 1967; Kemler Nelson, 1993; Ruskin & Kaye, 1990). Perception of variation in saturation and brightness on a single colour patch have in previous studies (e.g. Hyman & Well, 1967; 1968) been shown to be better described using the Euclidean compared to the city-block metric. The height- (size-) and tilt- dimensions of parallelograms is an example of separable dimensions (Tversky & Gati, 1982). Tversky and Gati found such pairs to be better described using the city-block metric compared to the Euclidean. In some cases, their data suggested a Minkowski-r even somewhat smaller than r = 1.

How, then, could subjects' phenomenological similarity/dissimilarity between parallelograms varying in height, tilt, saturation and brightness be described? With reference to the metric properties of the underlying pairs of dimensions, information integration behaviour may be expected to be best described by a single metric somewhere between the city-block and the Euclidean metrics. With reference to the relative complexity of the stimuli, it could be that subjects focus more heavily on the dimension where the stimuli differ the most - and so the dominance metric could be expected to be the most adequate. It also, again with reference to the different metric properties of the underlying pairs of dimensions, makes sense to divide the stimuli space into two separate subspaces - one describing the aspects of shape of the stimuli (i.e. height and tilt) - *the shape space* - and one the colour aspects (i.e. saturation and brightness) - *the colour space*. In this case it could be that two different metrics should be applied, one for the shape space and one for the colour space. Further, since it is previously known that height and tilt are separable, and saturation and brightness are integral, the city-block metric should apply to the shape space, whereas the Euclidean metric should apply to the colour space. Regarding how the separate subspaces should be combined into a holistic measure, simple addition could be expected. The line of thought underlying this is that the suggested subspaces better fit the description of separability compared to integrality, i.e. they could be processed independently, and separable dimensions are known to be best described using the city-block metric.

In the remainder of this paper, combination rules, or metrics, such that the same Minkowski-r applies to the whole stimuli space, will be referred to as *homogenous* rules. Rules or metrics such that one Minkowski-r, say r_1, applies to one subspace, and one Minkowski-r, say r_2, applies to another, and that the holistic measure is obtained by combining the sub-metrics separably, will in the following be referred to as *heterogeneous* rules or metrics[2].

Now, how could we determine which of the listed alternatives is the best when we want to describe similarity/dissimilarity judgements of parallelograms varying in height, tilt, saturation and brightness?

[2] There are quite few examples of true integral dimensions in the literature (Grau & Kemler Nelson, 1988). This fact does not, however, undermine the possible practical importance of heterogenous models, since perception of many dimensional pairs fall between the endpoints of a continuum of dimensional separability (Smith & Kilroy, 1979; Smith, 1980).

General Method

Finding out the metric for a similarity space is not an easy task, or as Dunn puts it:

> "Despite over 30 years of research, there is no single, agreed upon method for determining the metric of a similarity space."
> (Dunn, 1983, p. 244)

In order to investigate the relationship between dimensional integrality and the combination rule used in a similarity/dissimilarity judgement task, Dunn (1983) adopted an extension of the strategy used by Attneave (1950). The method used in this paper will be in line with the one adopted by Dunn, but generalised in order to deal with stimuli with more than two underlying dimensions.

The basic idea is to divide the set of dissimilarity ratings into unidimensional and bidimensional ratings, reduce them to distances between points in a predefined dimensional space and then determine the Minkowski-r - or just r - that best predicts the bidimensional dissimilarities from the unidimensional ones.

Assumptions

In order to reduce ratings to distances correspondence, interdimensional additivity, intradimensional subtractivity and linearity must be assumed.

Correspondence

The term "correspondence" means here that there should be a correspondence between physical and psychological dimensions. The assumption may be violated if, for example, subjects attend only to a true subset of the physical dimensions used. This can be detected by analysing the dimensional weights assigned by subjects. Ideally, the weights should be non-zero and relatively equal. The assumption will also be violated if subjects encode the stimuli with the use of an alternative dimensional structure. Such a violation is hard to detect, but will in most cases lead to a violation against interdimensional additivity (Dunn, 1983).

Interdimensional Additivity

This assumption means that perceived differences on each of the dimensions must be independent, i.e. a constant difference in one dimension must not be perceived differently depending on the levels of other dimensions.

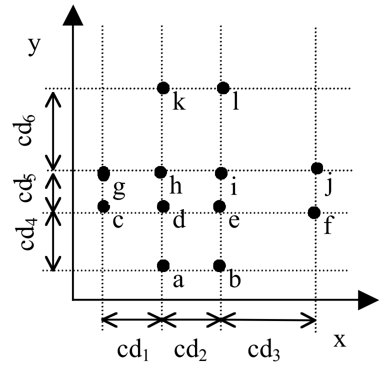

Figure 1: 12 stimuli and their component distances (cd_1-cd_6) in a two-dimensional space. (Based on Figure 1. p. 245 in Dunn, 1983).

In order for interdimensional additivity to hold, each individual component distance (marked with cd_{1-6} in Figure 1) should be constant[3]. Put alternatively, with reference to Figure 1, the perceived difference between the object *c* and *g* should be the same as that between e.g. *i* and *e* (note that symmetry is assumed, and so the order is irrelevant). In the same way, the bidimensional pairs *h-l* and *i-k* should be perceived as equally different.

In order to test that interdimensional additivity is not violated, pairs of judgements are compared and classified as either "+" or "-" differences. Dunn (1983) proposes that unidimensional judgements in line with an augmentation effect (i.e. when the difference increases with the magnitude of the irrelevant dimension) are classified as "+":es, whereas judgements in line with a negative augmentation effect are classified as "-":es. For bi-dimensional differences, if the pair positively correlated with the constituent dimensions is judged to be more dissimilar than the negatively correlated pair a "+" is given, if it is judged to be less dissimilar a "-" sign is given, if equal no sign is given. This principle of classification could be exemplified with reference to Figure 1: If a subject rate *g* and *h* to be more dissimilar than *c* and *d*, this is classified as a "+" difference. A situation like this is visualised in Figure 2 (a) below. If the pair *c* and *d* is rated as less dissimilar compared to *g* and *h*, a "-" difference is noted. The third and last possibility, that *c* and *d* are rated to be as dissimilar as *g* and *h*, does not lead to a classification that could be included in a nonparametric statistical test (Dunn, 1983), and no sign is given. In exactly the same way, if *d* and *h* are experienced as more dissimilar than *c* and *g*, a "+" is given. If less dissimilar, a "-" is given and in the case of no difference, no sign is given. The corresponding "-" situation is visualised in Figure 2 (b). When it comes to bidimensional pair comparisons like comparing the differences between *c* and *h* versus *d* and *g*, respectively, a "+" sign is given if *c* and *h* are rated as more dissimilar compared to *d* and *g*. A

[3] Different component distances (both within and between dimensions) may be of different magnitudes.

situation like this is visualised in Figure 2 (c), and it is easy to see that the constituting dimensions do not meet at a right angle in space (i.e. the point should form a rectangle), which is a requirement for using any Minkowski metric.

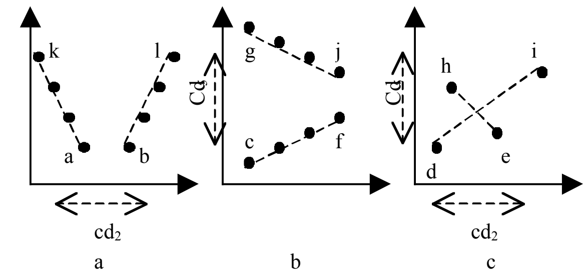

*Figure 2: Visualised examples of violation against interdimensional additivity: (*a*) The component distance cd_2 is positively correlated with the magnitude of the irrelevant dimension y, (*b*) The component distance cd_5 is negatively correlated with the magnitude of the irrelevant dimension* x, *and (*c*) The combination of the component distances cd_2 and cd_5 is positively correlated with the constituent dimensions.*

In order for interdimensional additivity to hold, the number of "+" and "-" differences should be equal for each set of equality tests, i.e. for each combination of constituting dimensions. This could be tested using a two-tailed sign test aimed at finding out whether the proportion of e.g. "+" differences are significantly different from an expected value of .5, or not.

Intradimensional Subtractivity

The distance between two stimuli, differing in exactly one dimension, is a result of a subtraction of the perceived values on that dimension. As an example, suppose we have three stimuli, *d*, *e* and *f* (see Figure 1) differing only in level of one dimension such that *e* has a level between the levels of *d* and *f*. In such a case, intradimensional subtractivity requires that

(2) $d(d,f) = d(d,e) + d(e,f)\,;\ \ r \geq 1$

where *d(x,y)* is the distance between object *x* and *y*.

In other words, any unidimensional distance can be decomposed into the sum of smaller component distances.

The unidimensional distances for the stimuli presented in Figure 1 could be rewritten as

(3) $$d(a,b) = \sum_{i=1}^{6} w_{iab} cd_i \quad ;w_{iab} = 0 \text{ or } 1$$

where *d(a,b)* is the distance between object *a* and *b*, w_{iab} refers to the weight of the component distance cd_i.

Linearity

In Dunn (1983), the function relating dissimilarities to distances is assumed to be linear. Thus, the dissimilarities between the stimuli in Figure 1 above could be expressed as

(4) $$\delta(a,b) = \sum_{i=1}^{6} w_{iab} cd_i + A \quad ;w_{iab} = 0 \text{ or } 1$$

where *δ(a,b)* is the perceived dissimilarity between object *a* and *b*, w_{iab} refers to the weight of the component distance cd_i, and *A* is an additive constant.

Eq. 4 specifies a multiple regression equation in which the weights define a set of dummy variables, the component distances form the regression coefficients and A is the additive constant. The multiple correlation coefficient derived using Eq. 4 can be used for testing the assumptions of both intradimensional subtractivity and linearity: if the assumptions are valid, the square of the multiple regression coefficient will be equal to, or greater than, an estimate of the specific reliability of the unidimensional dissimilarities (c f. (Dunn, 1983)).

Determining the Spatial Metric

Performing a multiple regression analysis, in line with Eq. 4, on unidimensional dissimilarities, provides an estimate of the component distances and the additive constant. From these, it is straightforward to calculate any Minkowski distance or - in this context - estimate any "Minkowski dissimilarity". In order to determine the "best" describing metric for a particular subject, Dunn (1983) compared the mean observed and the mean predicted bidimensional dissimilarity using a certain value of r: overestimation of r lead to underestimation of the observed mean, whereas underestimation of r lead to overestimation of the observed mean. The "best" Minkowski-r will generate a set of estimated/predicted dissimilarities which is not significantly different from the corresponding observed set.

Methodology Adopted

The basic methodology outlined by Dunn will be adopted here with some exceptions and additions as outlined below. The present paper aims to investigate whether the machine learning community could gain from using different Minkowski metrics for different subspaces rather than a single metric applied to the whole space. Seen from this perspective, the various tests suggested by Dunn, and their results, are not central here. The main reason for carrying out (some of) them at all, is to give the reader an opportunity to get at least some opinion of how well/bad the requirements are met.

Interdimensional Additivity

Especially when dealing with more than two dimensions, the test for detecting violations against interdimensional additivity, as suggested by Dunn (1983), possess some weaknesses in addition to the fact that sign-tests are not sensitive. One is that it is not straightforward to generalise the classification procedure described above to handle stimuli with more than two underlying dimensions. When only two underlying dimensions are used and the unidimensional differences are to be classified, it is completely clear what is meant by "correlation with the irrelevant dimension". In a three dimensional case, however, there are two irrelevant dimensions, and so there are two possible situations for unidimensional cases; a rating could be: 1) positively (negatively) correlated with both irrelevant dimensions, or 2) positively (negatively) correlated with one of the irrelevant dimensions and negatively (positively) with the other. The first of these two situations is clearly in line with an augmentation effect, and the classification could be done as done by Dunn (1983). The second situation, however, is ambiguous, and will not be analysed here. This problem becomes even harder when three or more underlying dimensions are used. Under such circumstances situations as skew proportions of positive versus negative correlations with irrelevant dimensions arise, and classification of such could certainly be discussed. Within the frame of the present paper, however, such situations will not be analysed. The same principle of merely analysing non-ambiguous pairs applies for treatment of pairs differing in from 2 up to N dimensions. However, even though selected pairs are all non-ambiguous, it could be argued that different pairs possess different qualities. For example, pairs with single dimensional differences (e.g. pairs differing only in saturation), are qualitatively different from, for instance, comparisons "between" known dimensional pairs (e.g. height/saturation). It could be argued that eventual violations against interdimensional additivity should be regarded differently depending on to which "category" the difference belong.

Another anomaly that occurs when the number of dimensions increases is that the number of sets to be analysed increases exponentially, something that in turn increases the risk for experiment wise type I error. For example, when four underlying dimensions are used, 15 different tests are needed. One way to deal with this problem could be to apply a Bonferroni correction (see e.g. Keppel, 1991) in which the new per comparison significance level equals the desired family wise error divided by the number of comparisons.

In the test as suggested by Dunn (1983), pairs with no assigned sign (i.e. pairs with no difference) are excluded from the analysis. An intuitive and simple way of decreasing the problems caused by neglecting the comparisons without an assigned sign is rather than setting N to the total number of violations, to distribute the number of pairs without a sign evenly between the "+" - and the "-" - categories and to set N to the total number of comparisons.

In summary: neither the presence nor the "non-presence" of violations against interdimensional additivity found with the test described above, give a true picture of to what extent interdimensional additivity holds.

EXPERIMENT I (PILOT EXPERIMENT)

Subjects

A total of 9 undergraduates at the University of Skövde participated for a payment of 250 SKr (this corresponds to roughly £20 or $30). Also, 3 persons from the authors' circle of acquaintances participated without payment. These three people were originally thought of as pilots in order to find out if something (e.g. instructions) in the experiment needed to be changed. Since this was not the case, they were included in the forthcoming analysis, which means that a number of 12 subjects completed the experiment.

Stimuli

The stimuli varied in four dimensions, height (h), tilt (t), saturation (s) and brightness (b) of a parallelogram (Figure 3).

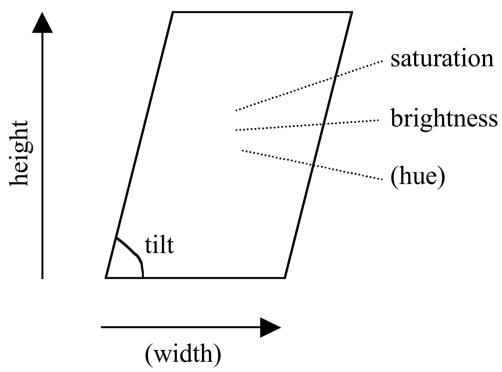

Figure 3: The dimensions "height", "tilt", "width", "saturation", "brightness" and "hue" of a parallelogram.

Each dimension varied in three levels, h: (4, 5 or 6 units of length), t: (40, 50 or 60 degrees), s: (60, 70 or 80% of maximum saturation) and b: (60, 70 or 80% of maximum brightness). The width and the hue of the parallelograms were held constant (4 units of length and 240° of the colour circle, respectively). The complete domain formed according to these specifications consist of the 81 different parallelograms presented in Appendix IA.

Of 81 stimuli, 3240 non-ordered pairs can be formed, a number far too high for a single rating session and it had to be reduced in order to not exhaust the subjects. For the purposes of this paper, it was judged to be more important that subjects made as many different ratings as possible rather than making a smaller number of ratings twice. The advantage of this choice is that it produces a richer material upon which the calculated component distances are based, and against which the polydimensional estimations are evaluated. The loss is that specific reliability coefficients (see section Linearity above) cannot be calculated, and therefore a test of intradimensional subtractivity and linearity cannot be performed.

In order to decrease the number of pairs that where to be rated, 20% (i.e. 648) of the total number of pairs where chosen randomly. Also the order of the pairs were randomised. With this selection the material involves 62 unidimensional, 185 bidimensional, 275 three-dimensional and 125 four-dimensional differences. Two pairs unfortunately became exactly the same due to a typing error, and in the forthcoming analysis one of the ratings of the pair that were rated twice was selected randomly for each subject.

Procedure

The experimental sessions were performed individually in a quiet room with drawn curtains. Except for the subject only the experiment leader was present.

Each subject started with a computerised[4] test, meant to find out if she/he could discriminate between the colours that were to be used in the real experiment. During the test, a two-digit number in one colour appeared at a random place on a background in another colour. In order to pass the test, a subject needed to identify the numbers, for all ordered combinations of colours, correctly.

Having passed the colour test, the experiment leader started the experiment program and asked the subject to read and follow the instructions given on the screen.

The experiment consisted of several phases:

- *Instruction and information-phase*: Subjects were informed that the goal of the experiment was to investigate how people judge similarity/dissimilarity between objects, and that it in this case was about coloured parallelograms. They were instructed to use a 20-graded scale for judging how similar/different they thought the parallelograms in each pair were, and they were informed that larger numbers should correspond to larger difference. Further, they were told that they would be shown all the parallelograms, and that they would go through a short training session, before they started with the judgements.
- *Stimulus presentation*: Diminished versions of all stimuli were presented simultaneously in a randomised layout. The basic reason for presenting this material was that subjects should be able to calibrate their scale faster.
- *Training phase*: Subjects were asked to make similarity/dissimilarity judgements for ten pairs of coloured parallelograms varying in the same four dimensions as the real stimulus material. However, the levels of the dimensions of the training stimuli did not coincide with the levels of the real material. The training phase was there for different reasons, one being that the subjects would get a clearer understanding of what they were supposed to do. During this phase they were allowed to ask questions.
- *Instruction phase*: An instruction phase as above was repeated. This time subjects were also informed that the judgement sessions would be divided into six parts in the following way: 108 judgements - 10 minute break - 108 judgements - 20 minute break - 108 judgements - 60 minute break - 108 judgements - 10 minute break - 108 judgements - 20 minute break - 108 judgements.
- *Stimulus presentation*: Subjects were again presented with the complete stimulus material. The reason for presenting this a second time was that it was believed that subjects would be able to calibrate their scales better after having completed the training phase.
- *Judgement phase*: The 648 stimulus pairs were presented in a random order that was the same for all subjects. With breaks included, the experiment took about 3 hours.

All subjects passed the colour test.

Results and analysis

The stimulus pairs selected, and the response from subjects are jointly presented in Appendix IA.

Correspondence

Table 1 below presents the average component distances (see Figure 1 above) per dimension, and the coefficient of determination for the collapsed data (see Appendix IB for corresponding information over individuals). The average component distances, which could be interpreted as the relative saliency of each dimension (Dunn, 1983), should be non-zero and approximately equal for the assumption of correspondence to be valid.

Table 1: Component distances averaged for each dimension, and coefficient of determination – Experiment I.

Avg_h	**Avg_t**	**Avg_s**	**Avg_b**	R^2
5.338	4.011	1.106	2.245	0.652

By inspecting Table 1 it becomes clear that the collapsed data not is ideal with respect to correspondence. Regarding the approximate equalness

[4] The tests as well as the real experiments was performed on a standard personal computer with a 15-inch colour screen in 32-bit colour mode.

required, it is hard to set an exact limit of what could be accepted; it is rather the question of a continuous scale. What is clear, however, is that the saturation and brightness dimensions have shorter component distances (are less weighted) compared to height and tilt. A probable explanation for this unequal weighting is that subjects perceived the variation in height and tilt as larger compared to the variation in saturation and brightness.

The coefficient of determination is not very large, indicating that a general linear model misses to account for a considerable proportion of the variance of the data.

Interdimensional additivity

All equality tests regarding interdimensional additivity for each subject are presented in Appendix IB, in which rows representing significant deviations from an expected value of .5 by a uncorrected two-tailed sign test with _ = .05 are marked in italics and bold.

For the collapsed data, 2 out of 15 separate tests significantly deviated from the expected value, and thus represented violations against interdimensional additivty.

Even though the number of significant violations against interdimensional additivity would decrease if for example Bonferroni correction should be used, it is clear that the data collected are not perfectly described with any Minkowski metric. It could, however, also be noted that the data does not represent the “worst case” either, since the number of tests showing significant violations (2) are few compared to the number of tests performed (15). With reference to what was mentioned above, it is in the present case not obvious how the results of the equality tests should be interpreted.

In summary, with respect to the validity of the assumptions of correspondence and interdimensional additivity, the data collected are not ideally described with any Minkowski metric. However, there are differences in how well these requirements are fulfilled between subjects (Appendix IB).

Determining the Spatial Metric

When there are just two underlying dimensions, as for the data analysed by Dunn (1983), it is obvious that distances/dissimilarities should be estimated and evaluated for stimuli differing in two dimensions. However, as the number of underlying dimensions increases, so does the number of possibilities. In the present case, when four underlying dimensions were used, stimuli pairs differing in two or more dimensions were analysed.

Justifying the Measure of Error

In order to possibly improve the process of determining the spatial metric, two alternative measures of error for a particular r were contrasted. One was in line with Dunn's method: deviation of the absolute difference between the mean observed dissimilarity and the mean predicted/estimated dissimilarity from the mean observed dissimilarity - in the following referred to as DEV. The other was in line with the measure used by Ronacher (1998) for the same purpose[5]. The latter is here referred to as the mean squared error (MSE), and is defined as

$$(5) \qquad MSE = \left(\sum_{\substack{a,b \\ a \neq b}} \left(\delta(a,b) - \tilde{\delta}(a,b) \right)^2 \right) / N$$

where $\delta(a,b)$ is the perceived, $\delta(a,b)$ is the predicted/estimated - dissimilarity between object a and b, and N is the number of stimuli pairs.

For each of the homogenous rules: city-block, Euclidean and dominance, and all non-ordered combinations of heterogeneous rules, where the subspaces where formed by the city-block, Euclidean or dominance metric[6], the distances between all non-ordered combinations of stimuli were calculated from physical descriptions of the stimuli. By regarding the distances as fictive dissimilarities, and by estimating the dissimilarities as described above for different rules, the errors according to DEV and MSE were calculated. The same subset and physical descriptions as used in the present experiment were analysed. Further, in line with this, the estimated distances/dissimilarities were scaled into a discrete scale ranging from 1 to 20. Since the underlying rule was known in each case, the two alternative measures of error could be evaluated against each other.

For the homogeneous models, both DEV and MSE suggested the same - and correct - underlying model. For the heterogeneous models MSE suggested the correct model in all cases. The use of DEV, however, was clearly systematically ambiguous. In all cases when the underlying model could be described as *metric A applies to subspace 1 and metric B applies to subspace 2*, **both** the correct model and the model such that *metric B applies to subspace 1 and metric A applies to subspace 2*, were suggested. The explanation is that that the sum of absolute deviations for the two models necessarily is the same for a balanced set of stimuli.

In summary, based on this analysis, MSE appear to be the better measure for the purposes of this paper.

---

[5] Ronacher (1998) did not use the number of stimuli pairs in the denominator, but the measure is basically the same as the one used here.

[6] Note that the heterogenous rule where both subspaces are formed by the city-block metric exactly corresponds to the city-block homogenous rule.

Spatial Metric

Candidates for describing the individual subjects' data were evaluated using MSE as the measure of error. In addition to the rules used when evaluating the two error measures (DEV and MSE) above, i.e.

- the homogenous rules: city-block, Euclidean and dominance – in the following referred to as *Hom cit*, *Hom euc* and *Hom dom*, respectively,
- all non-ordered combinations of heterogeneous rules, where each of the subspaces were formed by the city-block, Euclidean or dominance metric – in the following referred to as *Het citeuc*, *Het citdom*, *Het euccit*, *Het euceuc*, *Het eucdom*, *Het domcit*, *Het domeuc* and *Het domdom*, respectively,

errors were calculated for values of Minkowski-r ranging in small discrete steps from r = 1.0 to r = 50.0 applied to the whole stimuli space (the homogenous model with the r giving the lowest error will in the following be referred to as *Hom opt*), the shape subspace and the colour subspace respectively. The heterogenous model where the separately optimized r for the shape space is applied to "shape" and where the separately optimized r for the colour space is applied to "colour", will here be referred to as *Het sepHTsepSB*. Finally, the combination of r:s, one for the shape subspace and one for the colour subspace, when optimised simultaneously with a heterogeneous rule – here referred to as *Het simHTsimSB*.

Even though the variation between individual subjects was relatively extensive (see Appendix IC) Table 2, showing the errors for the models tested for, for the average ratings, captures some general characteristics.

Table 2: Models, associated r:s and errors for average data, sorted after errors – Experiment I.

	r	Err
Hom opt	1.6	3.240
Hom euc	2	4.065
Het simHTsimSB	2.85;50	4.066
Het sepHTsepSB	2.2;50	4.207
Het eucdom	2;50	4.376
Het domdom	50;50	4.572
Het domeuc	50;2	4.909
Het euceuc	2;2	5.017
Het domcit	50;1	6.812
Het euccit	2;1	7.616
Hom dom	50	9.996
Het citdom	1;50	13.545
Het citeuc	1;2	14.892
Hom cit	1	19.096

One is that the optimal homogenous rule (*Hom opt*) gave the lowest overall error, something that was true also for most of the individual cases. In the present case the Minkowski-r for the rule was between the city-block and the Euclidean metrics. Note however, that the r-value (1.6) was closer to the Euclidean compared to the city-block: the Minkowski-r of a rule giving distances halfway between the city-block and the Euclidean metric is not the intuitive 1.5, but rather approximately 1.2. This is the explanation for the large difference wrt the errors for *Hom euc* and *Hom cit*.

For the individual data (see Appendix IC), the corresponding r were between 1 and 2 for seven of the twelve cases, and somewhat larger than 2 for the remaining five. Another general characteristic is that the shape subspace tend to have lower Minkowski-r:s compared to the r:s for the colour subspace, and this is the case both when the r:s are optimised for the subspaces one by one (*Het sepHTsepSB*) and when they are optimised for both spaces simultaneously (*Het simHTsimSB*). This fact, that the r for the separable shape space, is lower compared to the r for the integral colour space is in line with previous research. However, in this case both r:s are larger than what has been found in previous research, i.e. when the subspaces are not combined. A probable explanation, especially for the large r for the colour space, is the unequal weighting of dimensions (see Correspondence above).

Focusing merely upon rules based on r = 1.0, r = 2.0 and r = 50.0 the results for the average subject data in Table 2, again, captures a general characteristic: in this case, as well as generally for the individual data, the homogenous Euclidean model (*Hom euc*) has a lower error compared to other models.

Errors and r:s were calculated also for the heterogenous rules combining the "odd", or counterintuitive, subspaces height/saturation and tilt/brightness on one hand and height/brightness and tilt/saturation on the other. Table 3, presenting the heterogenous models with the lowest errors from the three subspace divisions for the average data, are representative for the complete results (presented in Appendices ID and IE).

Table 3: The best heterogenous models and associated errors for average data for the three possible subspace divisions - Experiment I.

Subspace division	Model	Err
height/tilt; sat./bri.	**Het simHTsimSB**	**4.066**
height/sat.; tilt/bri.	**Het simHSsimTB**	**6.923**
height/bri.; tilt/sat.	**Het simHBsimTS**	**6.683**

For the average data, the errors for the heterogenous models for the "odd" subspace divisions are considerably larger compared to the error for the original division. For the individual data, the corresponding difference was true for 11 out of 12 cases (c.f. Appendices ID and IV). This difference in errors for the original and odd subspace divisions indicate that the intuitive division into subspaces of shape and colour makes sense.

EXPERIMENT II

A second experiment, with the same underlying space, was conducted in order to investigate if, as hinted above, a larger variation in the colour space would suggest a heterogenous model with lower Minkowski-r:s.

Subjects

A total of 14 students (the majority were undergraduates) at the University of Skövde participated for a reward of two cinema tickets (the value corresponded to 140 SKr, roughly £11 or $17).

Stimuli

As in Experiment I (see above), the stimuli varied in four dimensions, height (h), tilt (t), saturation (s) and brightness (b) of a parallelogram. Each dimension varied in three levels, h: (4, 5 or 6 units of length), t: (40, 50 or 60 degrees), s: (40, 60 or 80% of maximum saturation) and b: (40, 60 or 80% of maximum brightness). The width and the hue of the parallelograms were held constant (4 units of length and 240°, respectively).

The complete domain formed according to these speci-fications consist of the 81 different parallel-ograms presented in Appendix IIA.

The same pairs (w.r.t. the numbers of the stimuli), and order between pairs as in Experiment I were used – with the difference that one pair of the redundant pairs where changed into a new pair (c.f. Appendix IIA).

Procedure

The experiment was conducted in the same way as Experiment I above, except for some details:

- All subjects attending to Experiment I reported the colour test to be simple, and since the colours in Experiment II were more different, the test was replaced by a simple question whether subjects had normal colour vision or not.
- Several subjects attending to Experiment I reported that the ”forced” breaks felt too long, wherefore the minimum length of breaks in Experiment II were shortened to half the time compared to Experiment I.

All subjects reported they had normal colour vision.

Results and analysis

The stimulus pairs selected, and the response from subjects, are jointly presented in Appendix IIA.

Correspondence

Although not perfectly equal, the average component distances for the collapsed data in Experiment II (Table 4 below), are more equal compared to the corresponding distances for Experiment I (Table 1 above). However, for the collapsed data as well as for the individual data (Appendix IIB), it is still the case that the saturation and brightness are less weighted compared to height and tilt.

Table 4: Component distances averaged for each dimension, and coefficient of determination - Experiment II.

Avg_h	**Avg_t**	**Avg_s**	**Avg_b**	**R^2**
4.160	2.907	1.214	2.360	0.762

For the collapsed data, the coefficient of determination was larger for this experiment compared to Experiment I. However, there is still a considerable proportion of the variance of the data that remains unexplained by a general linear model.

Interdimensional additivity

As for Experiment I, the collapsed data in Experiment II to some extent contends violations against interdimensional additivity. In the present case, 3 out of 15 separate tests significantly deviated from the expected value of .5. The corresponding equality tests regarding interdimensional additivity for individuals are presented in Appendix IIB.

In summary, as for Experiment I, there exist violations against the assumptions of correspondence and interdimensional additivity for the data from Experiment II.

Spatial Metric

The same candidate models as evaluated in Experiment I were evaluated, and the errors for the collapsed data are presented in Table 5 below. The results for individual subjects are presented in Appendix IIC.

Contrary to Experiment I, in this case a hetero-genous model combining a rule between the city-block and the Euclidean metrics (though closer to the Euclidean) for the shape space, and a rule roughly corresponding to the Euclidean metric for the colour space (*Het simHTsimSB* and *Het sepHTsepSB*), gave a lower error than the best of the homogenous models (*Hom opt*), which had a Minkowski-r = 1.2, i.e. halfway between the city-block and the Euclidean metrics. This was true irrespectively of whether the r:s were optimised separately or simultaneously (see section Spatial metric under Experiment I above). The optimal heterogenous Minkowski-r:s found in the present case were, as for Experiment I, lower for the shape space compared to the colour space. However, an important difference is that the Minkowski-r:s found in Experiment II were more in line with the levels identified by previous research when two-dimensional stimuli have been used compared to Experiment I. Even though the Minkowski-r:s for Experiment II were closer to the city-block metric for the shape space and the Euclidean metric for the colour space compared to Experiment I, the values were somewhat higher compared to what has been identified for these spaces before. It may be the case that the r-value goes up when the dimensionality increases. This speculation makes some sense considering the fact that we have limitations in terms of how many dimensions we can process simultaneously, and that larger values of r corresponds to focusing more on the dimension where the stimuli-pair at hand differ the most.

Table 5: Models, associated r:s and errors for average data, sorted after errors – Experiment II.

	r	Err
Het simHTsimSB	1.55;2.25	2.146
Het sepHTsepSB	1.55;2.2	2.146
Het euceuc	2;2	2.339
Hom opt	1.2	2.481
Het eucdom	2;50	2.601
Het euccit	2;1	2.894
Het domeuc	50;2	3.838
Het domcit	50;1	3.905
Het citdom	1;50	3.948
Het citeuc	1;2	4.194
Het domdom	50;50	4.313
Hom cit	1	5.907
Hom euc	2	7.805
Hom dom	50	16.644

Another important difference is that the error levels for Experiment II generally are substantially lower compared to Experiment I, making the results from Experiment II more reliable. Worth to mention is also that the common homogenous Euclidean rule (*Hom euc*) gave a substantially worse error than both the best heterogenous rule and the best homogenous rule. However, the still somewhat unequal weightings of the dimensions defining the two subspaces (see Correspondence above) probably causes the pecularity that *Het euccit* produces an error lower than that for *Het citeuc*. The fact that there still are differences in weighting indicate that there are differences in salience between dimensions.

In summary, the increase (compared to Experiment I) in variety for the saturation and brightness dimensions seem to have caused a significant difference in the outcome of the experiment: the unequality of the weightings between dimensions has decreased and the error levels have decreased considerably, thus, the outcome is more reliable. Finally, there are clear indications that a heterogenous rule or model better describes the data compared to a homogenous one.

As for Experiment I, errors and r:s were calculated also for the heterogenous rules combining the "odd" subspaces height/saturation and tilt/brightness on one hand, and height/brightness and tilt/saturation on the other. The heterogenous models with the lowest errors for the average data for each of the three subspace divisions are presented in Table 6. The individual results are presented in Appendices IID and IIE.

Table 6: The best heterogenous models and associated errors for average data for the three possible subspace divisions - Experiment II.

Subspace division	Model	Err
height/tilt; sat./bri.	**Het simHTsimSB**	2.146
height/sat.; tilt/bri.	**Het simHSsimTB**	3.030
height/bri.; tilt/sat.	**Het simHBsimTS**	2.861

As for Experiment I, for the average data in Experiment II, the errors for the "odd" subspaces are larger compared to the error for the original division. For the individual data, the corresponding difference was true for 8 out of 12 cases with a least one r <> 1.0.

EXPERIMENT III

In Experiments I and II, the heterogenous r:s found were larger than what have been found in earlier research. A reasonable question to ask oneself is if the element of non-separability together with the increased dimensionality causes such effects. A third experiment was conducted in order to investigate if the possible factor of integrality (non-separability) could be eliminated as an explanation or not. Contrary to Experiments I and II, for which the underlying dimensions were a mix of of separable and integral dimensions, the underlying dimensions in the present experiment are purely separable.

Subjects

A total of 12 students (the majority were undergraduates) at the University of Skövde participated for a reward of two cinema tickets (the value corresponded to 140 SKr, roughly £11 or $17).

Stimuli

The stimuli varied in four dimensions (see Figure 4 below), height (h), tilt (t), width of a stripe parallell to the horisontal axes (st) and brightness (b) of a parallelogram. These dimensions differ from the ones used in Experiments I and II above in some crucial aspects. One is that they do not form intuitive subspaces. Another is that all possible pairs of dimensions match the description of separable dimensions.

Each dimension varied in three levels, h: (4, 5 or 6 units of length), t: (40, 50 or 60 degrees), st: (1, 2 or 3 units of width) and b: (40, 60 or 80% of maximum brightness). The width, hue and saturation of the parallelograms were held constant (4 units of length, 240° and 60% of maximum saturation, respectively).

The complete domain formed according to these specifications consists of the 81 different parallelograms presented in Appendix IIIA.

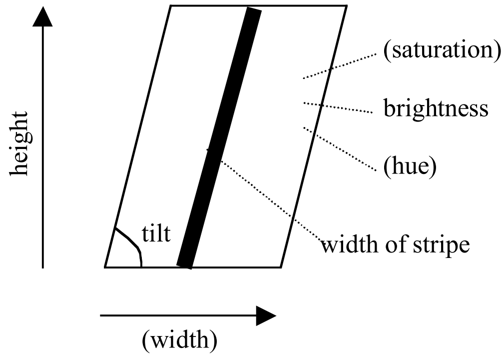

Figure 4: The dimensions "height", "tilt", "width", "width of stripe", "saturation", "brightness" and "hue" of a parallelogram.

The same pairs (w.r.t. the numbers of the stimuli), and order between pairs as in Experiment II were used.

Procedure

The experiment was conducted in the same way as Experiment II above.

Results

The stimulus pairs selected, and the response from subjects are jointly presented in Appendix IIIA.

Correspondence

The average component distances for the collapsed data in Experiment III (Table 7 below, see Appendix IIIB for the corresponding data for individuals), are not perfectly equal, especially the brightness dimension is weighted less compared to the others.

Table 7: Component distances averaged for each dimension, and coefficient of determination - Experiment III.

Avg_h	Avg_t	Avg_st	Avg_b	R^2
2.089	2.381	1.530	0.625	0.541

The coefficient of determination is very low for the collapsed data, hence a linear model does not apply well.

Interdimensional additivity

The equality tests regarding Experiment III are presented in Appendix IIIB. The collapsed data to some extent contends violations against interdimensional additivity since 1 of the 15 tests significantly deviated from the expected value of .5.

In summary, because there exist violations against correspondence and interdimensional additivity, the data from Experiment III are not either (compare Experiments I and II above) ideally described by any Minkowski metric.

Spatial Metric

The same candidate models as evaluated in the previous experiments were evaluated. The resulting errors are presented in Table 8 (for collapsed data) and Appendix IIIC (for individual subjects).

Table 8: Models, associated r:s and errors for average data, sorted after errors – Experiment III.

	r	Err
Het simHTsimSTB	1.1;1	3.483
Hom opt	1	3.523
Hom cit	1	3.523
Het sepHTsepSTB	1.6;1	4.010
Het citeuc	1;2	4.213
Het euccit	2;1	4.434
Het citdom	1;50	4.638
Het euceuc	2;2	5.573
Het domcit	50;1	5.885
Het eucdom	2;50	6.185
Het domeuc	50;2	7.234
Het domdom	50;50	7.935
Hom euc	2	11.333
Hom dom	50	17.219

It is clear that the best rule, of the ones tested for, for describing the collapsed data in Experiment III is close to a city-block rule (*Het simHTsimSTB* (r=1.1;1), *Hom opt* (r=1) and *Hom cit*). It is not, in this special case, possible to view this as supporting either of homogenous or heterogenous models. This is since the city-block metric simply is the sum of the differences for the constituting dimensions, and there is therefore no difference between a homogenous city-block rule and a heterogenous rule where city-block rules are used within all subspaces.

As opposed to experiments I and II, the Minkowski-r values (for the best models) did not increase in magnitude with increased dimensionality.

The heterogenous models with the lowest errors for the average data for each of the three subspace divisions are presented in Table 9, and the corresponding results for the individual data can be found in Appendices IIID and IIIE. As, for the collapsed data, the optimal "heterogenous" rule for the "original" subspace division was close to the city-block metric for both subspaces, this was necessarily the case also for the “odd” subspace divisions.

Table 9: The best heterogenous models and associated errors for average data for the three possible subspace divisions - Experiment III.

Subspace division	Model	Err
height/tilt; str./bri.	**Het simHTsimSTB**	3.483
height/str; tilt/bri.	**Het simHSTsimTB**	3.523
height/bri.; tilt/str.	**Het simHBsimTST**	3.523

EXPERIMENT IV

The stimulus material and the collection of the data used in Experiment IV have previously been described in detail elsewhere (Johannesson, 1996).

Subjects

Ten subjects (most of them undergraduates at the University of Skövde) participated without credit.

Stimuli

The stimuli were designed in order to look like beetles. They were created by Niklas Mellegård, Chalmers University, who did the artistic work, and the author, who stood for the entomological details. The stimuli beared, on purpose, no resemblance to any particular type of existing beetle (a sample is presented in Figure 5).

The stimuli varied in three physical dimensions: the absolute size of the head (H), the length of the abdomen (L) and the width of the abdomen (W) (see Figure 6). The head varied between two parameter values, 0.75 and 1.25. The length and width of the abdomen varied between three parameter values, 0.5, 1.0 and 1.5. The parameter values of the head could be interpreted as (parameter value * 100) % of area compared to the prototype beetle (i.e. the “original” beetle used for rendering the rest). The values for length and width have the same interpretation, except that they are relative to the prototype’s length and width, respectively, rather than to the area.

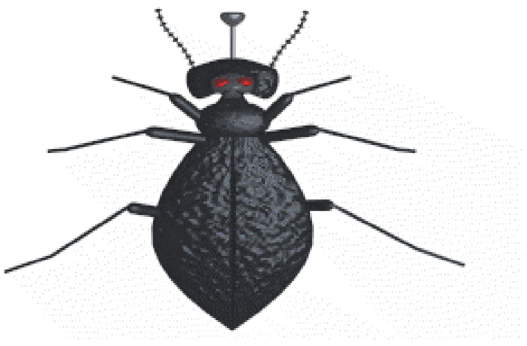

Figure 5: Sample of a beetle-stimuli.

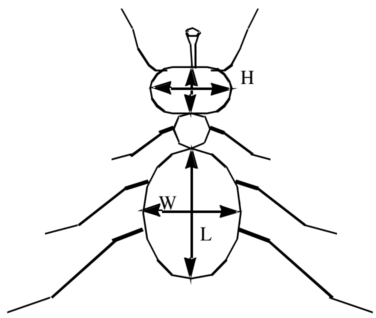

Figure 6: The dimensions H - absolute size of the head, L - the length of the abdomen and W - the width of the abdomen.

The complete domain formed according to these specifications consist of the 18 different beetles presented in Appendix IVA.

All 153 possible (non-ordered) pairs of stimuli were presented randomly (different for different subjects).

Procedure

The experimental sessions were performed individually. The scene differed somewhat between subjects, but the environment was in all cases quiet.

The data were collected during a session with PsyScope (Cohen, MacWhinney, Flatt & Provost, 1993). The session mainly consisted of three different phases: an instruction phase, a presentation phase and a similarity rating phase.

During the instruction phase, subjects first received general information about the purpose with the test and were then informed about the nine-grade scale they should use: “1” corresponded to “large similarity” whereas “9” corresponded to “large dissimilarity”. Subjects were also instructed not to spend too much time on particular pairs - and phenomenological dissimilarities were at interest.[7]

During the presentation phase all beetles were presented pairwise in a randomised order that was the same for all subjects.

During the rating phase subjects rated each of the 153 pairs of beetles with respect to their similarity/dissimilarity by pressing the corresponding figure-button (a nine-grade scale with marked endpoints was shown below each pair).

For further details regarding the procedure, see (Johannesson, 1996).

[7] Note, however, that this instruction may have affected the result.

Results

The stimulus pairs selected, and the response from subjects, are jointly presented in Appendix IVA.

Correspondence

The average component distances for the collapsed data in Experiment IV (Table 10 below, see Appendix IVB for the corresponding data for individuals), are relatively equal, meaning that no single dimension seems to be significantly more salient compared to the others. The coefficient of determination is quite large meaning that a general linear model describes the data quite well.

Table 10: Component distances averaged for each dimension, and coefficient of determination - Experiment IV.

Avg_L	Avg_W	Avg_H	R^2
2.794	2.489	3.347	0.890

Interdimensional additivity

The equality tests regarding Experiment IV are presented in Appendix IVB.

The collapsed data in Experiment IV contain quite severe violations against interdimensional additivity. In the present case, 4 out of 7 separate tests significantly deviated from the expected value of .5.

The corresponding equality tests regarding inter-dimensional additivity for individuals are presented in Appendix IVB.

Spatial Metric

The same candidate models as evaluated in the previous experiments were evaluated. Note, however, that for this three-dimensional stimulus, in the hetero-genous cases, it is not meaningful to talk about a Minkowski-r for the solitary dimension. The resulting errors are presented in Table 11 (for collapsed data) and Appendix IVC (for individual subjects).

Table 11: Models, associated r:s and errors for average data, sorted after errors – Experiment IV.

	R	Err
Het simLW_H	17.2;_	1.005
Het dom_H	50;_	1.008
Hom opt	1.7	1.039
Het sepLW_H	3.6;_	1.111
Hom euc	2	1.141
Het euc_H	2;_	1.606
Hom dom	50	2.954
Het cit_H	1;_	5.645
Hom cit	1	5.645

In this case a heterogenous model combining a rule close to the dominance metric for the length/width-space with the solitary headsize-dimension gave a slightly lower error than the best of the homogenous models (*Hom opt*), which had a Minkowski-r = 1.7, i.e. close to the Euclidean metric. This was only the case when the r for the length/width-space was optimised simultaneously.

As opposed to the previous experiments, the common homogenous Euclidean rule (*Hom euc*) seems to describe the proximity data about as well as the best heterogenous rule.

In summary, a heterogenous model provide the best fit according to the error measure used, but the differences between this and the best homogenous model are small.

The heterogenous models with the lowest errors for the average data for each of the three subspace divisions are presented in Table 12, and the corresponding results for the individual data can be found in Appendices IVD and IVE. Of the possible divisions of subspaces it is apparent that the “original” intuitive division, where length/width are clustered together and headsize is solitary, was clearly the best.

Table 12: The best heterogenous models and associated errors for average data for the three possible subspace divisions - Experiment IV.

Subspace division	Model	Err
length/width; headsize	**Het simLW_H**	1.005
headsize/length; width	**Het simHL_W**	3.004
headsize/width; length	**Het simHW_L**	3.127

EXPERIMENT V

As for Experiment IV, the stimulus material and the collectction of the data used in Experiment V have previously been described in detail elsewhere (Johannesson, 1996).

Subjects

Eleven subjects with different backgrounds partici-pated in the study without any credit.

Stimuli

Computer generated pictures of mollusc shells (see Figure 7 for a sample) developed by and used by Gärdenfors and Holmqvist (1994), were used.

The shells (Figure 8) varied in three physical dimensions: the rate E of whorl expansion, which determines the curvature of the shell, the rate V of vertical translation along the coiling axis and the expansion rate R of the generative curve of the shell.

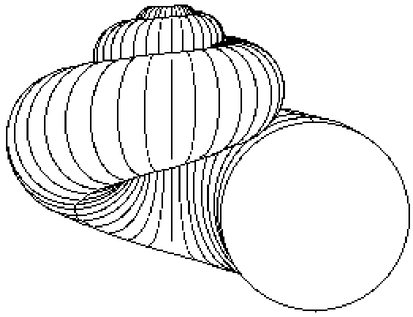

Figure 7: Sample of a shell-stimuli.

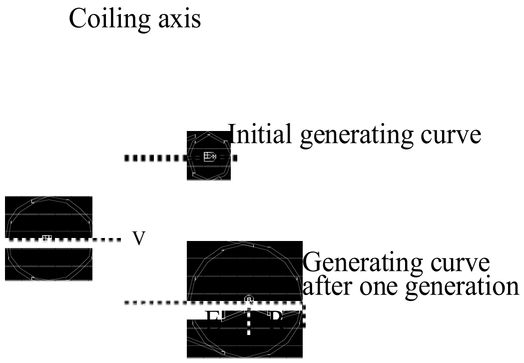

Figure 8: The dimensions V, E and R of a shell.

For parameters V and E, three levels (1.1, 1.25 and 1.4) were used. Parameter R had two levels (1.1 and 1.2).

The complete domain formed according to these specifications consists of the 18 different beetles presented in Appendix VA.

All 153 possible (non-ordered) pairs of stimuli were presented randomly (different for different subjects).

Procedure

The experimental sessions were conducted as for Experiment IV presented above. For further details regarding the procedure, see (Johannesson, 1996).

Results

The stimulus pairs selected, and the response from subjects are jointly presented in Appendix VA.

Correspondence

Table 13, presenting the average component distances found for the shells in Experiment V, indicates clear violations against correspondence: parameter V is obviously more weighted than parameter E, which in turn is more weigthed than parameter R. The coefficient of determination indicates, however, that a general linear model can describe the collapsed data (see Appendix VB for the corresponding data for individuals) relatively well.

Table 13: Component distances averaged for each dimension, and coefficient of determination - Experiment V.

Avg_V	Avg_E	Avg_R	R^2
2.117	1.479	0.861	0.865

Interdimensional additivity

The equality tests regarding Experiment V are presented in Appendix VB. The collapsed data in Experiment V to a large extent contends violations against interdimensional additivity in that 5 out of 7 separate tests significantly deviated from the expected value of .5.

In summary, for the Experiment V-data, there exist severe violations against the assumptions of correspondence and interdimensional additivity.

Spatial Metric

As opposed to the previous experiments, there were in the present case no division into subspaces that appeared to be more intuitive than others. Rather, the subspace to be discussed (V/E and R) was chosen to be the division with the lowest error for the simultaneously optimised heterogenous model (cf. Table 15 below).

The same candidate models as evaluated in Experiment IV were evaluated. The resulting errors are presented in Table 14 (for collapsed data) and Appendix VC (for individual subjects).

Table 14: Models, associated r:s and errors for average data, sorted after errors – Experiment V.

	R	Err
Hom opt	2.5	0.343
Hom euc	2	0.372
Hom dom	50	0.424
Het simVE_R	6.2;_	0.561
Het dom_R	50;_	0.564
Het sepVE_R	3;_	0.591
Het euc_R	2;_	0.720
Het cit_R	1;_	2.305
Hom cit	1	2.305

In the present case it is evident that a homogenous rule is more applicable compared to a heterogenous one. The optimal homogenous rule is between the Euclidean and the dominance metrics (somewhat closer to the Euclidean).

The heterogenous models with the lowest errors for the average data for each of the three subspace divisions are presented in Table 14, and the corresponding results for the individual data can be found in Appendices VD and VE. It is obvious that there are differences with respect to levels of error for the different divisions. However, since a heterogenous model obviously wasn't the best model, it is hard to tell whether this result is a coincidence or not.

Table 15: The best heterogenous models and associated errors for average data for the three possible subspace divisions - Experiment V.

Subspace division	Model	Err
V/E; R	Het simVE_R	0. 561
R/V; E	Het simRV_E	1.586
R/E; V	Het simRE_V	1.465

GENERAL DISCUSSION

The aim of this paper is to argue that one could gain from not just taking the common Euclidean metric for granted or to use it by tradition, but instead taking aspects like the nature of the objects of interest into consideration. The idea that division of features, or dimensions, of objects into separate subspaces – when applicable - possibly could increase descriptive power was investigated.

Experiments I and II both involved pairs of dimensions previously found to be combined best by two different metrics (the city-block and the Euclidean metric, respectively). In Experiment I, the best fitting metric found was a homogenous rule with r = 1.6, which could be described as a trade-off between integrality and separability. However, this result should be interpreted with care: the error levels in Experiment II were considerably lower, thus making the outome more reliable. In Experiment II, the Euclidean rule turned out to badly describe the data. Instead, a heterogenous rule combining the two subspaces formed by the intuitive division, was found to provide the best description. The Minkowski-r:s for the two subspaces found in this experiment rhymes with previous research in that they really possess different metric properties and that the r for saturation/brightness was higher than for height/tilt. However, both r:s found were somewhat larger compared to what has been found previously for the separate two-dimensional subspaces (see e.g. Hyman & Well, 1967; 1968 and Tversky & Gati, 1982, respectively). As opposed to Experiments I and II, the dimensions involved in Experiment III were all expected to be pairwise separable. Also in the four-dimensional case, the best describing metric turned out to be the city-block rule. Because of the nature of the city-block metric, it does not make any difference whether the space is divided into subspaces or not w.r.t. the resulting distances. Experiments IV and V differ from Experiments I - III in that there are no previous research indicating how pairs of the constituent dimensions are combined. In the case of Experiment IV it was more intuitive that the length and width of the abdomen "go together" compared to other possible divisions. The best describing rule was, again, not the Euclidean metric, but rather a heterogenous one where length and width of the abdomen combined with a dominance metric separately from the size of the head. It is possible that another representation of the dimensions involved (e.g. shape), and another rule, would have managed to describe the data even better. That is, however, beyound the scope of this paper. For Experiment V, there were no division into subspaces that was judged to be more intuitive than others because all dimensions - at least according to the author - interacted in a complicated manner. The best heterogenous rule for the subspace division giving the lowest error was, however, almost double compared to the best homogenous rule, which in fact was relatively close to the Euclidean rule.

The general pattern that could be identified from the experiments is that when stimuli are not subject to confusion (i.e. when they are sufficiently different and when the subject can see the contribution of the constituent dimensions without much effort), their phenomenological similarity/dissimilarity can be more accurately described by a heterogenous (i.e. combinatorial) rule taking aspects of the stimuli into consideration, compared to a homogenous Minkowski -metric. Thus, the idea presented received some support. Of interest is also that the experiments where a homogenous rule was found to be the best (i.e. Experiments I and V) tended to violate the assumptions of correspondence and interdimensional additivity more than others.

There are a number of open questions. For example, given that it is adequate to divide different aspects of objects into separate subspaces, a relevant issue is, besides how the dimensions within subspaces are combined, how the subspaces themselves are combined. In this paper, only one of many possible ways of doing this was investigated

Another open question concerns the relatively large Minkowski-r:s found in Experiment II. Since the r:s estimated in Experiment III were not larger compared to what could be expected for pairwise combinations of the constituent dimensions, it is apparent that the increase in magnitude of r:s as found in Experiment II, is not generalisable to all complex stimuli. However, in the developmental literature it is well documented that the separability of dimensions is not fixed but rather changes with experience (see e.g. Smith, 1980), with the direction from integrality to separability. This pattern also apply to short term learning (Johannesson, 2001). A possible reason for the relatively large r:s in Experiments I, II, IV and V and the stable r:s in Experiment III could thus be that stimuli with contents of integrality are harder to "learn" than stimuli composed just by separable dimensions. If so, the r:s could possibly stabilise at a lower magnitude for sufficiently experienced subjects. If not, it could simply be that the specific metric properties associated with integral/separable dimensions only are true in the context of single pairs of dimensions, i.e. depending on if they are combined or not. An example of stimuli, to start with, that could be used in order to explore this issue (and others) further is the multimodal stimuli composed of the pairwise integral dimensions of pitch/loudness and hue/saturation.

Further, the outcome that each of the experiments presented to some extent violated the assumptions of correspondence and interdimensional additivity, is not very surprising in light of the fact that some individuals' data in Dunn's (1983) experiments also contained such violations, even though the stimuli were simpler. Relevant questions are when, to what extent and why such violations occur.

The results presented clearly motivates further research on the idea that information integration could be described as a combination of distances within different subspaces. More research on if, how and when information integration behaviour can be described in terms of combinations of subspaces may shed light on how we interact with the inherently high-dimensional real world. For example, Edelman and Intrator (1997) discuss the necessity of low dimensionality for learning in perceptual tasks – known as 'the curse of dimensionality'. However, even if we always use low-dimensional representations internally, even for cognition, if these representations involve more than two dimensions, cognitive science have interesting problems to solve.

ACKNOWLEDGEMENTS

This project have been financed by the Department of Computer Science at University of Skövde and by a grant to the Center of Learning Systems at University of Skövde from the Foundation for Knowledge and Competence Development (1507/97), Sweden. I wish to thank Anders Eklund and Nils Jansson, both University of Skövde, for practical help, Professor Lars Niklasson, University of Skövde, and Professor Peter Gärdenfors, Lund University Cognitive Science, for support and constructive comments.

REFERENCES

Attneave, F. (1950). Dimensions of similarity. American Journal of Psychology, 63, 516-556.

Dunn, J. C. (1983). Spatial metrics of integral and separable dimensions. Journal of Experimental Psychology: Human Perception and Performance, 9 (2), 242-257.

Edelman, S. and Intrator, N. (1997). Learning as formation of low-dimensional representation spaces. In Shafto, M. G. and Langley, P., editors, Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society, Erlbaum, Mahwah, NJ, 199-204.

Garner, W. R. (1974). The processing of information and structure, New York, Wiley.

Garner, W. R. (1977). The effect of absolute size on the separability of the dimensions of size and brightness. Bulletin of the Psychonomic Society, 9 (5), 380-382.

Gottwald, R. L. and Garner, W. R. (1975). Filtering and condensation tasks with integral and separable dimensions. Perception & Psychophysics, 18 (1), 26-28.

Gärdenfors, P. (1992). A geometric model of concept formation. In S. Ohsuga et al. (Eds.) Information Modelling and Knowledge Bases III, 1-16, IOS Press, Amsterdam.

Gärdenfors, P. and Holmqvist, K. (1994). Concept formation in dimensional spaces. Lund University Cognitive Studies, 26, Lund University Cognitive Science, Lund University, Sweden.

Gärdenfors, P. (2000). Conceptual Spaces. Bradford Books, MIT Press.

Hyman, R. and Well, A. (1967). Judgments of similarity and spatial models. Perception & Psychophysics, 2 (6).

Hyman, R and Well, A. (1968). Perceptual separability and spatial models. Perception & Psychophysics, 3, 161-165.

Johannesson, M. (1996). Obtaining Psychologically Motivated Spaces with MDS. Lund University Cognitive Studies – LUCS 45. Lund, Sweden.

Johannesson, M. (2001). Toward Separability During Learning. Submitted for publication.

Kemler Nelson, D. G. (1993). Processing integral dimensions: The whole view. Journal of Experimental Psychology: Human Perception and Performance, 19 (5), 1105 - 1113.

Keppel, G. (1991). Design and Analysis: A Researcher's Handbook., 3rd ed., Engelwood Cliff, Prentice Hall.

Palmer, S. E. (1978). Fundamental aspects of cognitive representation, In Cognition and Categorization, Rosch, E. and Lloyd, B. B., (eds.), Lawrence Erlbaum Associates, Hillsdale, New Jersey, pp. 259 - 303.

Ronacher, B. (1998). How do bees learn and recognize visual patterns? Biol. Cybern., 79, 477-485.

Ruskin, E. M. and Kaye, D. B. (1990). Developmental differences in visual processing: Strategy versus structure. Journal of experimental child psychology, 50, 1-24.

Tversky, A. and Gati, I. (1982). Similarity, separability, and the triangle inequality. Psychological Review, 89 (2), 123-154.