Invariance and Meaningfulness in Phenotype spaces

Simon M. Huttegger · Philipp Mitteroecker

Abstract Mathematical spaces are widely used in the sciences for representing quantitative and qualitative relations between objects or individuals. Phenotype spaces—spaces whose elements represent phenotypes—are frequently applied in morphometrics, evolutionary quantitative genetics, and systematics. In many applications, several quantitative measurements are taken as the orthogonal axes of a Euclidean vector space. We show that incommensurable units, geometric dependencies between measurements, and arbitrary spacing of measurements do not warrant a Euclidean geometry for phenotype spaces. Instead, we propose that most phenotype spaces have an affine structure. This has profound consequences for the meaningfulness of biological statements derived from a phenotype space, as they should be invariant relative to the transformations determining the structure of the phenotype space. Meaningful geometric relations in an affine space are incidence, linearity, parallel lines, distances along parallel lines, intermediacy, and ratios of volumes. Biological hypotheses should be phrased and tested in terms of these fundamental geometries, whereas the interpretation of angles and of phenotypic distances in different directions should be avoided. We present meaningful notions of phenotypic variance and other statistics for an affine phenotype space. Furthermore, we connect our findings to standard examples of morphospaces such as Raup's space of coiled shells and Kendall's shape space.

S. M. Huttegger (✉)
Department of Logic and Philosophy of Science, University of California, 3151 Social Science Plaza A, Irvine, CA 92697, USA
e-mail: shuttegg@uci.edu

P. Mitteroecker
Department of Theoretical Biology, University of Vienna, Vienna, Austria
e-mail: philipp.mitteroecker@univie.ac.at

Introduction

Mathematical spaces are widely used in the sciences for representing quantitative and qualitative relations between objects or individuals. Biological applications range from representations of stretches of DNA, RNA, or proteins as elements of a sequence space (e.g., Maynard Smith 1970; Stadler et al. 2001, 2002), to phenotypes as equivalence classes of genotypes (e.g., Fontana and Schuster 1998a, b), to different approaches of capturing similarity and closeness of morphologies (see Sneath and Sokal 1973; Bookstein 1991; McGhee 1999; Mitteroecker and Gunz 2009, and references therein). A *phenotype space* is a space whose elements represent phenotypes such as RNA folding patterns, morphological, physiological, or behavioral traits. Specifying the relations between phenotypes in terms of a formal space can lead to effective analyses and visualizations of complex connections and variation patterns that would not be easily accessible by tabulations of pairwise relations. For example, phenotype spaces have proven relevant to investigate the evolutionary accessibility of phenotypes or to compare developmental trajectories and growth patterns (Fontana and Schuster 1998a; Stadler et al. 2001; Mitteroecker et al. 2004). Fitness landscapes or adaptive landscapes represent the distribution of fitness or adaptiveness across a phenotype space (Wright 1932; Arnold et al., 2001). For all these types of spaces, the range

of useful scientific inferences is determined by the geometrical or topological properties of the formal space that appropriately represents the empirical data.

To put it very generally, a mathematical space consists of some set of elements (often called "points") together with particular relations describing the structure of the space. Since structures can be more or less specific, spaces differ in their degree of generality. *Euclidean space* consists of points in a vector space together with the notions of angle and distance. *Affine space*, for example, is one generalization of Euclidean space where distances in different directions are incommensurable. (We shall describe Euclidean and affine geometry in more detail below.) Issues of "meaningfulness" in biology and other scientific disciplines arise very often because, as expressed in Suppes et al. (1989, p. 40), "whenever a Cartesian coordinate system is used, one tends to assume that all Euclidean relations, and only those, are defined and usable." It is, however, very often the case that a geometric structure that adequately represents a domain of interest (like a phenotype space) is not Euclidean but, in fact, less specific. In these cases, Euclidean geometry is too "rich". It allows one to draw conclusions that do not correspond to features of the empirical domain of interest, as the empirical structure permits a wider class of transformations than just Euclidean transformations (see below). Conclusions drawn within the framework of Euclidean geometry thus are not guaranteed to be scientifically meaningful.

We shall focus in this paper on necessary conditions for scientifically meaningful conclusions that are based on the structure of phenotype spaces. Our examples are morphological spaces or morphospaces, i.e., spaces based on morphological characteristics. Our conclusions, however, apply to other phenotype spaces as well. We shall argue that several kinds of phenotype spaces are, in fact, affine spaces and not Euclidean. Thus, many of the concepts familiar from Euclidean geometry, such as distance and angle, will not be meaningful in most phenotype spaces. We will address notions of similarity and disparity of phenotypes as well as different geometries of phenotypic trajectories, and discuss meaningful quantifications and statistics of these properties.

Vector Spaces, Affine Spaces, and Euclidean Spaces

This section briefly reviews some standard mathematical structures, which can be used to represent relations between phenotypes. We limit our discussion to spaces with a vector space structure and point to implicit assumptions about the underlying empirical structures they are supposed to represent. We refer the interested reader to Suppes et al. (1989) for more details. This section may be skipped by readers with some background in geometry.

Vector Spaces

Vector spaces are widely used to represent very different kinds of empirical structures. A vector space $(V, +, \cdot)$ is an algebraic structure defined over a field F. Elements of V can be added by performing the addition operation $+$. They can also be multiplied with scalars, i.e., elements of the field F. Essentially, the operations of addition and scalar multiplication in general vector spaces are similar to the familiar case of a real vector space, where F is the field of real numbers and where we have component-wise addition

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n)$$

and multiplication with real numbers

$$r(x_1, \ldots, x_n) = (rx_1, \ldots, rx_n).$$

Real vector spaces are the most common type of vector spaces to represent empirical biological structures. However, the field F may also be finite or countably infinite (rationals), and it may be ordered (rationals, reals) or unordered (complex numbers, various finite fields).[1]

Affine Spaces

Affine spaces add more structure to vector spaces by identifying points, lines, planes, and so on. These objects are called *linear varieties*. A linear variety in an n-dimensional vector space is a set of the form

$$\left\{ \mathbf{y} + \sum_{i=1}^{k} \mu_i \mathbf{x}_i \,\middle|\, \mu_i \in F, \quad i = 1, \ldots, k \right\},$$

where $0 \leq k \leq n$ and $\mathbf{y}, \mathbf{x}_1, \ldots, \mathbf{x}_k \in V$ are fixed. If $k = 0$, the linear variety consists of a point $\mathbf{y}$. If $k = 1$ or $k = 2$, then the linear variety is a line or a plane through $\mathbf{y}$, respectively. By fixing k and varying $\mathbf{y}$ and $\mathbf{x}_1, \ldots, \mathbf{x}_k$ we get all linear varieties $\mathcal{L}_k$ of dimension k. For instance, we get all points, lines, or planes. Furthermore, an *incidence relation I* is introduced for linear varieties $L_k \in \mathcal{L}_k$. If $L \in \mathcal{L}_j$ and $L' \in \mathcal{L}_k$ and if $j < k$, then $I(L, L')$ if L' contains L, i.e., if $L \subset L'$. In geometric terms, this means that L lies on L'. Think of a point lying on a line or on a plane. An n-dimensional vector space $(V, +, \cdot)$ together with all linear varieties $\mathcal{L}_0, \ldots, \mathcal{L}_n$ and the incidence relation I constitutes an *affine geometry*.

An *affine transformation* ϕ is characterized by mapping k-dimensional linear varieties to k-dimensional linear varieties, and by preserving incidence relations:

---

[1] When an underlying empirical structure is unordered, the use of an ordered field would contain more structure than its empirical counterpart and may lead to unwarranted conclusions drawn from the mathematical representation.
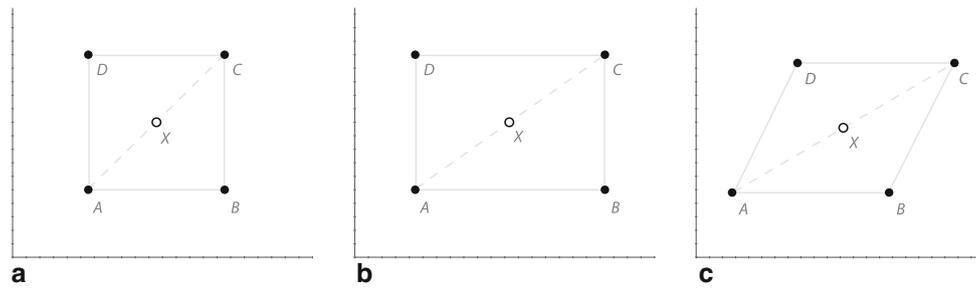
Fig. 1 a Four points (*A–D*), constituting a square in $\mathbb{R}^2$, along with their centroid *X*. In **b** and **c** this point configuration is modified by two different affine transformations. In **b** the transformation is a scaling of the vertical axis and in **c** a shear along the vertical axis. Notice that distances, relative distances, and angles have been changed by these transformations. But in all three figures the lines *AB* and *DC* are parallel and the ratio of the lengths of these lines is constant. Furthermore, the points *AXC* are collinear before and after the affine transformations, and *X* remains the exact midpoint of *A* and *C*. The areas of the triangles *ABC* and *ACD* are affected by the transformation, but the ratio of these areas is invariant

$$\phi(L) \in \mathcal{L}_k \Longleftrightarrow L \in \mathcal{L}_k$$
$$I(\phi(L), \phi(L')) \Longleftrightarrow I(L, L')$$

This means, for instance, that points are mapped to points, lines to lines, and planes to planes. Moreover, points that are on the same line before the transformation will also be on the same line after the transformation.

The definition of affine transformations is equivalent to saying that an affine transformation is an automorphism of *V* together with a translation, i.e.,

$$\phi(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{y},$$

where **A** is a non-singular $n \times n$ matrix and **y** is an arbitrary element of *V*. Geometrically, affine transformations comprise translation, rotation, reflection, scaling, and shearing (Fig. 1).

For an ordered field *F* (such as the reals) the concept of *betweenness* on a line or of a *midpoint* can be defined in affine spaces. More generally, affine transformations leave barycentric combinations invariant; these are given by $\sum_{j=0}^{n} \alpha_j \mathbf{x}_j$, where the weights α_j sum up to 1. E.g., if $n = 2$ and $\alpha_1 = \alpha_2 = 1/2$, then the resulting combination is the midpoint between $\mathbf{x}_1$ and $\mathbf{x}_2$; weights of $1/n$ for *n* points gives the centroid or the center of gravity of these points. Consequently, the concept of a mean is affine invariant (i.e., unchanged by affine transformations).

Euclidean Spaces

In *Euclidean spaces* the concept of *perpendicularity* is added to the affine structure of a vector space over $\mathbb{R}$. In an *n*-dimensional real Euclidean vector space, perpendicularity is commonly defined by

$$\sum_{i=1}^{n} x_i y_i = 0, \tag{1}$$

where $\mathbf{x} = (x_1, \ldots, x_n), \mathbf{y} = (y_1, \ldots, y_n) \in \mathbb{R}^n$. If (1) holds, then **x** and **y** are said to be perpendicular. The left-hand side of (1) is the standard *inner product* of two vectors in $\mathbb{R}^n$. The concept of an inner product can be generalized to vector spaces $(V, +, \cdot)$ over the ordered field $\mathbb{R}$. An inner product $\langle \cdot, \cdot \rangle$ is a mapping from $V \times V$ to $\mathbb{R}$ which is *symmetric* (the order of arguments does not matter), *positive definite* ($\langle \mathbf{x}, \mathbf{x} \rangle > 0$ except for $\mathbf{x} = \mathbf{0}$) and *bilinear* (linear in both arguments). Hence, perpendicularity can be defined for general real vector spaces by defining that **x** is perpendicular to **y** if, and only if, $\langle \mathbf{x}, \mathbf{y} \rangle = 0$. A Euclidean geometry can then be defined as consisting of a vector space over the field $\mathbb{R}$ together with an affine geometry and a perpendicularity relation based on an inner product. An inner product also defines a distance measure *d* by setting

$$d(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle^{\frac{1}{2}}.$$

Congruence is a central notion for Euclidean spaces which is defined in terms of *d*. Two points **x**, **y** are said to be congruent to two other points $\mathbf{x}'$, $\mathbf{y}'$ if, and only if, $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{x}', \mathbf{y}')$. The distance *d* is also a measure of similarity. In particular, if $\mathbf{x} \neq \mathbf{y}$ then $d(\mathbf{x}, \mathbf{y}) > 0$. Moreover *d* is symmetric and satisfies the triangle inequality. Thus, *d* is a *metric*.[2]

Euclidean geometries are preserved under a certain group of transformations, called *similarities*. (Traditional treatments often use invariance relative to *motions*, i.e., similarities which leave distances invariant.) A similarity ψ is an affine transformation which preserves perpendicularity and congruence:

$$\langle \psi(\mathbf{x}) - \psi(\mathbf{0}), \psi(\mathbf{y}) - \psi(\mathbf{0}) \rangle = 0 \Longleftrightarrow \langle \mathbf{x}, \mathbf{y} \rangle = 0$$
$$d(\psi(\mathbf{x}), \psi(\mathbf{y})) = d(\psi(\mathbf{x}'), \psi(\mathbf{y}')) \Longleftrightarrow d(\mathbf{x}, \mathbf{y}) = d(\mathbf{x}', \mathbf{y}')$$

In perhaps more familiar terms, a similarity is an orthogonal transformation (a rotation) together with a translation

---

[2] A vector space together with a distance measure is a special case of a metric space. A metric space does in general not have to be a vector space.
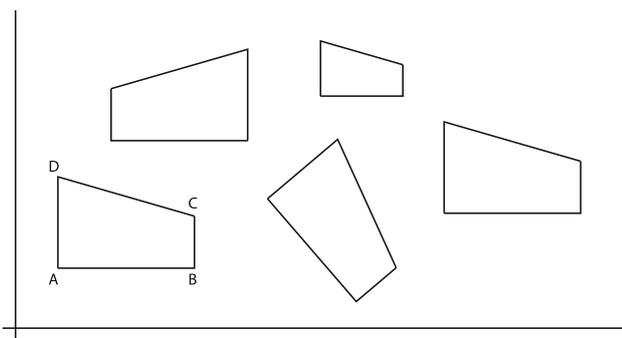
Fig. 2 In Euclidean space, all geometric properties invariant to similarity transformations (translation, uniform scaling, rotation, reflection) are meaningful. The figure shows four two-dimensional points (*A–D*) and several similarity transformations of them. All the affine invariant relations (linearity, parallelity, intermediacy) as well as angles and relative distances are preserved by these transformations

and multiplication by a scalar (uniform scaling; Fig. 2). This invariance also means that we can define Euclidean geometry in terms of the similarity group, like we can define affine geometry in terms of the group of affine transformations, reflecting the dual nature of geometries. They can either be thought of as given by a particular relational structure, from which other relations can be defined. Or they can be conceived as being defined by some group of transformations. The relational structure of the geometry is implicitly given by requiring that it is invariant under the group of transformations (Klein 1872; Narens 2002).

Finally, we would like to make a short remark on the metrizability of affine spaces. As was pointed out above, neither distances nor orderings of distances are invariant to affine transformations. However, distance along a line is an appropriate concept for an affine geometry since orderings of distances on a line are preserved under affine transformations. Similarly, distances of segments of parallel lines can be compared. This extends to certain metrics based on norms (Minkowski geometries, cf. Busemann 1955). It should be pointed out that such metrics do not exhibit strong invariance properties that go beyond the partial one for lines.

Meaningfulness

The group of transformations associated with a geometry provides us with a powerful criterion to evaluate statements about the represented empirical structure. *Meaningful statements* should be invariant under the relevant class of transformations (see Narens 2002, for a comprehensive account of meaningfulness). Performing a transformation of this class produces a new space that is the same as the old one in all relevant respects. A conclusion which holds in one space but fails to hold after transforming the space into an equally suitable one does not reflect significant properties of the represented empirical structure. Thus, all meaningful conclusions that we draw about the empirical structure from its representing mathematical structure must be invariant across all equivalent versions of the mathematical structure.

A meaningful relation in Euclidean space is one that is preserved under similarities (Fig. 2). Distances are meaningful in this sense, since *the order of distances* (not the distances themselves) are invariant to similarity transformations (this follows immediately from the invariance of congruence). Another meaningful relation in Euclidean space are angles (an angle is a function of the ratio of two distances and hence invariant to similarities). However, distances and angles are not meaningful in affine spaces. Similarities are a subgroup of the group of affine transformations. There exist affine transformations which do not preserve Euclidean relations such as angle and distance (Fig. 1). Yet, if a relation is invariant to all affine transformations, then it will also be invariant to similarities. Thus, concepts and relations characteristic of affine spaces are also meaningful in Euclidean spaces. One such geometric concept is given by *parallels*. Parallelism is meaningful both in affine and Euclidean geometry since it is preserved under all affine transformations. An important concept which is meaningful in Euclidean but not affine spaces is *volume*. (Orders of) volumes are preserved under similarities but not under affine transformations, only ratios of volumes are affine invariant (see below).

For phenotype spaces, this translates into an analogous condition for a concept or a conclusion to be meaningful. Formal concepts used to infer biological conclusions, such as the distance between phenotypes or the angle between evolutionary trajectories, must be invariant relative to a suitable class of transformations. Common transformations in biometrics are changes of the origin of a coordinate system used for measurement, changes of the scales or units of variables, log transformations, or the replacement of measures by equivalent measures. The substantive conclusions drawn from the structure of phenotype spaces should be unaffected by these transformations which represent arbitrary choices in the course of research. For example, a meaningful conclusion should not depend on whether we express a measure in millimeters instead of centimeters, or a temperature in degree Celsius rather than in degree Fahrenheit (in statistics this is usually referred to as scale invariance), or whether a structure is represented by two or by three equivalent measurements. While this is not a sufficient criterion for scientific importance, it is a necessary condition for a meaningful biological interpretation of structures in phenotype space. In the following

sections we investigate how this criterion of meaningfulness can be applied to scientific statements derived from multidimensional phenotype spaces.

The Affine Geometry of Phenotype Spaces

It is well established that many physical properties such as length, volume, weight, and temperature can be represented by real numbers together with meaningful operations of addition and multiplication ("extensive measurement"), whereas this is not guaranteed to be the case for other phenotypic characteristics like for behavioral traits (e.g., Krantz et al. 1970). In this paper we limit the discussion to phenotypic properties that can be represented by real numbers and for which meaningful operations of addition and multiplication exist. This is the case for most morphological and physiological measurements. However, numerical representations of biological characteristics typically have empirical upper and lower limits, outside of which the numbers do not correspond to biological structures. E.g., the diameter of an organ cannot be smaller than the diameter of a single cell, or an organ cannot be larger than the whole organism (these considerations can have effects on scales; see Frank and Smith 2010).

To illustrate the properties of multidimensional phenotype spaces derived from a set of such measurements, we examine morphospaces typically used in morphometrics and evolutionary quantitative genetics. One prime example is the work by Raup and Michelson (1965) and Raup

(1966), who developed a simple geometric model of coiled shells based on three parameters: Distance between the cross-section of the shell and the coiling axis, the rate of translation of the cross-section along the axis per revolution, and the rate of increase in the size of the generated shell cross-section per revolution (Fig. 3). Raup used these three parameters to construct a three-dimensional real vector space representing the possible geometries of shells. He compared the regions in this space that were occupied by different taxa to those which appeared to be uninhabited. The size of the region in a morphospace occupied by a taxon is often called its "disparity" (e.g., Foote 1997, see also below). Morphospaces like Raup's have been influential because they allow one to raise the fundamental question of why certain areas in a morphospace seem to be empty. Two common explanations are that either the morphologies in the empty regions are selectively disadvantageous, or that internal constraints on the morphologies make these regions inaccessible (but see also Pie and Weitz 2005). For more examples of these types of morphospaces see McGhee (1999) and references therein.

Similar morphospaces have been put forward in the classical morphometric literature and in numerical taxonomy (e.g. Rohlf and Sokal 1965; Blackith and Reyment 1971; Sneath and Sokal 1973; Bookstein et al. 1985; Marcus 1990). Developments in multivariate statistics, such as principal component analysis, have been applied to morphological measures of all kinds, like distances and ratios of distances, angles, volumes, or counts. The measured values of these variables typically serve as Cartesian
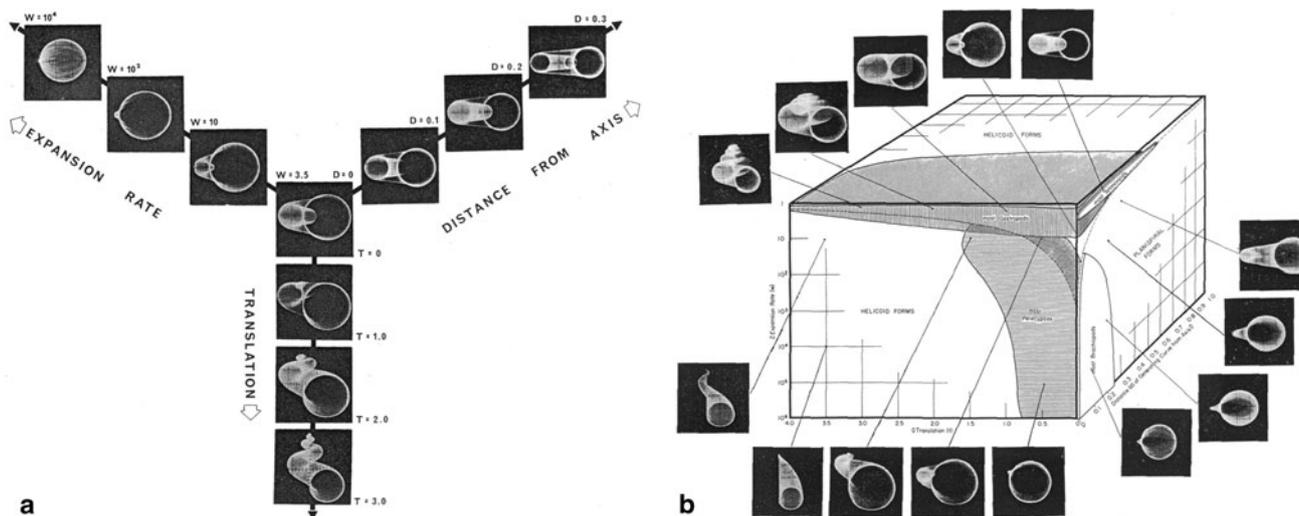

Fig. 3 Raup's (1966) space of coiled shells. To define orthogonal axes of his morphospace, Raup used three out of several parameters in a geometric model of shell geometry. These parameters are the distance between the cross section and the coiling axis, the rate of translation of the cross section along the axis per revolution, and the rate of increase in the size of the generated shell cross section per revolution (**a**). Raup identified regions in this space corresponding to the morphology of existing taxa, but most parts of the space remain empty. These geometrically possible forms have not been realized by Nature (**b**)

coordinates of the specimens in the underlying vector space. Raup's variables, by contrast, were the parameters of his geometric model, not any empirical measurements, and the location of forms in his space was based on visual comparison of actual and computed forms. Likewise, a vector space of measured or unmeasured traits is used in quantitative genetics to study evolutionary changes of phenotypes and the distribution of fitness over a phenotype space (e.g., Lande 1979; Lynch and Walsh 1998; Arnold et al. 2001).

A set of such measurements or parameters, represented by vectors of real numbers together with the operations of addition and scalar multiplication, induce a vector space (with certain empirical limits). An affine space additionally consists of linear structures such as lines or planes. In order to identify linear structures, a joint linear increase or decrease of the measurements must correspond to a feasible phenotypic change. This is not necessarily the case. For example, one side of a rectangle and the diagonal of this rectangle cannot both increase linearly. Likewise, Procrustes shape coordinates or the set of all possible Euclidean distances between a configuration of landmarks constitute a curved manifold rather than a linear space (Kendall 1984; Bookstein 1991; Dryden and Mardia 1998; Slice 2001). However, such manifolds can be unfolded into a linear space (consider, e.g., the Mercator map of the surface of the Earth) or approximated by a linear tangent space (e.g., a plane tangent to the surface at some point). For example, geometric morphometric analyses are usually carried out in the linear tangent space to the curved shape space; this has been shown to be a good approximation even for large biological variation (Rohlf 1999; Marcus et al. 2000).

In many practical applications, the vector space of the measured variables thus has—at least locally—an affine structure, but the variables do not necessarily posses commensurate units. For example, different measures such as distances, angles, or volumes share no common scales but are often used to construct a phenotype space. Two of Raup's three parameters are of the same unit, but they serve fundamentally different roles in his geometric model and there is no "natural" relation among their scales. Furthermore, some variables may be geometrically dependent so that it would be misleading—or at least arbitrary—to take them as orthogonal axes of a morphospace. For example, Raup's three parameters are not independent; a change in whorl expansion rate automatically leads to changes in the other two variables (Schindel 1990). Consider also two measured distances sharing the same starting point, or the angles and distances of a triangle. Similarly, two spatially closely adjacent measurements of an organisms cannot be considered as independent

(Mitteroecker and Bookstein 2007; Mitteroecker 2009). In some cases, like for landmark coordinates, variables may also lack a meaningful origin.

For such measurements there is no unique way of scaling the variables and of arranging the axes to construct a multidimensional space. Thus, if we choose a specific coordinate system we are free to scale each basis vector independently of the other basis vectors. Moreover, the basis vectors do not need to be orthogonal and we may choose an arbitrary origin. (Note that when the basis vectors are not orthogonal, scaling of one vector would affect other vectors too.) These operations correspond to the affine transformations of stretching or shrinking the phenotype space along its axes, to shearing the space, and to translation. Geometric relations characteristic for Euclidean geometry, such as distance and angle, are affected by scaling and shearing the vector space. They are not meaningful in an affine phenotype space.

We thus propose that classical morphospaces exhibit an affine geometry and are not Euclidean. The same applies more generally to spaces of more than one quantitative character as long as there is no substantial relationship between the characters that would justify more geometrical structure, in particular, perpendicularity and congruence. The classical morphospaces and similar phenotype spaces therefore pose certain restrictions on which concepts and relations can be used to describe phenotypes in a meaningful way.

In the next two sections we provide two different formal versions of this argument, the first one starting with a set of measured variables, and the second one starting with a set of phenotypic distances between specimens.

Affine Transformation of Measurements

Consider the vector $\mathbf{x}_i$, consisting of p phenotypic measurements of individual i, where i goes from 1 to n. These vectors can be represented as n points in a p-dimensional phenotype space (a so-called Q-space). In the above introduction to mathematical spaces we have seen that one characterization of an affine space is in terms of affine transformations

$$\phi(\mathbf{x}_i) = \mathbf{A}\mathbf{x}_i + \mathbf{b}, \tag{2}$$

where $\mathbf{A}$ is any regular square matrix.

In the absence of any natural relation between the scales of the measurements we are free to change the scales independently. A change of scale for one variable by the factor k corresponds to a multiplication of the vectors by a diagonal matrix with the value k in the corresponding diagonal element and 1s in all other diagonal elements. Take three variables as an example. Multiplying $\mathbf{x}_i$ with the matrix

$$\begin{pmatrix} k & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{3}$$

multiplies the first measurement by k and leaves all other measurements unchanged (this corresponds to Fig. 1b).

We further argued that the geometric dependence between variables may be a result of the measurement scheme (e.g., interlandmark distances with one landmark in common) or of the model used to generate variables (e.g., the parameters in Raup's model). When the angles between the axes of a phenotype space do not explicitly reflect these relationships, we are also free to separately modify the orientation of the axes. This corresponds to a multiplication by a matrix that differs from the identity matrix by one off-diagonal element. E.g., when multiplying $\mathbf{x}_i$ with the matrix

$$\begin{pmatrix} 1 & k & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{4}$$

the first new variable is a linear combination of the first and the second measurements, whereas all other variables stay the same (Fig. 1c). We are also free to exchange any two axes of the phenotype space, which corresponds to a multiplication by a matrix that differs from the identity matrix by exchanging two rows (or two columns). Multiplying $\mathbf{x}_i$ with

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{5}$$

for instance, exchanges the first two measurements.

These three types of matrices are called *elementary matrices*. It is well known that any nonsingular matrix $\mathbf{A}$ can be decomposed into elementary matrices $\mathbf{A} = \mathbf{T}_q\mathbf{T}_{q-1}\ldots\mathbf{T}_2\mathbf{T}_1$. Each elementary matrix $\mathbf{T}_j$ can be obtained from the identity matrix $\mathbf{I}$ in one step. This step either consists of (i) multiplying a row of $\mathbf{I}$ with a scalar k, (ii) replacing a row of $\mathbf{I}$ by the sum of that row and a multiple of another row, or (iii) exchanging two rows of $\mathbf{I}$.

Geometrically, (i) corresponds to scaling one of the coordinate axes in phenotype space, (ii) represents a shear of two axes, and (iii) corresponds to exchanging two coordinate axes. This means that we can expand the affine transformation in (2) into transformations involving only these three kinds of elementary steps together with a translation. Conversely, performing transformations of type (i), (ii) or (iii) consecutively will always result in a linear (i.e., affine) transformation. (The result

will be a nonsingular matrix since we have excluded projections.)

There is an important relationship between (i) and (ii). Whenever two variables are geometrically dependent, scaling of one variable will also affect the scale of the other variable. In this case, the corresponding axes or basis vectors are not orthogonal and scaling of one axes (i) is performed together with a shear (ii). In fact, the possibility of independent scaling of variables leads to a criterion for orthogonality of the corresponding axes in phenotype space.

If we suppose that our phenotype space has a vector space structure, it will be affine instead of Euclidean if all possible $\mathbf{T}_j$ can be applied to the vector space in arbitrary order without violating any of the information we have about the relationships between the measured specimens (points in the phenotype space). This implies that (a) there is no natural relation between the scale of the variables, allowing for arbitrary rescaling, (b) the geometric dependencies between the variables are unknown, allowing for arbitrary shearing, (c) there is no inherent order in the variables, i.e., the associated coordinate axes can be permuted, and (d) there is no distinguished origin, which allows arbitrary translation of the phenotype space. Whenever these conditions are met, the phenotype space will be a vector space with an affine geometry.

It is instructive to compare these conditions to admissible transformations in Euclidean spaces, i.e., to similarity transformations. A similarity is obtained from an orthogonality (rotation) by a translation and a multiplication by a scalar. Geometrically, an orthogonality rotates a vector space in such a way that a coordinate system with perpendicular axes is transformed into another coordinate system with perpendicular axes (angles between coordinate axes are preserved). This essentially means that the geometric dependencies (angles) between the variables are known and are expressed in the same way in all admissible representations of the phenotype space. Multiplication by a scalar (uniform scaling) corresponds to a scaling of all axes by the same factor. This implies that there is a natural relation between the scales of all variables which is preserved in all admissible representations of the phenotype space. In an affine space, by contrast, the axes can be scaled separately. Both in Euclidean spaces and in affine spaces the coordinate system can be translated and the axes can be permuted.

Weighted Phenotypic Distances

Instead of referring to the measured variables as Cartesian coordinates, a phenotype space can also be constructed by

ordinating a set of phenotypic distances between individuals (using methods such as principal coordinate analysis or multidimensional scaling; Sneath and Sokal 1973; Mardia et al. 1979). This approach has been central to numerical taxonomy, which is aimed at classifying organisms by means of phenotypic similarity (Sokal 1961; Rohlf and Sokal 1965). Cain and Harrison (1960) introduced the term to describe an approach in which forms are spatially arranged in some phenetic space or diagram "by overall similarity, based on all available characters without any weighting" (p. 3). Likewise, the concept of an overall phenotypic distance plays an important role in many morphometric applications. Equal weighting of every character in an overall similarity measure or in a distance function is fundamental in these approaches.

The most frequently used index of overall (dis)similarity between two phenotypes is the Euclidean distance between the two vectors of measurements, which is the 2-norm of the difference vector $\mathbf{x} - \mathbf{y}$:

$$\|\mathbf{x} - \mathbf{y}\|_2 = \left(\sum_{i=1}^{p} (x_i - y_i)^2 \right)^{1/2}. \tag{6}$$

It has also been referred to as taxonomic distance by Sokal and Sneath (1963) and Rohlf and Sokal (1965). In geometric morphometrics, Procrustes distance is usually approximated by the Euclidean distance between two sets of Procrustes shape coordinates (Dryden and Mardia 1998; Rohlf 1999). Euclidean distance is a member of a family of metrics, the Minkowski metrics

$$\|\mathbf{x} - \mathbf{y}\|_k = \left(\sum_{i=1}^{p} |x_i - y_i|^k \right)^{1/k}, \tag{7}$$

but only Euclidean distance ($k = 2$) is invariant to arbitrary rotations of the coordinate system. When $k = 1$ the overall distance is simply given by the sum of the absolute univariate distances, which is referred to as Manhattan distance, city block distance, or mean character difference when divided by the number of variables (Cain and Harrison 1960). It is particularly used for categorical variables but has also been applied to continuous variables.

The central dogma of equal weighting of characters is reflected by the unweighted variables in the distance functions (6) and (7). Sneath and Sokal (1973) defended equal weighting because no principled a priori weighting of characters can be supported. However, equal weighting is of course one particular form of weighting and is, ultimately, an arbitrary decision. A measure of phenotypic distance may thus be supplied with any non-zero weightings a_i for each variable:

$$\begin{aligned} \|\mathbf{x} - \mathbf{y}\|_k' &= \left(\sum_{i=1}^{p} |a_i(x_i - y_i)|^k \right)^{1/k} \\ &= \left(\sum_{i=1}^{p} |a_i x_i - a_i y_i|^k \right)^{1/k}. \end{aligned} \tag{8}$$

It is easy to see that weighting of the variables in the distance function is equivalent to a linear scaling of the ith variable by the factor a_i. When constructing a matrix $\mathbf{D}$ with the factors a_i in the diagonal and zeros in all off-diagonal elements, we can thus write $\|\mathbf{x} - \mathbf{y}\|_k' = \|\mathbf{D}(\mathbf{x} - \mathbf{y})\|_k = \|\mathbf{Dx} - \mathbf{Dy}\|_k$.

When we cannot scale the variables independently because of geometric dependencies between the variables, the linear influence of variable j on variable i may be represented by a_{ij}. The non-zero factor a_{ii} is the same as a_i in (8) and corresponds to a linear scaling of the ith variable. The weighted distance is then given by

$$\|\mathbf{x} - \mathbf{y}\|_k'' = \left(\sum_{i=1}^{p} | \sum_{j=1}^{p} a_{ij} x_j - a_{ij} y_j |^k \right)^{1/k}. \tag{9}$$

When building a regular matrix $\mathbf{A}$ with the elements a_{ij} from formula (9), we can write $\|\mathbf{x} - \mathbf{y}\|_k'' = \|\mathbf{A}(\mathbf{x} - \mathbf{y})\|_k = \|\mathbf{Ax} - \mathbf{Ay}\|_k$. Because distances are unaffected by changes of the origin, the weighting in (9) represents all possible affine transformations of the variables.

In most classical morphometric and taxonomic approaches the weighting is largely arbitrary or unknown so that meaningful inferences should be unaffected by the weighting of variables in a distance function; that is to say, they should be invariant under affine transformations of the variables. A meaningful overall distance function or similarity measure is possible only in the presence of a principled weighting scheme and a specification of geometric dependencies. When the weightings in a distance function are unknown, only affine invariant structures can be interpreted in an ordination analysis.

Redundancy of Variables

Equal weighting is also arbitrary because the number of measurements per anatomical unit (and the definition of units or characters) already represents a form of weighting. An anatomical region has a larger effect on a phenotypic distance if it is covered by many measurements, as compared to regions assessed by less measurements. Consider, for example, simple organisms composed of two parts. Two such organisms might be similar in the first part but clearly differ in the second part. If there were many measurements of the first part and only few of the second part,

the organisms would appear quite similar relative to a third one. In the opposite situation—few measurements of the first part and many of the second one—the two organisms would appear more different. One might tend to cover both parts by the same number of measurements and hence give them equal weight, but this decision is arbitrary and not applicable in practice where biological signals and anatomical units cannot be identified in such an idealized way.

More formally, consider three different variables a, b, c and augment them with a fourth variable d that is identical to c except for mean differences. For example, the first three variables might be measures of different parts of an organism while measurement d is spatially closely adjacent to c. Variable c, hence, is affecting the distance measure twice in the four-variable scenario as compared to the first three variables only. For the Euclidean distance this gives

$$\sqrt{\Delta_a^2 + \Delta_b^2 + \Delta_c^2 + \Delta_d^2} = \sqrt{\Delta_a^2 + \Delta_b^2 + 2\Delta_c^2}$$
$$= \sqrt{\Delta_a^2 + \Delta_b^2 + (\sqrt{2}\Delta_c)^2},$$

where Δ_a stands for the difference $a_i - a_j$ between two individuals i and j and similarly for the other variables. One can see that including the variable c twice in the distance function is equivalent to a linear scaling of this variable by the factor $\sqrt{2}$ (or $\sqrt[k]{2}$ for other Minkowski metrics). The vector space of the four variables is of course not a linear transformation of the three-variable space; it simply has one more dimension. But the matrix of four variables still is of rank three, and the first three principal components of a, b, c, d are identical to the principal components of $a, b, \sqrt{2}c$, which are an affine transformation of the original three variables. For real data, variables are of course not perfectly correlated and a change in the redundancy of variables only approximately corresponds to an affine transformation.

In careful morphometric studies, many measurements are spatially closely adjacent and hence highly correlated. In particular, modern geometric morphometric approaches often are based on a large number of spatially closely adjacent landmarks and semilandmarks, giving rise to effective exploratory studies (e.g., Polly and McLeod 2008; Adams et al. 2004; Gunz et al. 2009; Mitteroecker et al. 2005; Coquerelle et al., in press). Adjacent measurements (landmarks), or measurements of bilateral symmetric traits, might thus be reduced to a smaller number of measurements or they could be augmented by additional measurements without leading to major changes of the captured information. But the distribution of measurements across an organism is a kind of linear weighting of anatomical regions or biological signals in any quantification of overall phenotypic distance and in any low-dimensional ordination of such distances. Since the choice of measurements and especially the number of adjacent measurements often is largely arbitrary, meaningful results should be affine invariant.

Applications

In the previous sections we have argued that many phenotype spaces exhibit an affine geometry. We now apply the corresponding condition for meaningfulness to particular statistics of phenotypic variation and disparity. Furthermore, we analyze typical transformations of variables in terms of invariance to affine transformations.

Phenotypic Variation and Disparity

In many morphospace approaches the size or the volume of a region occupied by some taxa is of central interest and interpreted as a measure of phenotypic variability. Phenotypic variability within and between taxa is often referred to as morphological disparity or diversity, and has been assessed frequently in paleobiological and ecological contexts (Foote 1997; Roy and Foote 1997; Wills 2001; Zelditch et al. 2004). Factors responsible for the creation, maintenance, and canalization of phenotypic variance and covariance within a population have been studied both from evolutionary and developmental perspectives (e.g., Lynch and Walsh 1998; Hallgrimsson and Hall 2005; Arnold et al. 2008; Mitteroecker and Bookstein 2009).

Several measures of disparity have been suggested, such as total variance (trace of the covariance matrix or, equivalently, sum of its eigenvalues), the summed Euclidean distances of all specimens from their mean, or the sum of all pairwise Euclidean distances between the specimens. Another common multivariate measure of variance is the determinant of the covariance matrix, which is referred to as generalized variance. The determinant is equal to the product of all eigenvalues and can thus be interpreted as the volume of the orthotope (the multidimensional box) spanned by the scaled eigenvectors.

For a single variable x, the variance $\mathrm{Var}(x)$ is affected by linear scaling of the variable but not by translation:

$$\mathrm{Var(ax + b)} = (n-1)^{-1} \sum_{i=1}^{n} (ax_i + b - a\bar{x} - b)^2$$
$$= (n-1)^{-1} \sum_{i=1}^{n} a^2(x_i - \bar{x})^2 \qquad (10)$$
$$= a^2 \mathrm{Var(x)},$$

where $a \neq 0$ and b are real numbers. However, ratios of variances are invariant to linear scaling:

$$\frac{\mathrm{Var}(ax)}{\mathrm{Var}(ay)} = \frac{a^2\mathrm{Var}(x)}{a^2\mathrm{Var}(y)} = \frac{\mathrm{Var}(x)}{\mathrm{Var}(y)}.$$

Similarly, all multivariate measures of variance or disparity are affected by affine transformations. Only ratios of generalized variances are affine invariant. As mentioned above, generalized variance relates to the volume spanned by the scaled eigenvectors in phenotype space and ratios of volumes are invariant to affine transformations:

$$\frac{\det(\mathbf{A}^T\mathbf{S}_1\mathbf{A})}{\det(\mathbf{A}^T\mathbf{S}_2\mathbf{A})} = \frac{(\det \mathbf{A})^2 \det \mathbf{S}_1}{(\det \mathbf{A})^2 \det \mathbf{S}_2} = \frac{\det \mathbf{S}_1}{\det \mathbf{S}_2}$$

Here, $\mathbf{S}_1$ and $\mathbf{S}_2$ are two covariance matrices, $\mathbf{A}$ is a regular matrix, and the superscript ^{T} denotes the transpose of a matrix. Ratios of total variance are only invariant to affine transformations if the covariance matrices are proportional. That is, if $\mathbf{S}_2 = k\mathbf{S}_1$ then $\mathrm{Tr}(k\mathbf{S}_1)/\mathrm{Tr}(\mathbf{S}_1) = Tr(\mathbf{A}^T k\mathbf{S}_1\mathbf{A})/Tr(\mathbf{A}^T\mathbf{S}_1\mathbf{A}) = k$. In all other cases, ratios of total variance need not be invariant relative to affine transformations, just like summed Euclidean distances (see the "Appendix" for details).

Another affine invariant extension of the notion of ratios of variances are relative or generalized eigenvalues. For two covariance matrices $\mathbf{S}_1$ and $\mathbf{S}_2$, generalized eigenvalues are equal to the eigenvalues of $\mathbf{S}_2^{-1}\mathbf{S}_1$, which are the same as the eigenvalues of $(\mathbf{A}^T\mathbf{S}_2)^{-1}(\mathbf{A}^T\mathbf{S}_1\mathbf{A})$, where $\mathbf{A}$ is again a nonsingular matrix. This follows from the fact that the matrices $\mathbf{S}_2^{-1}\mathbf{S}_1$ and $(\mathbf{A}^T\mathbf{S}_2\mathbf{A})^{-1}(\mathbf{A}^T\mathbf{S}_1\mathbf{A})$ are similar. To see this, set $\mathbf{B} = \mathbf{A}^{-1}$, which exists since $\mathbf{A}$ is nonsingular. Then

$$\mathbf{B}^{-1}((\mathbf{A}^T\mathbf{S}_2\mathbf{A})^{-1}(\mathbf{A}^T\mathbf{S}_1\mathbf{A}))\mathbf{B} = \mathbf{A}\mathbf{A}^{-1}\mathbf{S}_2^{-1}\mathbf{S}_1\mathbf{A}\mathbf{A}^{-1} = \mathbf{S}_2^{-1}\mathbf{S}_1.$$
$$(11)$$

Furthermore, as shown in (10), the translation term of an affine transformation does not affect the calculation of the transformed covariance matrices. Hence, measures based on generalized eigenvalues, such as the metric for phenotypic covariance matrices proposed in Mitteroecker and Bookstein (2009), are meaningful in an affine space. The space of covariance matrices, a Riemannian manifold in the form of a convex cone, thus is a metric space that is locally Euclidean, even when the underlying phenotype space is affine. Also, likelihood ratio tests for comparing multivariate means or covariance matrices are affine invariant as they are based on generalized eigenvalues (see the "Appendix" for more details).

Transformation of Variables

Logarithmic Transformations

It is a common practice in traditional (i.e., non-geometric) morphometrics and quantitative genetics to transform variables to their logarithm (to any base; see, e.g.,

Bookstein et al. 1985; Marcus 1990; Falconer and Mackay 1996). Logarithms have been justified partially from the multivariate generalization of the bivariate allometric equation (Jolicoeur 1963). Furthermore, variances and covariances of logged data are unit free and hence independent of linear scaling factors, i.e., $\mathrm{Var}(\log x) = \mathrm{Var}(\log \lambda x)$ for any $\lambda \neq 0$.

Logarithmic transformations might thus appear as a natural solution to the problem of incommensurable units. However, such transformations also modify many relevant geometric relationships in a phenotype space. For example, while preserving incidence relationships, linear trajectories in the original space may be curved when the variables are log transformed; parallel trajectories may be divergent or convergent and vice versa; many curved trajectories (those with exponential relationships between the variables) would be linear in the logged space. In general, linear varieties do not map onto linear varieties under log transformation. Absolute and relative distances, and hence also variances and covariances, are affected by logarithmic transformations (see Fig. 4 for an example).

This implies that most conclusions drawn from the empirical distribution within a morphospace would not be invariant to logarithmic transformations of the variables. It is thus crucial to specify the correct geometry prior to the analysis. For example, a straight trajectory should indicate a constant process during a certain (evolutionary or developmental) time period. For the raw variables such a straight trajectory results from an additive process, whereas for log transformed variables it follows from a multiplicative process, such as exponential growth of cell populations. Furthermore, log transformed variables should have a meaningful origin since $\mathrm{Var}(\log x) \neq \mathrm{Var}(\log (b + x))$ for any $b \neq 0$. However, many morphometric studies are about bone morphology, where growth and remodeling due to osteoblast and osteoclast activity may more closely resemble an additive rather than a multiplicative process (e.g., Hall 2005).

Mahalanobis Distance

Another common transformation is to divide a variable by its standard deviation σ, so that differences between two specimens or between a specimen and the population mean are expressed as a multiple of σ. The multivariate extension of this *statistical distance* is usually called Mahalanobis distance or generalized statistical distance (Mahalanobis 1936),

$$d_M(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^T \mathbf{S}^{-1}(\mathbf{x} - \mathbf{y})}, \qquad (12)$$

where $\mathbf{x}$ and $\mathbf{y}$ are vectors representing two specimens and $\mathbf{S}$ is a variance-covariance matrix. In most applications $\mathbf{S}$ is
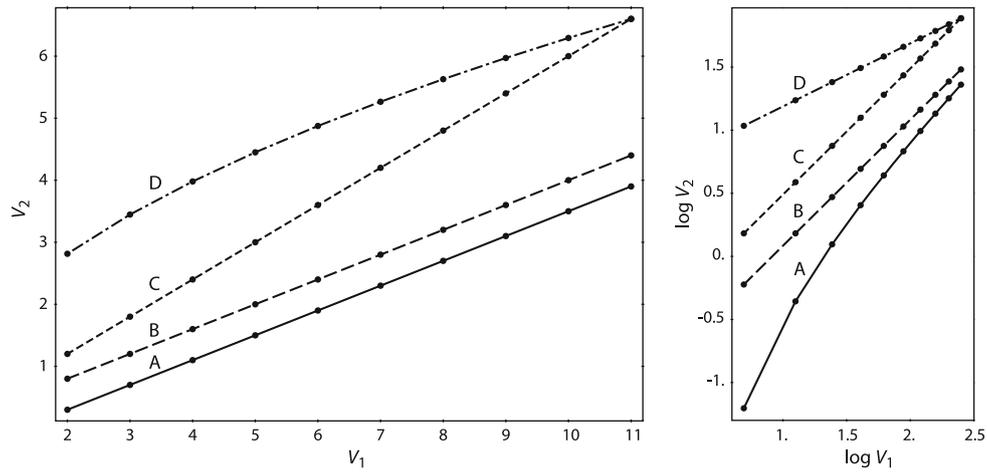
Fig. 4 Four trajectories (*A–D*) of ten specimens each are plotted for the two variables V_1 and V_2 (*left*) and for the natural logarithm of these variables (*right*). While the linear trajectories *A* and *B* are parallel for the original variables, they are oblique for the logged data and *A* is even curved. In contrast, *B* and *C* are linearly diverging for the untransformed data and parallel for the logged variables. The trajectory *D* is linear only after a log transformation. Note also that the variable V_1 has a larger variance than V_2, whereas the opposite is the case for the transformed variables

the (pooled) within-group covariance matrix. The Mahalanobis distance between two points is equal to the Euclidean distance when **S** is the identity matrix or when the data points are transformed by $\mathbf{S}^{-1/2}$, the inverse square root of the covariance matrix (e.g., Mitteroecker and Bookstein 2011). This linear transformation leads to an isotropic within-group distribution by standardizing the affine components of the data space (Fig. 5). Mahalanobis distance thus is invariant to affine transformation of the raw variables.

To show this in more detail, consider the one-dimensional case first. An affine transformation of the two values x and y is given by $ax + b$ and $ay + b$, where $a \neq 0$ and b are real numbers. Let the variance be denoted by σ^2. Then,

the squared Mahalanobis distance between the transformed values is

$$
\begin{aligned}
d_M^2(ax+b, ay+b) &= (ax+b-ay-b)\frac{1}{a^2\sigma^2}(ax+b-ay-b) \\
&= a(x-y)\frac{1}{a^2\sigma^2}a(x-y) \\
&= (x-y)\frac{1}{\sigma^2}(x-y) \\
&= d_M^2(x,y).
\end{aligned}
$$

The last expression is the same as the squared Mahalanobis distance between the original values x and y. The multi-dimensional case proceeds analogously, but for simplicity we may ignore the translation term as it

Fig. 5 a Two random variables for three groups of specimens with a common covariance matrix **S**. The three ellipses are the 90% equal frequency ellipses. When the space is transformed by $\mathbf{S}^{-1/2}$ as shown in **b**, Mahalanobis distance is equal to the Euclidean distance in this transformed space. The transformation is an affine map, which leads to an isotropic (*circular*) average within-group distribution
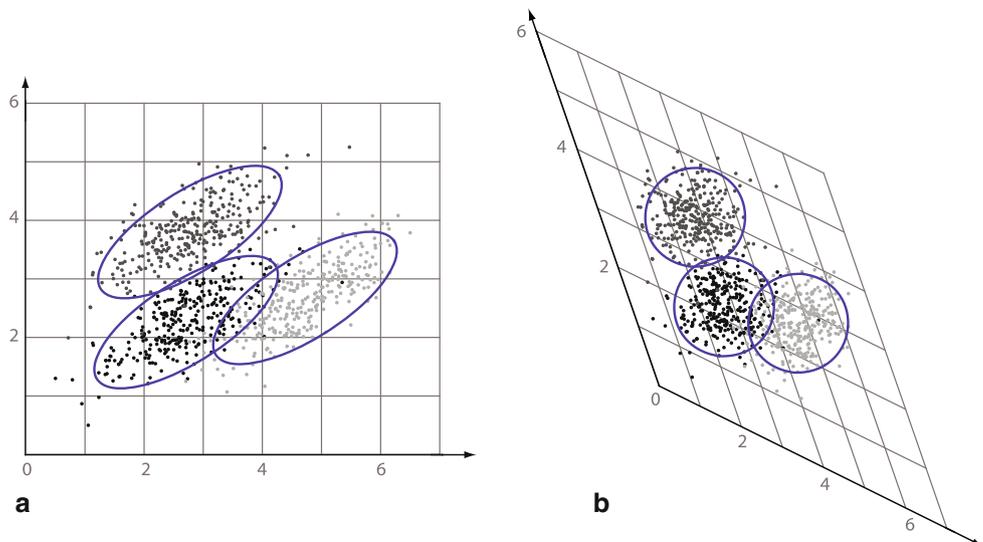
cancels out in the calculation of both distances and variances (see (10)):

$$d_M^2 = (\mathbf{Ax} - \mathbf{Ay})^T (\mathbf{ASA}^T)^{-1} (\mathbf{Ax} - \mathbf{Ay})$$
$$= (\mathbf{x} - \mathbf{y})^T \mathbf{A}^T (\mathbf{A}^T)^{-1} \mathbf{S}^{-1} \mathbf{A}^{-1} \mathbf{A} (\mathbf{x} - \mathbf{y}) \qquad (13)$$
$$= (\mathbf{x} - \mathbf{y})^T \mathbf{S}^{-1} (\mathbf{x} - \mathbf{y})$$

Intuitively, $d_M(\mathbf{x}, \mathbf{y})$ is a distance relative to a reference distribution $\mathbf{S}$, and affine transformations affect the squared Euclidean distance between $\mathbf{x}$ and $\mathbf{y}$ in the same way as $\mathbf{S}$, so that the *relative* distance remains constant.

The interpretation of the resulting Mahalanobis metric crucially depends on the choice of the reference distribution. There are examples in evolutionary quantitative genetics where Mahalanobis distance has a natural meaning (e.g., Lande 1979), but an interpretation in a biometric context can be difficult and computationally unstable. One typical application is the Mahalanobis distance between specimens and different group means in the context of linear discrimination (Mardia et al. 1979). Under the (relatively unrealistic) assumption of homogenous within-group covariance matrices, the squared Mahalanobis distances between a specimen and different group means are proportional to the log likelihoods for the specimen in these groups. The computation of d_M is problematic when $\mathbf{S}$ is singular or nearly singular, which is typically the case in modern morphometrics (Mitteroecker and Bookstein 2011). However, as Euclidean space differs from a space based on Mahalanobis distance by a linear (i.e., affine) transformation, all affine invariant relations are the same in both spaces and do not depend on the choice of a reference distribution.

Discussion

In this paper, we investigated the meaningfulness of biological statements derived from a representing mathematical structure, which are geometric spaces in our case. There is a long history of these considerations, going back to Felix Klein's Erlanger program (Klein, 1872). Klein defined geometrical concepts not by bridging them to physical space, but by tying them to inherent mathematical principles. His basic idea was to identify a geometry with the invariance under certain transformational groups. The fundamental properties of a geometry remain unaffected by the associated group of transformations, whereas other properties may change. As an example consider the case of angles and distances, which are invariant under similarities (Euclidean geometry) but not under affine transformations (affine geometry). The meaningfulness of geometric concepts depends on the groups of transformations defining the geometry (Narens 2002).

This view of meaningfulness is closely related to the problem of measurement scales in the physical and social sciences (Krantz et al. 1970; Suppes et al. 1989; Luce et al. 1990). The study of physical measurement processes based on the assignment of numbers to qualitative (empirical) objects dates back at least to von Helmholtz (1887) and was further worked out by Hölder (1901). It also played an important role in Hilbert's foundations of geometry (Hilbert 1899). While these early approaches to measurement are confined to the measurement of physical quantities like length or temperature, there is also a long standing tradition of measurement studies in mathematical psychology and economics. Stevens (1946) proposed a classification of psychological measurement scales that has immediate consequences for the problem of meaningfulness, for it identifies scales like ordinal scales or ratio scales with particular groups of transformations. These, in turn, determine permissible statistics, which are precisely those that are invariant under the corresponding group of transformations. Similar approaches have been important in economics, particularly for modern utility theory, which goes back to von Neumann and Morgenstern's theory of cardinal utility (von Neumann and Morgenstern 1944). Utility is a numerical representation of a qualitative preference ordering, giving rise to particular measurement scales and thus to implicit conditions for meaningfulness in terms of invariance.

Even though specifying meaningful geometric relations by groups of transformations has a long history in several scientific disciplines, it had less impact on the concept of phenotype spaces (but see, e.g., Bookstein 1991; Dryden and Mardia 1998; Lele and McCulloch 2002). Once an appropriate representing structure is identified, invariance yields a powerful tool to justify the use of certain geometric and statistical methods and to establish limits on what can meaningfully be said about an empirical structure. However, geometric representations of empirical structures are only instances of a more general class, including topological (and weaker) representations as well (see, e.g., Stadler et al. 2001; Stadler et al. 2002; Mitteroecker and Huttegger 2009).

The Fundamental Relations and Properties in a Phenotype Space

We showed that typical phenotype spaces, such as Raup's space of coiled shells or the spaces used in traditional morphometrics and numerical taxonomy, often are characterized by an affine rather than an Euclidean geometry. Kendall's shape space, the central phenotype space in landmark-based geometric morphometrics, is a space in which a point corresponds to the shape of a landmark configuration. Kendall (1984) showed that this space is a Riemannian manifold, which can be approximated locally by a Euclidean tangent space. But given the more or less

arbitrary spacing and number of landmarks, even this phenotype space should be considered as a structure that locally is similar to an affine space. The concepts of distance, angle, and volume are not meaningful in an affine space, but several other concepts are invariant to affine transformations.

The most fundamental geometric relationship in an affine space is *incidence*. Phenotypes inhabiting the same position in a space are identical; phenotypes lying on another structure such as a line or a plane are elements of these sets of phenotypes. For example, statements about the overlap of groups (the intersection of convex hulls) are affine invariant. A specimen within a cluster of specimens belongs to this cluster even under affine transformation. Raup's finding that different taxa inhabited different areas in his morphospace whereas other areas remain empty thus is a meaningful statement. Furthermore, intersection, overlap, or identity of phenotypic trajectories are affine invariant relations.

Linearity is a fundamental property in an affine space. A linear trajectory remains linear under all affine transformations and indicates a constant additive process. A curved developmental trajectory, by contrast, would indicate a nonlinear processes or, alternatively, a combination of two or more additive developmental processes. For log transformed variables a linear trajectory would indicate a continual multiplicative process. Furthermore, a point lying in between two other points is an affine invariant relation so that phenotypic *intermediacy* is a meaningful property.

Parallel lines remain parallel under affine transformations. Two parallel trajectories represent the same additive processes whereas oblique trajectories indicate different processes. For instance, overlapping and parallel developmental trajectories can be indicative of heterochronic developmental processes (Mitteroecker et al. 2005). Furthermore, *distances along parallel lines* can be compared meaningfully, i.e., for two parallel trajectories (identical processes), the length of these trajectories (the extent or magnitude of these processes) can be assessed. The following is an example of a meaningful statement: "Two organisms grow in the same way (direction), but one organism experiences more growth along this common pattern than the other one". However, distances along different directions, like the amount of growth or evolution along different directions, cannot be related meaningfully in a quantitative way.

Ratios of volumes are constant under affine transformations even though volumes per se are not meaningful in an affine space. As a consequence, no measures of disparity or total phenotypic variance can be interpreted directly, only ratios of generalized variance (determinant of the covariance matrix) are meaningful (take Raup's morphospace as a paradigmatic example). Moreover, the metric for

phenotypic covariance matrices proposed in Mitteroecker and Bookstein (2009) as well as other statistics based on generalized eigenvalues are affine invariant.

Distances relative to a reference distribution, such as Mahalanobis distances, are affine invariant. That is to say, although distances are not generally invariant relative to affine transformation, they may become so by supplementing more information about the distribution of data points, as in the case of the Mahalanobis distance. Therefore, the likelihood of a specimen belonging to a certain group can be computed, and discrimination and classification problems can be meaningfully approached (see the "Appendix").

All these listed geometric relations are invariant under a relatively large class of arbitrary choices leading to affine transformations of the phenotype space, whereas phenotypic distances, angles, and volumes are not guaranteed to be meaningful. They are also unaffected by small modifications of the number and spacing of measurements. Scientific hypotheses should thus be phrased and tested in terms of these geometric relationships within phenotype spaces. The interpretation of angles and phenotypic distances in different directions should be avoided.

Phenotypic Distances

We showed that in many cases it will not be possible to define a meaningful distance measure between phenotypes. This is a controversial position, and it deserves some further comments.

Overall measures of phenotypic similarity or dissimilarity are averages over a range of different characters and signals, so that the emerging measure might lack a meaningful biological interpretation. Consider two mice with 5 mm difference in the length of their tails and another two mice with 5 mm difference in head length. Is it valid to say that both pairs of mice have the same phenotypic distance? What if the skull had been measured by 20 different measurements and the tail only by one, so that skull differences would have a much larger impact on a phenotypic distance than tail differences? A similar ambiguity arises when comparing the difference in, e.g., a length measure to a difference in an angle. The inability to define a meaningful general notion of phenotypic distance leads to an affine phenotype space rather than a Euclidean one. In an affine space only distances along parallel directions are comparable, that is, distances along the same characteristics or combination of characteristics. When two organisms both grow 5 mm in the forelimbs and 8 mm in the hindlimbs, it is meaningful to say that they have the same additive growth pattern (parallel trajectories of the same length in phenotype space). Another organism with 10 mm forelimb growth and 16 mm hindlimb growth would have

the same direction or pattern of growth as the previous ones, but a twofold amount of growth (a parallel trajectory of twofold length). Likewise, a notion of intermediacy is meaningful in affine phenotype spaces. For instance, in the tendency to reduce facial size and increase brain size, *Australopithecus* is approximately in between chimpanzees and humans. It is, however, ambiguous to assess how much humans differ from chimpanzees as compared to the morphological distance between chimpanzees and gorillas because they differ in other characteristics (distances in different directions in phenotype space).

A meaningful general phenotypic distance can only be defined under certain (relatively unrealistic) conditions. Derived from some functional or developmental model, it might be possible to select a small number of measurements with known geometric dependencies that capture the expected signals or determine a specific function. When these measurements are of the same unit and of equal weight, they could be construed as the axes of a Euclidean phenotype space. It is also tempting to asses biological structures by a large set of equally spaced landmarks or semilandmarks (e.g., Schaefer et al., 2006; Polly and McLeod 2008; Gunz et al. 2009) and to treat them as descriptors of equal weight and with similar geometric dependencies. Results based on such phenotypic distances should still be interpreted with great care. For example, larger elements would be covered by more landmarks and hence would have a larger impact on phenotypic distance than smaller structures. In general, the distance between two phenotypes A and B is guaranteed to be larger than the distance between A and C only if this holds for all reasonable weightings or transformations, that is, for all linear combinations of characters.

We showed that distances relative to a reference distribution, such as Mahalanobis distance, are affine invariant. The space induced by Mahalanobis distance preserves all affine invariant relations (Euclidean space relates to this space by an affine transformation), but the interpretability of all additional geometric relations, including distance, depends on the scientific significance and the computational stability of the reference distribution. Logarithmic transformations, by contrast, modify most geometric relations in phenotype space.

In general, the properties of the representing geometric structure (the phenotype space) depends on the empirical structure in question. Particular empirical structures (certain measures or descriptions, particular relations of interest) may also result in groups of transformations different from affine or Euclidean transformations. In such cases, the meaningfulness of notions like distance has to be assessed again. It may be the case that some notion of distance can be found that is meaningful with respect to the given group of transformations. Thus, in a certain sense affine space constitutes a worst-case scenario for morphospaces. Its structure could be improved by new information; for example, specifications of the geometric dependencies between the variables could fix the angles between the axes in a coordinate system. Without such additional information, however, affine spaces appear to be the most appropriate choice of geometry for many phenotype spaces.

One topic that we only have implicitly touched in this paper is the problem of how to determine whether a mathematical representation matches an empirical structure. This is at the heart of the problem of meaningfulness. We have motivated the choice of affine geometries in the case of classical morphospaces, and we have pointed out that we expect similar structures to be appropriate for other phenotype spaces. A deeper justification would proceed along the lines of establishing representation theorems between synthetic and analytic geometries (Suppes et al. 1989). This requires an axiomatization of the properties and relations governing the empirical structure and showing that it is uniquely represented by some geometry (or some other mathematical structure). Such an approach will yield much deeper insights than the one offered in our paper and is left for future work.

Appendix: Affine Invariant Statistics

Incidence relations, averages, and also distances relative to a statistical distribution are affine invariant. Consequently, many statistical tests can be applied to data with an affine structure. For example, the classical test to compare multivariate means is based on the Hotelling's T^2 statistic

$$T^2 = n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^T \mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0),$$

where n is the number of cases, $\bar{\mathbf{x}}$ a p-dimensional vector representing the estimated mean, $\boldsymbol{\mu}_0$ the hypothesized mean, and $\mathbf{S}$ is a $p \times p$ sample covariance matrix. If $\mathbf{x}$ is a random variable with a multivariate normal distribution and $\mathbf{S}$ has a Wishart distribution with $m = n - 1$ degrees of freedom and is independent of $\mathbf{x}$, then T^2 has a Hotelling's T^2 distribution with the parameters p and m. The Hotelling's T^2 statistic is equal to the squared Mahalanobis distance (11) multiplied by n. We have shown in (12) that Mahalanobis distance is invariant to affine transformation and hence T^2 is affine invariant too. It follows that multivariate means can be tested meaningfully even if their underlying geometric structure is affine instead of Euclidean.

An extension of the Hotelling's T^2 distribution is the Wilks' lambda distribution. For two covariance matrices $\mathbf{S}_1$

and $\mathbf{S}_2$ with Wishart distributions of m and n degrees of freedom, Wilks lambda is

$$\Lambda = \frac{\det(\mathbf{S}_1)}{\det(\mathbf{S}_1 + \mathbf{S}_2)} = \frac{1}{\det(\mathbf{I} + \mathbf{S}_1^{-1}\mathbf{S}_2)}$$
$$= \prod\left(\frac{1}{1+\lambda_i}\right) \sim \Lambda(p, m, n),$$

where $\mathbf{I}$ is the identity matrix and λ_i is the ith eigenvalue of $\mathbf{S}_1^{-1}\mathbf{S}_2$. In several likelihood ratio tests (e.g., in the context of MANOVA) $\mathbf{S}_1$ is the error variance and $\mathbf{S}_2$ the variance explained by some model. As shown in (11), the eigenvalues of $\mathbf{S}_1^{-1}\mathbf{S}_2$ are equal to the eigenvalues of $(\mathbf{A}^T\mathbf{S}_1\mathbf{A})^{-1}\mathbf{A}^T\mathbf{S}_2\mathbf{A}$, so that Λ is affine invariant.

For the likelihood ratio test of homogeneity of covariance matrices, i.e., of $H_0: \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \cdots = \mathbf{\Sigma}_k$, the maximum likelihood estimate of $\mathbf{\Sigma}_i$ is $\mathbf{S} = n^{-1}\sum n_i \mathbf{S}_i$ under H_0 and $\mathbf{S}_i$ under the alternative, where n_i is the sample size of the ith group and $n = \sum n_i$. The likelihood ratio

$$-2 \log \lambda_a = \sum_{i=1}^{k} n_i \log \det(S_i^{-1}S),$$

has an asymptotic χ^2 distribution with $\frac{1}{2}p(p+1)(k-1)$ degrees of freedom (Mardia et al. 1979, p. 140). As the eigenvalues of $\mathbf{S}_i^{-1}\mathbf{S}$ are affine invariant (11), this test is meaningful also for data with an affine structure. Similarly, the metric for covariance matrices

$$d_{cov}(\mathbf{S}_1, \mathbf{S}_2) = \sqrt{\sum_{i=1}^{p}(\log \lambda_i)^2},$$

where λ_i is the ith eigenvalues of $\mathbf{S}_2^{-1}\mathbf{S}_1$, is invariant to affine transformation (Mitteroecker and Bookstein 2009).

For the multivariate multiple regression of $\mathbf{Y}$ on $\mathbf{Z}$, the least squares estimates of the regression coefficients are given by $\boldsymbol{\beta} = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{Y}$ and the predicted values are $\hat{\mathbf{Y}} = \mathbf{Z}\boldsymbol{\beta}$. Affine transformation of the predictors $\mathbf{Z}$ has no affect on the prediction because

$$\hat{\mathbf{Y}} = \mathbf{Z}\mathbf{A}(\mathbf{A}^T\mathbf{Z}^T\mathbf{Z}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{Z}^T\mathbf{Y}$$
$$= \mathbf{Z}\mathbf{A}\mathbf{A}^{-1}(\mathbf{Z}^T\mathbf{Z})^{-1}(\mathbf{A}^T)^{-1}\mathbf{A}^T\mathbf{Z}^T\mathbf{Y}$$
$$= \mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{Y}.$$

As Fisher's linear discriminant function is computationally equivalent to a multiple regression of a group variable on the phenotypic variables, the success of linear discrimination is unaffected by affine transformations (see Mitteroecker and Bookstein 2011, for an explicit proof). When the dependent variables $\mathbf{Y}$ are transformed into $\mathbf{Y}\mathbf{A}$, the predicted values result from the same transformation $\hat{\mathbf{Y}}\mathbf{A} = \mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{Y}\mathbf{A}$.

We have argued that only ratios of generalized variance (determinant of the covariance matrix) are affine invariant,

whereas generalized variance itself and also total variance (trace of the covariance matrix) and ratios of total variance are affected by affine transformations. We demonstrate this here by an example. Consider two covariance matrices

$$\mathbf{S}_1 = \begin{pmatrix} 2.3 & 1.2 \\ 1.2 & 1.8 \end{pmatrix}, \mathbf{S}_2 = \begin{pmatrix} 2.1 & 0.6 \\ 0.6 & 2.4 \end{pmatrix}$$

and the transformation matrix

$$\mathbf{A} = \begin{pmatrix} 0.8 & -0.3 \\ 0.2 & 1.2 \end{pmatrix}.$$

The total variance of $\mathbf{S}_1$ is $Tr(\mathbf{S}_1) = 4.10$ and that of the transformed data $Tr(\mathbf{A}^T\mathbf{S}_1\mathbf{A}) = 3.86$, and similarly for the generalized variances $Det(\mathbf{S}_1) = 2.70$ and $Det(\mathbf{A}^T\mathbf{S}_1\mathbf{A}) = 2.81$. The ratios of total variance are $Tr(\mathbf{S}_1)/Tr(\mathbf{S}_2) = 0.91$ and $Tr(\mathbf{A}^T\mathbf{S}_1\mathbf{A})/Tr(\mathbf{A}^T\mathbf{S}_2\mathbf{A}) = 0.80$, whereas the ratio of the generalized variance is affine invariant: $Det(\mathbf{S}_1)/Det(\mathbf{S}_2) = Det(\mathbf{A}^T\mathbf{S}_1\mathbf{A})/Det(\mathbf{A}^T\mathbf{S}_2\mathbf{A}) = 0.58$. Ratios of total variance are affine invariant only if $\mathbf{S}_2 = k\mathbf{S}_1$.

References

Adams, D., Rohlf, F. J., & Slice, D. E. (2004). Geometric morphometrics: Ten years of progress following the "revolution". *Italian Journal of Zoology, 71*(9), 5–16.

Arnold, S. J., Bürger, R., Hohenlohe, P. A., Ajie, B. C., & Jones, A. G. (2008). Understanding the evolution and stability of the G-matrix. *Evolution, 62*, 2451–2461.

Arnold, S. J., Pfrender, M. E., & Jones, A. G. (2001). The adaptive landscape as a conceptual bridge between micro- and macroevolution. *Genetica, 112–113*, 9–32.

Blackith, R. E., & Reyment, R. A. (1971). *Multivariate morphometrics*. New York: Academic Press.

Bookstein, F., Chernoff, B., Elder, R. L., Humphries, J. M., Jr., Smith, G. R., & Strauss, R. E. (1985). Morphometrics in evolutionary biology: The geometry of size and shape change, with examples from fishes. Academy of Sciences of Philadelphia Special Publication 15.

Bookstein, F. L. (1991). *Morphometric tools for landmark data: Geometry and biology*. Cambridge, UK: Cambridge University Press.

Busemann, H. (1955). *The geometry of geodesics*. New York: Academic Press.

Cain, A. J., & Harrison, G. A. (1960). Phyletic weighting. *Proceedings of the Zoological Society London, 135*, 1–31.

Coquerelle, M., Bookstein, F. L., Braga, J., Halazonetis, D. J., Weber, G. W., & Mitteroecker, P. (in press). Sexual dimorphism of the human mandible and its association with dental development. *American Journal of Physical Anthropology*.

Dryden, I. L., & Mardia, K. (1998). *Statistical shape analysis*. New York: Wiley.

Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to quantitative genetics*. Essex, UK: Longman.

Fontana, W., & Schuster, P. (1998a). Continuity in evolution: On the nature of transitions. *Science, 280*, 1451–1455.

Fontana, W., & Schuster, P. (1998b). Shaping space: The possible and the attainable in RNA genotype-phenotype mapping. *Journal of Theoretical Biology, 194*, 491–515.

Foote, M. (1997). The evolution of morphological diversity. *Annual Review of Ecology and Systematics, 28*, 129–152.

Frank, S. A., & Smith, D. E. (2010). Measurement invariance, entropy, and probability. *Entropy, 12*, 289–303.

Gunz, P., Bookstein, F. L., Mitteroecker, P., Stadlmayr, A., Seidler, H., & Weber, G. W. (2009). Early modern human diversity suggests subdivided population structure and a complex out-of-africa scenario. *Proceedings of the National Academy of Sciences, USA, 106*(15), 6094–6098.

Hall, B. K. (2005). *Bones and cartilage: Developmental skeletal biology*. San Diego: Academic Press.

Hallgrimsson, B., & Hall, B. K. (2005). *Variation: A central concept in biology*. New York: Elsevier.

Hilbert, D. (1899). *Grundlagen der Geometrie*. Stuttgart: Teubner.

Hölder, O. (1901). Die Axiome der Quantität und die Lehre vom Maß. Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig. *Mathematisch-Physikalische Classe, 53*, 1–64.

Jolicoeur, P. (1963). *The multivariate generalization of the allometry equation. Biometrics, 19*, 497–499.

Kendall, D. G. (1984). Shape manifolds, Procrustean metrics and complex projective spaces. *Bulletin of the London Mathematical Society, 16*(2), 81–121.

Klein, F. (1872). *Vergleichende Betrachtungen über neuere geometrische Forschungen*. Erlangen: Verlag von Andreas Deichert.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement vol I. Additive and polynomial representations*. San Diego, CA: Academic Press (reprinted by Dover 2007).

Lande, R. (1979). Quantitative genetic analysis of multivariate evolution, applied to brain-body size allometry. *Evolution, 33*, 402–416.

Lele, S. R., & McCulloch, C. E. (2002). Invariance, identifiability, and morphometrics. *Journal of the American Statistical Association, 97*, 796–806.

Luce, R. D., Krantz, D. H., Suppes, P., & Tversky, A. (1990). *Foundations of measurement vol III. Representation, axiomatization, and invariance*. San Diego, CA: Academic Press (reprinted by Dover 2007).

Lynch, M., & Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sunderland, MA: Sinauer Associates.

Mahalanobis, P. C. (1936). On a generalised distance in statistics. *Proceedings of the National Institute of Science, India, 2*, 49–55.

Marcus, L. F. (1990). Traditional morphometrics. In F. J. Rohlf & F. L. Bookstein (Eds.), *Proceedings of the Michigan morphometrics workshop* (pp. 77–122). Ann Arbor, Michigan: Univ. Michigan Museums.

Marcus, L. F., Hingst-Zaher, E., & Zaher, H. (2000). Application of landmark morphometrics to skulls representing the orders of living mammals. *Hystrix, Italian J. Mammology, 11*(1), 27–47.

Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis*. London: Academic Press.

Maynard Smith, J. (1970). Natural selection and the concept of a protein space. *Nature, 225*, 563–564.

McGhee, G. R. (1999). *Theoretical morphology: The concept and its applications*. New York: Columbia University Press.

Mitteroecker, P. (2009). The developmental basis of variational modularity: Insights from quantitative genetics, morphometrics, and developmental biology. *Evolutionary Biology, 36*, 377–385.

Mitteroecker, P., & Bookstein, F. (2009). The ontogenetic trajectory of the phenotypic covariance matrix, with examples from craniofacial shape in rats and humans. *Evolution, 63*(3), 727–737.

Mitteroecker, P., & Bookstein, F. (2011). Classification, linear discrimination, and the visualization of selection gradients in modern morphometrics. *Evolutionary Biology, 38*, 100–114.

Mitteroecker, P., & Bookstein, F. L. (2007). The conceptual and statistical relationship between modularity and morphological integration. *Systematic Biology, 56*(5), 818–836.

Mitteroecker, P., & Gunz, P. (2009). Advances in geometric morphometrics. *Evolutionary Biology, 36*, 235–247.

Mitteroecker, P., Gunz, P., & Bookstein, F. L. (2005). Heterochrony and geometric morphometrics: A comparison of cranial growth in pan paniscus versus pan troglodytes. *Evolution & Development, 7*(3), 244–258.

Mitteroecker, P., Gunz, P., Bernhard, M., Schaefer, K., & Bookstein, F. L. (2004). Comparison of cranial ontogenetic trajectories among great apes and humans. *Journal of Human Evolution, 46*, 679–697.

Mitteroecker, P., & Huttegger, S. M. (2009). The concept of morphospaces in evolutionary and developmental biology: Mathematics and metaphors. *Biological Theory, 4*(1), 54–67.

Narens, L. (2002). *Theories of meaningfulness*. Mahwah, NJ: Lawrence Erlbaum Associates.

Pie, M. R. & Weitz, J. S. (2005). A null model of morphospace occupation. *American Naturalist, 166*(1), E1–13.

Polly, P. D., & McLeod, N. (2008). Locomotion in fossil carnivora: An application of eigensurface analysis for morphometric comparison of 3D surfaces. *Palaeontologia Electronica, 11*(2), 10A:13p.

Raup, D. M. (1966). Geometric analysis of shell coiling: General problems. *Journal of Paleontology, 40*, 1178–1190.

Raup, D. M., & Michelson, A. (1965). Theoretical morphology of the coiled shell. *Science, 147*, 1294–1295.

Rohlf, F. J. (1999). Shape statistics: Procrustes superimpositions and tangent spaces. *Journal of Classification, 16*, 197–223.

Rohlf, F. J., & Sokal, R. (1965). Coefficients of correlation and distance in numerical taxonomy. *University of Kansas Science Bulletin, 45*, 3–27.

Roy, K., & Foote, M. (1997). Morphological approaches to measuring biodiversity. *Trends in Ecology and Evolution, 12*(7), 277–281.

Schaefer, K., Lauc, T., Mitteroecker, P., Gunz, P., & Bookstein, F. L. (2006). Dental arch asymmetry in an isolated Adriatic community. *American Journal of Physical Anthropology, 129*(1), 132–142.

Schindel, D. E. (1990). Unoccupied morphospace and the coiled geometry of gastropods: Architectural constraints or geometric covariation? In R. A. Ross, & W. D. Allmon (Eds.), *Causes of Evolution* (pp. 270–304). Chicago: University of Chicago Press.

Slice, D. E. (2001). Landmark coordinates aligned by Procrustes analysis do not lie in Kendall's shape space. *Systematic Biology, 50*(1), 141–149.

Sneath, P. H. A., & Sokal, R. R. (1973). *Numerical taxonomy: The principles and practice of numerical classification*. San Francisco: W. H. Freeman.

Sokal, R. (1961). Distance as a measure of taxonomic similarity. *Systematic Zoology, 10*, 70–79.

Sokal, R. R., & Sneath, P. H. A. (1963). *Principles of numerical taxonomy*. San Francisco: W. H. Freeman.

Stadler, B. M. R., Stadler, P. F., Shpak, M., & Wagner, G. P. (2002). Recombination spaces, metrics, and pretopologies. *Zeitschrift für physikalische Chemie, 216*, 217–274.

Stadler, B. M. R., Stadler, P. F., Wagner, G. P., & Fontana, W. (2001). The topology of the possible: Formal spaces underlying patterns of evolutionary change. *Journal of Theoretical Biology, 213*, 241–274.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*, 677–680.

Suppes, P., Krantz, D. H., Luce, R. D., & Tversky, A. (1989). *Foundations of measurement Vol II. Geometrical, threshold and probabilistic representations*. San Diego, CA: Academic Press (reprinted by Dover 2007).

von Helmholtz, H. (1887). *Zählen und Messen erkenntnistheoretisch betrachtet*. Leipzig: Philosophische Aufsätze Eduard Zeller gewidmet.

von Neumann, J., Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.

Wills, M. A. (2001). Morphological disparity: A primer. In J. M. Adrain, G. D. Edgecombe, & B. S. Lieberman (Eds.), *Fossils, phylogeny, and form—An analytical approach*. New York: Kluwer.

Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proceedings of the sixth international congress of genetics, 1*, 356–366.

Zelditch, M. L., Swiderski, D., Sheets, D. S., & Fink, W. L. (2004). *Geometric morphometrics for biologists*. San Diego: Elsevier.