Individual and Group Differences in Multivariate Domains:
What Happens When the Number of Traits Increases?

Marco Del Giudice

University of New Mexico

Address correspondence to Marco Del Giudice, Department of Psychology, University of New Mexico. Logan Hall, 2001 Redondo Dr. NE, Albuquerque, NM 87131, USA; email: marcodg@unm.edu

Abstract

The major domains of psychological variation are intrinsically multivariate, and can be mapped at various levels of resolution—from broad-band descriptions involving a small number of abstract traits to fine-grained representations based on many narrow traits. As the number of traits increases, the corresponding space becomes increasingly high-dimensional, and intuitions based on low-dimensional representations become inaccurate and misleading. The consequences for individual and group differences are profound, but have gone largely unrecognized in the psychological literature. Moreover, alternative distance metrics show distinctive behaviors with increasing dimensionality. In this paper, I offer a systematic yet accessible treatment of individual and group differences in multivariate domains, with a focus on high-dimensional phenomena and their theoretical implications. I begin by introducing four alternative metrics (the Euclidean, Mahalanobis, city-block, and shape distance) and reviewing their geometric properties. I also examine their potential psychological significance, because different metrics imply different cognitive models of how people process information about similarity and dissimilarity. I then discuss how these metrics behave as the number of traits increases. After considering the effects of measurement error and common methods of error correction, I conclude with an empirical example based on a large dataset of self-reported personality.

Keywords: distance metrics; high-dimensional statistics; individual differences; multivariate domains.

Introduction

The major domains of psychological variation are intrinsically multivariate. Personality, cognitive ability, interests, and values can all be represented as multidimensional trait spaces and mapped at various levels of resolution—from broad-band descriptions involving one or a few abstract traits to fine-grained representations based on many narrow, specific traits. For example, standard psychometric models of intelligence include a general factor (*g*), about 8-16 broad ability factors, and dozens of narrow abilities (see McGrew, 2009). In the field of personality, the canonical "Big Five"—*Extraversion, Agreeableness, Conscientiousness, Neuroticism/Emotional Stability, and Openness*—show a higher-order structure that can be summarized by two metatraits (*Stability* and *Plasticity*; DeYoung et al., 2002; Saucier, 2009), or even a single "general factor of personality" (though the substantive meaning of this factor is disputed; see Just, 2011; Davies et al., 2015; van der Linden et al., 2017). Descending in the hierarchy, the Big Five (known as "domains" in the terminology of the five-factor model) can be parsed into ten narrower "aspects", and further subdivided into 30-45 facets (DeYoung et al., 2007). Cattell's 16PF model describes five global factors and 16 primary factors (15 personality traits plus reasoning/intelligence; Cattell & Schuerger, 2003); while the six factors of the HEXACO model can be refined into 24 facets (Lee & Ashton, 2004). In recent years, some researchers have argued that single personality items—which range from dozens to hundreds in typical questionnaires—may describe trait-like patterns of behavior below the level of facets (e.g., Mõttus et al., 2019; Revelle et al., 2021).

The Strange World of High-Dimensional Spaces

When a psychological domain is represented geometrically as a multidimensional space, an individual's combination of traits or *profile* is described by a point in that space. Likewise, the location of a group of individuals can be summarized by its multivariate mean or *centroid*, which corresponds to the average profile for that group. As the number of traits used to map a given domain increases, the corresponding space becomes increasingly high-dimensional. The consequences for individual and group differences are profound—and yet, they have gone largely unrecognized in the psychological literature. Our geometric intuitions are inevitably based on two- and three-dimensional representations. Lacking direct visualization of higher-dimensional spaces, it is natural to assume that distributions in 10, 20, or 100 dimensions will behave in key respects like their low-dimensional counterparts. But this is a mistake. High-dimensional spaces are not just large: they are vast and sparse in ways that stretch imagination, giving rise to a host of important but counterintuitive phenomena. These phenomena are usually discussed as part of the so-called "curse of dimensionality" (or its mirror image, the "blessing of dimensionality;" see Altman & Krzywinski, 2018; Gorban et al., 2020).

Consider a normal distribution, which is a reasonable approximation for many psychometric traits. In the familiar uni- and bivariate cases, the mass of the distribution clusters around the mean (or centroid), and only a small proportion of points are located in the tails. But as dimensionality increases, a larger proportion of the probability mass becomes concentrated in the tail region, where the probability density is low. Stated otherwise, the majority of the points move far away from the centroid, along a progressively thinner "shell" that envelopes a mostly empty interior (Giraud, 2015; van Tilburg, 2019). As the points disperse further in space, the

distribution of distances gets narrower relative to their size (*distance concentration*), with the result that all the points in the distribution tend to become approximately equally distant from one another, as well as from the centroid (see Aggarwal et al., 2001; Altman & Krzywinski, 2018; Giraud, 2015). Meanwhile, if the points belong to two or more distributions with different centroids, small differences across multiple dimensions tend to cumulate, sharpening the overall separation between the distributions even if they overlap substantially on any individual variable. This often allows one to classify individual points into different groups or clusters with increasing accuracy as dimensionality increases (Bennett et al., 1999; Zimek et al., 2012; see also Gorban et al., 2020).

Implications for Individual and Group Differences

For an example of how these statistical phenomena can have important consequences in the real world, consider my and my colleagues' work on multivariate sex differences in personality. Until recently, the consensus in the literature was that overall sex differences in personality and behavior are small, because male and female distributions largely overlap on most psychological variables (Hyde, 2005, 2014; Zell et al., 2015). But this is only true if variables are considered one by one. When differences across the traits that make up a domain are aggregated into a multivariate effect size, the separation between the sexes can increase dramatically (Del Giudice, 2009; Del Giudice et al., 2012). Predictably, the effect is stronger when the domain is mapped with many narrow traits (e.g., the 30 facets of the Big Five) compared with a few broad traits (e.g., the Big Five). When personality is measured at the level of facets, the overall difference between the average male and female profiles in English-speaking countries is consistently larger than two standard deviations, corresponding to an overlap of less than 30% (Del Giudice, 2022; Del Giudice et al., 2012; Kaiser, 2019; Kaiser et al., 2020). For comparison, a detailed study of facial anatomy in males and females found an overall sex difference of approximately three standard deviations, corresponding to an overlap of about 10% between the distributions of male and female faces (Hennessy et al., 2005).

Another notable application of these concepts can be found in a recent paper by van Tilburg (2019), brilliantly titled "*It's not unusual to be unusual.*" The author noted that, as the number of traits used to describe personality increases, the frequency of "average" personality profiles (i.e., those close to the distribution centroid) can be expected to decrease very quickly. By including more than a handful of traits, one ends up in a paradoxical situation in which almost every individual in the population is highly "unusual" when compared with the average— in other words, the average personality profile ceases to be typical in any meaningful sense. The same phenomenon has been recognized for some time in the field of face perception, where it is known as the "typicality paradox" (Burton & Vokey, 1998). Individual faces can be represented as profiles of morphological features, and thus located in a multivariate *face space* that spans dozens if not hundreds of dimensions (Valentine et al., 2016; more on this below). Owing to the high-dimensional nature of this space, "average" faces turn out to be surprisingly rare, and observers rate the majority of people's faces as distinctive rather than typical (Burton & Vokey, 1998; Lewis et al., 2014; Valentine et al., 2016).

From these brief examples, it is clear that high-dimensional phenomena have important implications for our understanding of psychological variation. However, they have yet to be

explored in a systematic fashion. An especially important issue that awaits investigation is the impact of using alternative distance metrics. For example, van Tilburg (2019) employed the familiar Euclidean distance to measure differences between individual and average profiles. In contrast, research on multivariate sex differences relies on the Mahalanobis distance, which is the multivariate equivalent of Cohen's *d* (see Del Giudice, 2022). These metrics have different properties, and respond in different ways to patterns of correlations among traits (more on this below). This is important because, in most psychological domains, narrow traits are not just more numerous than their broad-band counterparts, but also more strongly correlated to one another. Besides the Euclidean and Mahalanobis distances, several other indices can be used to measure profile (dis)similarity (e.g., Carroll & Field, 1974; Cronback & Gleser, 1953; Skinner, 1978); not only do different metrics have different psychological implications, they also show distinctive behaviors in high-dimensional scenarios. Researchers dealing with multivariate domains should be fully aware of how the choice of a metric can affect the results and interpretation of their studies.

Overview of the Paper

In this paper, I offer a systematic yet accessible treatment of individual and group differences in multivariate domains, with a focus on high-dimensional phenomena and their theoretical implications. I begin by introducing four alternative distance metrics, reviewing their geometric properties, and examining the significance of those properties from a cognitive standpoint. I then discuss how these metrics behave as the number of traits increases, and their potential uses in describing individual and group variation. After considering the effects of measurement error and common methods of error correction, I conclude with an empirical example based on a large dataset of self-reported personality. Throughout the paper, I illustrate the concepts I present with examples from the study of faces and face perception (Bruce & Young, 2012). Human faces are familiar, psychologically salient, highly multidimensional, and show significant patterns of group differences (e.g., between males and females). For these reasons, they are a great source of analogies, and can be used to build reliable intuitions on a topic that is far from intuitive.

<div align="center">

Alternative Distance Metrics and Their Meaning

</div>

Geometrically, the difference between two multivariate profiles (of individuals and/or groups) corresponds to the distance between two points in the *k*-dimensional space defined by the *k* traits of interest. The question of how to best compare psychological profiles is not new; there is a venerable methodological literature on this topic going back to the 1930s, and more than a dozen (dis)similarity indices that have been proposed and tested over the years (Carroll & Field, 1974; Cronbach & Gleser, 1953; Furr, 2010; McCrae, 2008; Skinner, 1978; see also Jones & Furnas, 1987). In this paper I focus on four indices that satisfy the axioms of a metric[1] (meaning that they are proper distances from a geometric standpoint), are clearly interpretable, and—taken together—cover most of the useful ground: the *Euclidean distance* (D_2), the *Mahalanobis distance* (D_M), the *city-block distance* (D_1), and the *shape distance* (D_S).

---

[1] The three axioms are: (1) the distance between *x* and *y* is zero if and only if $x = y$; (2) the distance between *x* and *y* is equal to the distance between *y* and *x* (symmetry); and (3) the distance between *x* and *z* is less than, or equal to, the sum of the distances between *x* and *y* and between *y* and *z* (triangle inequality).

All these metrics can be used to make objective comparisons between individual and/or group profiles, with different costs and benefits depending on the task at hand. From this perspective, distance metrics are quantification tools that researchers use to turn differences between profiles into numerical values, which can then be subjected to various kinds of analyses. But the same metrics can also be used as *cognitive* models of how people process information about differences and similarities. This raises some fascinating questions: when does it make sense to use a particular metric to represent distances in psychological space? And what are the implications for the way in which individual and/or group differences are perceived and evaluated? Needless to say, the overarching assumption is that psychological perceptions of similarity *can* be adequately represented by geometric distances. While this is a common and generally well-supported assumption, geometric models of similarity are not without challenges and limitations, especially when dealing with abstract, non-perceptual representations (e.g., similarities between countries). For an overview of potential problems and alternative approaches, see Goldstone and Son (2013).

Euclidean Distance

The ordinary Euclidean distance D_2 is the length of the straight-line segment connecting two points:

$$D_2 = \left[\sum_{i=1}^{k} (x_{1i} - x_{2i})^2 \right]^{1/2} \qquad (1)$$

where x_{1i} and x_{2i} are the two scores or group means on the ith dimension (trait).

An important property of the Euclidean distance is that it is invariant under any rigid rotations of the axes. This means that the distance between any two points within a domain does not depend on the specific choice of the k dimensions used to describe that domain.

If the traits are measured on arbitrary scales, or there are other reasons to weigh all of them equally, D_2 can be computed from standardized scores:

$$D_2 = \left[\sum_{i=1}^{k} (z_{1i} - z_{2i})^2 \right]^{1/2} = \left[\sum_{i=1}^{k} \left(\frac{x_{1i} - x_{2i}}{S_i} \right)^2 \right]^{1/2} = (\mathbf{d}^{\mathrm{T}} \mathbf{d})^{1/2} \qquad (2)$$

where S_i is the standard deviation of the ith trait (pooled in the case of two groups), and $\mathbf{d}$ is a column vector of standardized univariate differences (i.e., values of Cohen's d in the case of two groups). Note that, as a general rule, the standardized D_2 lacks rotational invariance. This is because, after rotation, the rotated scores have to be re-standardized to their new standard deviations, which changes D_2 unless the traits are orthogonal with equal variance.

Cognitive Implications

The Euclidean distance D_2 is the default metric in many common models of psychological similarity, including those based on multidimensional scaling (MDS; see Arabie, 1991; Borg et al., 2018). From an information-processing perspective, D_2 is consistent with the

idea that the difference between two objects is evaluated holistically, so that multiple dimensions are integrated into a unitary representation. Dimensions that are processed in this way are usually described as *integral*; common examples are pitch and loudness for sounds and hue, brightness, and saturation for colors. In contrast, when differences on multiple dimensions are evaluated one by one and then combined to yield an overall judgment of similarity, the dimensions are called *separable* or *analyzable*. An example is that of geometric shapes that vary in their size and orientation (Attneave, 1950; Borg et al., 2018; Garner, 1974; Shepard, 1987; more on this below). According to some authors, genuine perceptions of similarity are always based on integral processing, whereas tasks that involve classification and judgment encourage analytic processing (Garner, 1974). Indeed, changing the task instructions can promote analytic processing even in domains that are spontaneously perceived as integral, such as colors or sounds (Kemler Nelson, 1993). In the field of face perception, the evidence indicates that the geometry of face space is approximately Euclidean (Meytlis, 2011; Tredoux, 2002; Wilson et al., 2002).[2]

The rotational invariance of D_2 implies that the specific dimensions chosen to describe the domain may be arbitrary rather than essential; one could rotate them at will, and the perceived distance between any two objects would remain the same. This lack of privileged axes is one of the defining features of integral processing. However, even domains that are spontaneously processed in a holistic fashion—and thus are well described by the Euclidean metric—may be represented more naturally along certain particular combinations of dimensions, which then serve as "weakly privileged" axes for analysis (Kemler Nelson, 1993). For example, pitch and loudness seem to represent true psychological dimensions in the auditory domain: when two sounds differ in pitch or loudness, they are easier to discriminate and classify than sounds that differ along a rotated dimension that combines both features. More generally, integral dimensions imply a Euclidean psychological space, but the converse is not necessarily true: certain cognitive processes may employ Euclidean representations even if the relevant dimensions can be analyzed separately in other contexts.

I am not aware of any research that directly investigated the geometry of psychological spaces for personality, cognitive ability, values, and similar domains. A series of studies on preferences for social and sexual partners has shown that desirable traits across domains (e.g., kindness, ambition, intelligence, physical attractiveness) are integrated in a way that is well approximated by a Euclidean algorithm—so that, for example, D_2 can be used to accurately quantify the perceived distance between any individual candidate and one's ideal partner (Conroy-Beam & Buss, 2017; Conroy-Beam et al., 2019a; Krems & Conroy-Beam, 2020). However, it is possible that the computational logic of partner choice is somewhat unique—after all, the goal is not simply to describe or classify people, but to *maximize* the desirability of one's friends and mates (see Conroy-Beam & Buss, 2017; Krems & Conroy-Beam, 2020; more on this below). As noted earlier, Euclidean representations do not imply that the relevant dimensions are always or necessarily processed in an integral fashion. Desirable traits such as intelligence, kindness, and physical attractiveness are clearly not arbitrary and can be analyzed separately with relative ease. A possible reason why they are integrated with a Euclidean algorithm is that,

---

[2] Even if differences between faces are usually perceived in a way that is well described by D_2, it is still possible to induce violations of Euclidean geometry by using strategically constructed stimuli (e.g., ambiguous faces; Laub et al., 2007).

compared with other metrics such as the city-block distance (see below), D_2 is disproportionately influenced by the largest discrepancies. As a result, the algorithm tends to discard potential partners who are severely lacking in one or a few key areas, even if they rate highly in the remaining ones (Conroy-Beam & Buss, 2017).

Another hint that domains of psychological variation may be represented in Euclidean form comes from the very existence of multiple models of personality and cognitive ability, derived from alternative rotations of a similar number of factors. Examples in the personality domain include the "Alternative Five" by Zuckerman et al. (1993), the five global factors of the 16PF (Cattell & Schuerger, 2003), and the six factors of the HEXACO (Lee & Ashton, 2004). The fact that experts find it hard to agree on the "true" basic dimensions of personality suggests that this domain may lack privileged or essential axes, and thus exhibit rotational invariance. At the same time, two key dimensions of social perception—usually labeled *dominance* and *nurturance,* or *agency* and *communion*—tend to emerge rather consistently across domains (e.g., first impressions from faces, judgments of familiar others, group stereotypes; see Stolier et al., 2020) and define a two-dimensional space called the "interpersonal circumplex". The same two dimensions can be recovered from a rotation of the Big Five traits Extraversion and Agreeableness (see et al., 2013), and may capture the same information with more psychological immediacy. Similar to colors and sounds, perceptions of people may be processed holistically as a default, but also possess some weakly privileged dimensions of analysis that facilitate comparison and classification.

Mahalanobis Distance

The Mahalanobis distance D_M is a generalization of the standardized Euclidean distance that takes correlations among variables into account (De Maesschalck et al., 2000; Huberty, 2005). Specifically, D_M corresponds to the length of the straight-line segment between two points, divided by the value of the standard deviation along the direction of that segment. The formula for D_M is:

$$D_M = [(x_1 - x_2)^T S^{-1} (x_1 - x_2)]^{1/2} = (d^T R^{-1} d)^{1/2} \tag{3}$$

where x_1 and x_2 are column vectors of scores or means, $\mathbf{d}$ is a column vector of standardized univariate differences, and $\mathbf{S}$ and $\mathbf{R}$ are the covariance and correlation matrices, respectively (pooled in the case of two groups). The difference between D_M and D_2 is illustrated in Figure 1.

It is easy to verify from Eq. 6 that, if the traits are all orthogonal, the correlation matrix reduces to the identity matrix and D_M reduces to the standardized Euclidean distance (Eq. 2). Importantly, the Mahalanobis D_M is invariant to axis rotation, just like the unstandardized D_2.
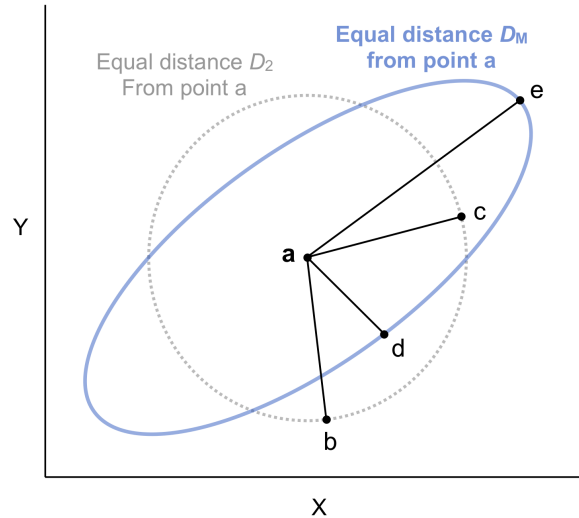
Figure 1. Illustration of the difference between the Euclidean distance D_2 and the Mahalanobis distance D_M. In the figure, X and Y are positively correlated. In the Euclidean plane, the circle (defined as the set of points at the same distance from a given point) has the familiar round shape. The Mahalanobis circle looks like an ellipse whose orientation depends on the correlation structure. Points *b* and *c* are equally distant from *a* according to D_2, but *c* is closer to *a* according to D_M. Conversely, points *d* and *e* are equally distant from *a* according to D_M, but *d* is closer to *a* according to D_2.

When it is used to compare the average profiles of two groups, D_M is the multivariate equivalent of Cohen's *d*, and has the exact same interpretation in terms of distribution overlap, classification accuracy, and so forth (assuming multivariate normality and equality of covariance matrices between groups). For example, consider two univariate normal distributions with $d = 0.50$ and two multivariate normal distributions with $D_M = 0.50$. In both cases, the overlap between distributions is 80%; this implies an expected classification accuracy of 60% with linear discriminant analysis (LDA), which approximates the optimal classifier under multivariate normality (see Del Giudice, 2022; James et al., 2013).[3] The formula for the overlapping coefficient *OVL* is:

$$OVL = 2\Phi(-D_M/2) \tag{4}$$

---

[3] The Mahalanobis distance D_M between two groups can be described equivalently as (a) the straight-line distance between the centroids, standardized by the value of the standard deviation in the direction of the line that connects the centroids; or (b) the standardized univariate difference between the two distributions, after they have been orthogonally projected on the discriminant axis (e.g., Thomas, 2003). This can be a source of confusion, because the line that connects the centroids does not coincide with the discriminant axis (except in special cases). The key is that description (a) refers to the standard deviation of the multivariate distribution *in a particular direction* (i.e., that of the line connecting the centroids), whereas description (b) refers to the standard deviation of a univariate distribution, obtained by orthogonally projecting the multivariate distribution onto the discriminant axis. If one orthogonally projects the multivariate distributions onto the line that connects the centroids, the distance between the resulting distributions is *not* equivalent to D_M (see Thomas, 1999, 2003). However, if one projects the multivariate distributions onto the line connecting the centroids, not orthogonally but *in the direction of the classification boundary*, the resulting univariate distributions are equivalent to those on the discriminant axis, and their standardized distance is D_M.

where $\Phi(\cdot)$ is the standardized normal cumulative distribution function (CDF). *OVL* is the proportion of each distribution that overlaps with the other distribution, and ranges from 0 to 1. Similarly, the expected classification accuracy or *probability of correct classification* (*PCC*) for equal-sized groups is given by:

$$PCC = \Phi(D_M/2). \tag{5}$$

If sample size is small relative to the number of traits (a simple rule of thumb is less than 100 cases per trait), sampling error can inflate the estimated size of D_M to a substantial degree. The upward bias in D_M can be corrected with this formula:

$$D_{\text{Mu}} = \left[max \left(0, \frac{N_1+N_2-k-3}{N_1+N_2-2} D_M^2 - k \frac{N_1+N_2}{N_1 N_2} \right) \right]^{1/2} \tag{6}$$

where N_1 and N_2 are the sample sizes for the two groups and k is the number of traits (see Del Giudice, 2022).

Cognitive Implications

Because D_M generalizes the Euclidean distance and shares the same basic properties (including rotational invariance), most of what I wrote about the cognitive implications of D_2 also applies to D_M. Like D_2, the D_M distance implies integral processing, but with a twist: multiple dimensions are integrated into a unitary representation *while taking their correlational structure into account* (Figure 1). Experimental studies show that participants are good at detecting correlations between multiple features of the stimuli, whether the latter are abstract like geometric shapes or ecologically relevant like faces (e.g., Ashby & Perrin, 1988; Jones & Goldstone, 2013). A recent study by Stolier et al. (2020) showed that people possess detailed, realistic knowledge about the way in which different personality traits correlate with one another. In turn, this knowledge guides inferences and perceptions across multiple domains, from group stereotypes to impressions from face pictures (Stolier et al., 2020).

Clearly, people are sensitive to correlational patterns and are able to integrate them in their judgments and decisions. The question is whether people *do* use information about correlations, particularly when dealing with differences and similarities between individuals. While this may seem an obvious strategy in presence of correlated traits, it can have some unexpected and potentially damaging consequences. In particular, D_M as a measure of dissimilarity leads to violations of the *dominance axiom*, which holds that a pair of objects that differ on two dimensions must be more dissimilar than the corresponding pair of objects that differ on only one of the dimensions (see Perrin & Ashby, 1991). Imagine a pair of almost identical faces *a* and *b* that differ only because *b* has a larger nose than *a*. Now imagine another face *c*, which has a larger nose *and* larger ears. The Euclidean distance would indicate that *a* and *c* are more dissimilar than *a* and *b*. But if the size of the nose and that of the ear are positively correlated, the D_M distance may judge *c* as more *similar* to *a* than *b*, thus violating the dominance axiom (see Figure 2).
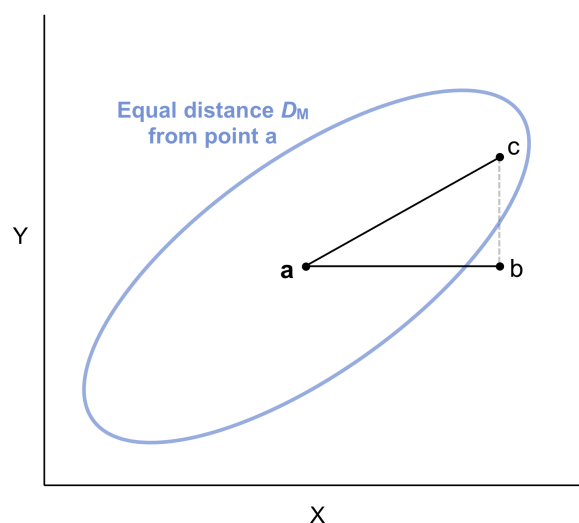
Figure 2. The Mahalanobis distance D_M violates the dominance axiom. Points *a* and *b* differ only on dimension X, while points *a* and *c* differ both on X (by the same amount) and on Y. To illustrate, *a* and *b* could be two faces that differ only in the size of the nose, while *a* and *c* differ in the size of the nose *and* that of the ears. The Euclidean distance would judge *a* and *b* as more similar than *a* and *c;* but since X and Y are correlated (e.g., larger noses tend to be associated with larger ears), *a* and *c* are more similar according to D_M.

To illustrate how dominance violations can have perverse effects under certain conditions, consider the integration of preferences in partner choice. In this context, using the Mahalanobis distance would be almost certainly suboptimal: since desirable traits tend to correlate positively with one another (Conroy-Beam et al., 2019b), a lower-value partner who rated poorly on two traits could be perceived as closer to the ideal—and hence more desirable—than a higher-value partner who rated poorly on only one trait. Thus, a process of preference integration based on D_M would systematically fail to maximize partner value, in contrast with the Euclidean criterion that people seem to follow (Conroy-Beam & Buss, 2017). I am not aware of any studies comparing D_2 and D_M as similarity criteria in psychological domains such as personality and cognitive ability. One way to do so would be to look for evidence of dominance violations, which are not expected if the relevant metric is Euclidean (see Perrin & Ashby, 1991).

The caveats and complications of using D_M as a measure of dissimilarity between *individuals* cease to apply when the task is to compare *groups*, or classify individuals as members of alternative groups. In group comparison and classification, the correlational structure of the traits provides critical information, and D_M is clearly preferable to Euclidean distances that ignore correlations. If the assumptions of multivariate normality and equal covariance matrices hold, D_M can be used to calculate the proportion of overlap between two distributions, in addition to several other indices of group difference (Del Giudice, 2022). Under the same assumptions, D_M is the optimal criterion for classifying individuals into groups, and yields the same expected accuracy of LDA (see Ashby & Perrin, 1988; Del Giudice, 2022; Thomas, 1999, 2003). In sum, it is plausible to expect that people will take correlations into

account when dealing with group differences and classification; however, I do not know of any studies testing this prediction with respect to group differences in psychological traits.
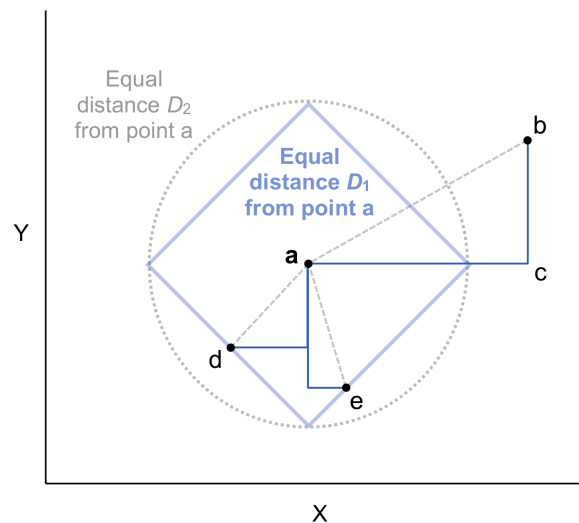
City-Block Distance

The city-block distance D_1 between two points is the sum of their absolute distances on each dimension:

$$D_1 = \sum_{i=1}^{k} |x_{1i} - x_{2i}| \tag{7}$$

or, in standardized form:

$$D_1 = \sum_{i=1}^{k} |z_{1i} - z_{2i}| = \sum_{i=1}^{k} \left| \frac{x_{1i} - x_{2i}}{s_i} \right|. \tag{8}$$

The name derives from the fact that, as it happens in cities with a North-South and West-East street grid, the shortest path between two points is not the diagonal but rather a sum of perpendicular segments (D_1 is also called the *taxicab* or *Manhattan* distance). The difference between D_1 and D_2 is illustrated in Figure 3. One notable point of comparison is that D_1 is not invariant to axis rotation, so that changing the orientation of the axes will also alter the distances between points.

Figure 3. Illustration of the difference between the Euclidean distance D_2 and the city-block distance D_1. The distance D_1 between *a* and *b* is not the straight line from *a* to *b* (dashed line), but the sum of the distance on dimension X and that on dimension Y (i.e., *a* to *c* and *c* to *b*). The city-block circle has the shape of a diamond; for example, *d* and *e* are equally distant from *a* according to D_1.

Both the Euclidean and city-block distances are special cases of the generalized *Minkowski distance:*

$$D_p = \left[\sum_{i=1}^{k} |x_{1i} - x_{2i}|^p\right]^{1/p}. \tag{9}$$

The exponent p determines the order of the distance, and can range from 1 to infinity (the Minkowski distance is no longer a metric when $p < 1$). With $p = 1$ one obtains the city-block distance D_1, whereas $p = 2$ returns the Euclidean distance D_2. Note that D_2 is the only Minkowski metric that exhibits rotational invariance. Also, as p increases above 1, the distance D_p becomes increasingly dominated by the traits showing the larger absolute differences (see Borg et al., 2018).

Cognitive Implications

Modeling psychological similarity with the city-block distance implies that people form their judgment by first assessing the distance of two objects on each of the k dimensions of the domain, and then adding them to yield an overall rating of dissimilarity. Hence, D_1 is the natural metric for fully analyzable dimensions, such as the height and width of rectangles or the size and orientation of geometric shapes (Borg et al., 2018; Garner, 1974; Shepard, 1987). When judgments reflect a mixture of integral and analytic processing, the most appropriate metric may be neither D_1 nor D_2, but a Minkowski distance with exponent p between 1 and 2 (for example $D_{1.5}$ as in Gronau & Lee, 2020; see Shepard, 1991).

Some researchers have argued that the city-block distance is often more cognitively plausible than its Euclidean counterpart, a notion that is not without empirical support (see Arabie, 1991; Kemler Nelson, 1993). For example, the studies of mate choice I discussed earlier generally favor a Euclidean model of preference integration, but also suggest that D_1 may better predict attraction to short-term sexual partners under some conditions (Conroy-Beam & Buss, 2017). Unfortunately, determining whether D_1 provides the best fit to similarity data is often technically challenging, and can be an intractable problem for certain statistical methods (Arabie, 1991; Gronau & Lee, 2020). Moreover, small deviations from a Euclidean space toward a city-block metric can be easily drowned out by measurement noise (Carroll & Wish, 1974).

Shape Distance

All the distance metrics reviewed until now seek to quantify the *overall* dissimilarity between two multivariate profiles. However, in some cases it can be useful to focus on the specific features that make two profiles more or less similar. To this end, overall similarity can be partitioned into three components: *elevation*, *scatter*, and *shape* (Cronbach & Gleser, 1953; Furr, 2010; Skinner, 1978). Elevation can be operationalized as the mean score across all the traits; scatter as the variance of scores around the mean; and shape as the pattern of "peaks and valleys" in the profile (Figure 4). When plotted in parallel coordinates as in Figure 4A, two profiles with the same shape and scatter but different elevations will appear as parallel lines; whereas two profiles with the same shape but different levels of scatter will look like the "amplified" and "muted" versions of each other. In general, shape captures the relations among traits *within* the profile. For example, people who score higher in verbal tasks compared with

visuospatial tasks may share some common psychological features, regardless of their overall intelligence level (that is, their profile elevation).
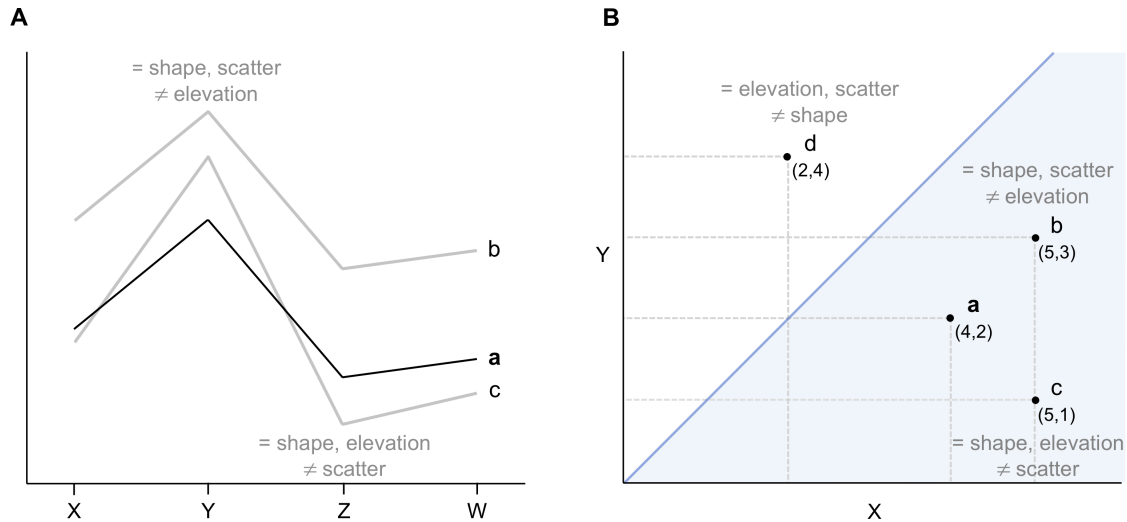
Figure 4. The (dis)similarity between two multivariate profiles can be decomposed into elevation, scatter, and shape. Panel (A) shows three four-dimensional profiles, plotted in parallel coordinates. Profile *b* has the same scatter and shape as *a*, but higher elevation; profile *c* has the same elevation and shape as *a*, but larger scatter. Panel (B) shows four two-dimensional profiles as points on a plane: *b* differs from *a* in elevation, *c* in scatter, and *d* in shape. Note that, in a two-dimensional space, all the points on the same side of the Y = X line have the same shape. Hence, all the points in the shaded area have the same shape as *a*, though they may differ in scatter and/or elevation.

A simple and straightforward way to quantify shape similarity is to calculate the Pearson correlation between two profiles (r_{12}). Because the correlation is unaffected by elevation and scatter, it is a "pure" index of shape similarity, and performs well even in comparison with more complex indices (Furr, 2010; McCrae, 2008). The value of r_{12} is +1 for maximum similarity, and reaches −1 (maximum dissimilarity) when the two profiles have opposite shapes (high when the other is low, and vice versa). The quantity $(1 − r_{12})$ is known as *correlation distance* and ranges from 0 to 2.

Although the correlation distance is not a proper metric, it can be turned into one with a simple transformation, yielding

$$D_S = [2(1 − r_{12})]^{1/2}. \tag{10}$$

The distance metric in Eq. 10 is used quite often in the literature (for example in applications of multidimensional scaling and clustering; see e.g., Borg, 2018; Revelle, 2018), but does not have a standard label.[4] For convenience, in this paper refer to it as *shape distance* or D_S. Like the distance correlation, D_S ranges from 0 (maximum similarity) to 2 (maximum

---

[4] In their classic paper, Cronbach & Gleser (1953) used D'' to denote the shape distance and distinguish it from the alternative metric D' (which combines information about shape and scatter). However, this notation is opaque and has not been followed in the more recent literature.

dissimilarity). D_S is equivalent to the Euclidean distance between two profiles, after each profile has been standardized with respect to its own mean and standard deviation (Cronbach & Gleser, 1953). Because of the square root transformation, D_S is especially sensitive to deviations from perfect similarity. To illustrate, the distance is $D_S = 0.45$ when $r_{12} = .90$, reaches the midpoint of the scale ($D_S = 1.00$) when $r_{12} = .50$, and goes up to $D_S = \sqrt{2} \approx 1.41$ when $r_{12} = 0$.

Cognitive Implications

While D_S and other correlation-based indices are commonly used in research on profile similarity (e.g., Furr, 2010; McCrae, 2008), their cognitive implications remain obscure. I surmise that focusing on shape information to the exclusion of elevation and scatter does not work as a general-purpose strategy, but can become extremely useful in the context of certain well-defined problems. For example, as people get older, their facial features tend to become more pronounced (or "caricatured"). In the geometry of face space, this corresponds to a movement away from the centroid—that is, a joint change in the elevation and scatter of the profile (see Deffenbacher et al., 1998). By filtering out these effects, the shape distance would make it easier to recognize the same person, despite the physical changes of aging. Hence, indices of shape (dis)similarity such as D_S may play a useful role in tracking the identity of specific individuals over time, even if the overall geometry of face space is best described as Euclidean (Meytlis, 2011; Wilson et al., 2002).

Are there aspects of social perception that would similarly benefit from a narrow focus on profile shape? One can certainly speculate. Consider the pervasive dimensions of dominance and nurturance that I introduced earlier, and assume that they can be treated as non-arbitrary psychological axes. The shape distance D_S cleaves this interpersonal space into two regions—one in which dominance exceeds nurturance (marked by descriptors such as "dominant", "arrogant", and "cold-hearted"), and one in which nurturance prevails ("warm", "unassuming", "submissive"; see DeYoung et al., 2013). This distinction is intuitively salient, and is naturally captured by indices of profile shape such as D_S.

Distances in Multivariate Domains

When the number of traits in a domain increases, the space of individual differences expands, and distances between profiles grow in predictable ways. In this section I examine how various distance metrics behave in high dimensions. I begin with patterns of individual differences (i.e., distances within a single distribution), then move on to consider group differences (i.e., distances between two distributions). The simulations I present are based on the standardized multivariate normal distribution (which approximates the distribution of many psychometric traits), but the same qualitative patterns apply more broadly, as long as distributions are reasonably symmetric without fat tails. The R code of the simulations is available at https://doi.org/10.6084/m9.figshare.13070576.

Patterns of Individual Differences

Orthogonal Traits

To begin, consider the case in which the k traits of interest are all orthogonal (Figures 5a-5d). With Euclidean and Mahalanobis metrics, the average distances between pairs of individuals, and between individuals and the centroid, increase as $\sqrt{k}$ (Figures 5A and 5B; Altman & Krzywinski, 2018; Giraud, 2015). Average city-block distances scale up as k, which is considerably faster (Figure 5C); at the same time, the distance concentration effect (relative to the average) occurs at a somewhat slower pace (Aggarwal et al. 2001), though in this particular case the difference is negligible (Figure 6A). As dimensionality increases, correlations among pairs of profiles tend to cluster more tightly around zero; as a result, the distribution of pairwise shape distances becomes more narrowly concentrated around $D_S = \sqrt{2} \approx 1.41$ (Figure 5D; Altman & Krzywinski, 2018). In contrast, the distribution of D_S between individuals and the centroid does not follow a predictable pattern, as it depends on the particular coordinates of the centroid (for this reason, it is not shown in the figure). Note that this distance can only be calculated from unstandardized scores, and is only interpretable when those scores can be meaningfully compared across traits (i.e., not measured on arbitrary scales).

Correlated Traits

When traits are not orthogonal but correlated, they become partly redundant, so that the domain contains less independent variation than implied by the number of observed traits (k). The *effective dimensionality* (ED) of a distribution or dataset is the equivalent number of orthogonal dimensions (with equal variance) that would produce the same overall pattern of covariation (Del Giudice, 2020). If the traits are perfectly correlated with one another, they can be represented by just one dimension of variation and their ED is 1; if the traits are orthogonal with equal variance, the ED equals k.
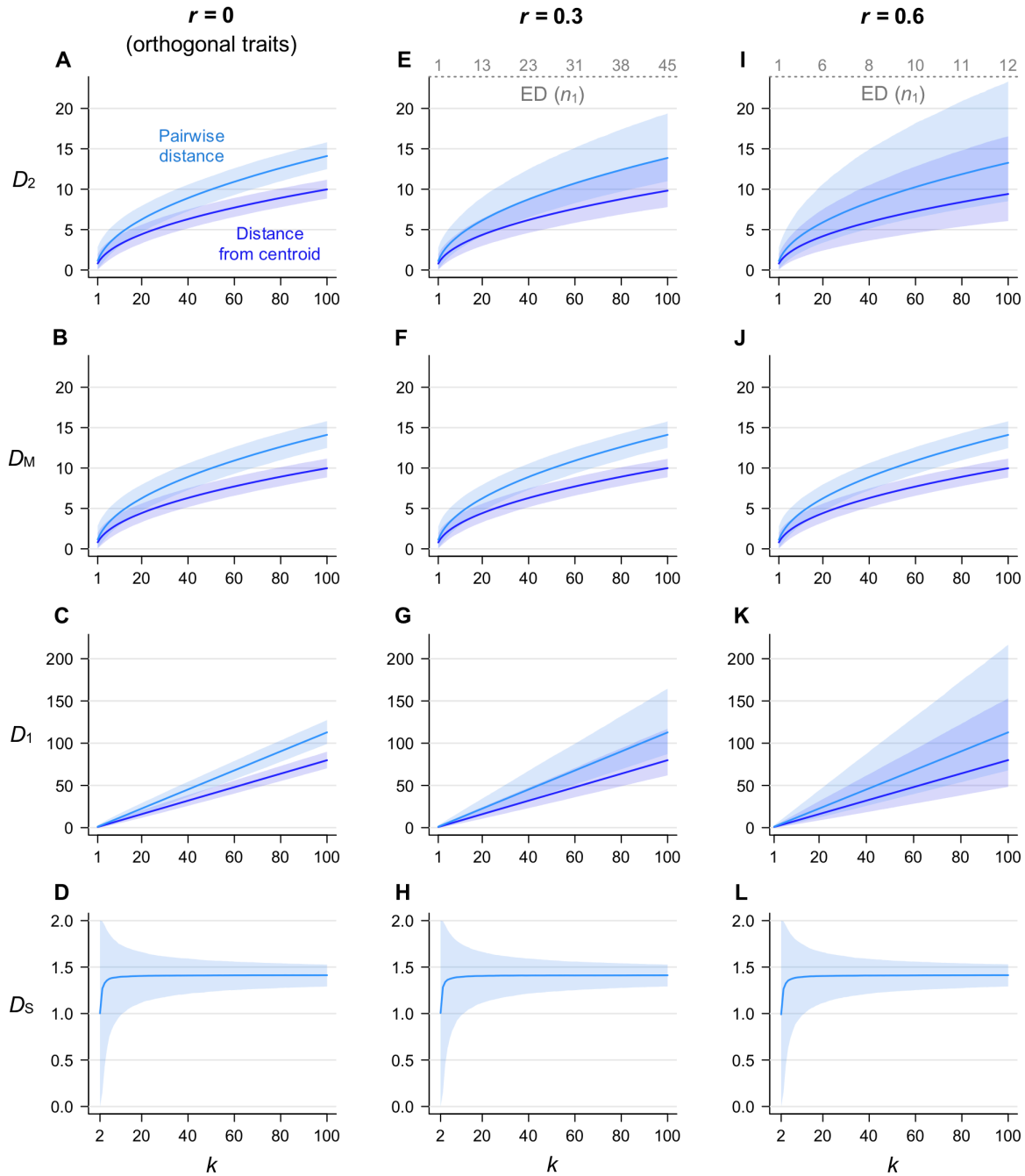
Figure 5. Simulated distances between pairs of points and from the centroid, measured with four alternative metrics. Plots are based on a standardized multivariate normal distribution (10,000 points) with $k = 1$ to 100 traits. Solid lines are average distances; shaded areas represent 90% of the distance distribution (5th to 95th percentile). In the left column, all traits are orthogonal. In the center and right columns, traits are uniformly correlated at $r = .30$ and $.60$; the top horizontal axis show the approximate effective dimensionality (ED) of the distribution, calculated with index n_1. D_2 = standardized Euclidean distance; D_M = Mahalanobis distance; D_1 = standardized city-block distance; D_S = shape distance (from standardized scores).
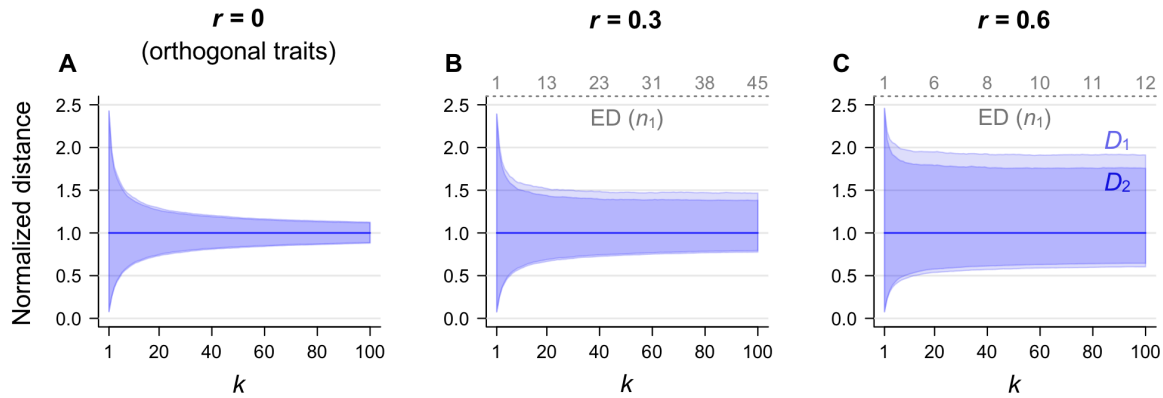
Figure 6. Distance concentration with D_1 and D_2. Plots are based on a multivariate normal distribution (10,000 simulated points) with $k = 1$ to 100 traits. Shaded areas show the 5th and 95th percentiles of the distance distribution (from the centroid), normalized by the average distance. The light shaded area represents D_1; the dark shaded area represents D_2. In panel (A), all traits are orthogonal. In panels (B) and (C), traits are uniformly correlated at $r = .30$ and .60; the top horizontal axis show the approximate effective dimensionality (ED) of the distribution, calculated with index n_1. D_1 = standardized city-block distance; D_2 = standardized Euclidean distance.

It is usually the case that measurement error becomes more severe when traits become narrower (e.g., because traits are measured with fewer items); if uncorrected, measurement error deflates the apparent size of correlations and increases the ED of the dataset (more on this below). In this paper, I quantify the ED with the n_1 index discussed in Del Giudice (2020).[5] The same paper reviews the practical issues involved in the estimation of ED, including the impact of small sample size, deviations from normality, and other factors.

Human faces offer an excellent illustration of the meaning of effective dimensionality. With principal component analysis (PCA), pictures of faces can be decomposed into orthogonal components (called *eigenfaces*) that encode different aspects of anatomy and appearance (see Meytlis & Sirovich, 2007; Sirovich & Meytlis, 2009a; Valentine et al., 2016). Eigenfaces can then be reassembled to yield a reconstructed version of the original pictures; using more components increases the dimensionality of the synthetic face space and yields more detailed reconstructions. The evidence indicates that about 70-100 components are sufficient to permit accurate face recognition (Burton et al., 2001; Meytlis & Sirovich, 2007; Sirovich & Meytlis, 2009; see Section S1 of the Supplement).

However, facial features do not vary independently but in a correlated fashion, as reflected in the unequal variance accounted for by the PCA components. This means that the ED of facial identity must be lower than 70-100 dimensions, perhaps substantially so. To get an initial estimate of the ED of this domain, I reanalyzed published data from Meytlis & Sirovich (2007) and Sirovich & Meytlis (2009b), which yielded a range of values from $n_1 = 27.8$ to 43.4. (Note that these values are rough approximations; see Section S1 of the Supplement for details.)

---

[5] The n_1 index is based on the Shannon entropy of the normalized eigenvalues of the correlation or covariance matrix. Other choices for entropy yield different indices of ED, but n_1 can be recommended as a balanced, general-purpose estimator (see Del Giudice, 2020).

These figures suggest that the facial traits involved in identity recognition span about 30-40 effective dimensions, which is still a vast space for individual variation. From another perspective, these ED values are consistent with average trait correlations around .30-.35. In an intriguing study, Sheehan and Nachman (2014) found that facial traits are markedly less intercorrelated than body traits (less than .20 on average,[6] compared with about .50 for body traits). Based on morphological and genetic data, the authors suggested that selection has acted to increase the variety and distinctiveness of human faces, as an adaptation to facilitate individual recognition (Sheehan & Nachman, 2014).

As correlations among traits grow stronger (and the ED decreases), within-group Euclidean distances grow somewhat less steeply with the number of dimensions, and concentrate at a slower pace than in the orthogonal scenario (Figures 5E and 5I; see also Del Giudice, 2020; Durrant & Kabán, 2009). City-block distances also concentrate more slowly, although their average values increase at the same rate (Figures 5G and 5K). As can be seen in Figure 6, the difference in the distance concentration patterns of D_1 and D_2 becomes more clear-cut when traits are correlated (Figures 6B and 6C), compared with the orthogonal scenario (Figure 6A). In contrast with D_1 and D_2, the Mahalanobis distance D_M automatically "corrects" for the correlational structure of the domain; accordingly, its within-group distribution is not affected by changes in trait correlations. Whereas Euclidean distances grow and concentrate more slowly when traits are more strongly correlated, the distribution of D_M between individuals depends only on the number of traits in the domain (k) and not on their ED (Figures 5F and 5J). Similarly, the distribution of pairwise D_S is insensitive to changes in correlations among traits, as illustrated in Figures 5H and 5L.

Patterns of Group Differences

Orthogonal Traits

The first and most obvious measure of the difference between two groups is the distance between their respective centroids. Figure 7 illustrates the case of orthogonal traits, assuming that groups differ by the same amount on each individual trait (0.5 standard deviations in this example). Because traits are orthogonal, D_M equals the standardized D_2 (Figure 7A); and because univariate distances are constant across traits, these distances grow as $\sqrt{k}$, whereas D_1 scales up as k (Figure 7B). The Mahalanobis D_M between two groups is the multivariate equivalent of Cohen's d; as such, it can be used as an effect size, or converted into other effect sizes such as the overlapping coefficient (see Figure 7A). Since the shape distance between the centroids is insensitive to differences in elevation between groups, it does not follow any predictable pattern; whether D_S increases, decreases, or remains constant as more traits are added depends entirely on the direction and (relative) size of each univariate difference. Again, the D_S between centroids can only be calculated from mean unstandardized scores, and its interpretability depends on whether those scores can be meaningfully compared across traits.

---

[6] Note that Sheehan & Nachman (2014) measured only 18 facial traits (such as nose width and length), way too few to capture individual appearance at any level of detail. Had they measured more fine-grained traits, the additional measurements would have become increasingly redundant, and the average correlation would have increased accordingly.
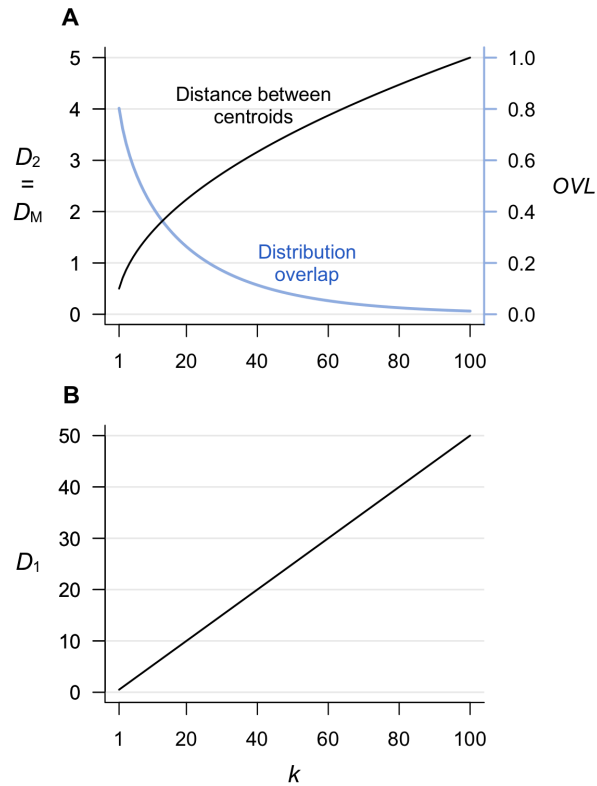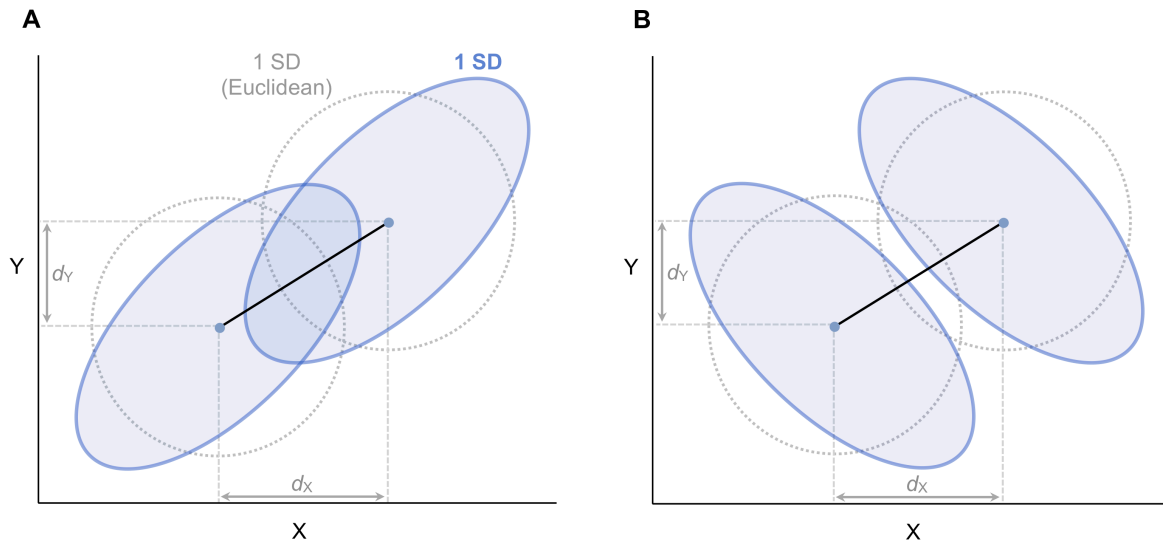
Figure 7. Distance between the centroids of two groups, measured with three alternative metrics. Plots are based on two multivariate normal distributions with $k = 1$ to 100 orthogonal traits, and a univariate difference of 0.5 standard deviations on each trait. Panel (A) also shows the proportion of overlap between distributions, measured with the overlapping coefficient *OVL* (right vertical axis). D_2 = standardized Euclidean distance; D_M = Mahalanobis distance; D_1 = standardized city-block distance.

Correlated Traits

When traits are correlated, D_M diverges from the standardized D_2 under most scenarios, and can become either larger or smaller than its Euclidean counterpart. Whether D_M is larger or smaller than D_2 depends on the direction along which the centroids are separated, relative to the axes of variation in the data. Figure 8 illustrates his concept. Both the univariate differences and the standardized D_2 are exactly the same in Figures 8A and 8B; the only difference is the correlation between the two traits. In Figure 8A, the connecting line is oriented along a major axis of variation, where the multivariate standard deviation is large. (Stated differently, the difference between the two groups goes "with the grain" of the correlational structure.) As a result, D_M is smaller than D_2. In Figure 8B, the connecting line is oriented along a minor axis of variation, where the standard deviation is small (i.e., the difference between the two groups goes "against the grain" of the correlational structure). The statistical separation between the distributions is sharper than in Figure 8A, and D_M becomes larger than D_2.

Figure 8. The Mahalanobis distance D_M between two groups can be larger or smaller than the standardized Euclidean distance D_2, depending on trait correlations. Dotted circles show the points at 1 standard deviation (SD) from the centroids, according to D_2. Solid ellipses show the points at 1 standard SD from the centroids according to D_M. In panel (A), the line connecting the centroids is oriented along a major axis of variation, and D_M is smaller than D_2. In panel (B), the line connecting the centroids is oriented along a minor axis of variation, and D_M is larger than D_2, reflecting the sharper statistical separation between the two distributions. Note that the univariate standardized distances on the two traits (d_X and d_Y) and the standardized D_2 are the same in the two panels.

A crucial implication of this principle is that increasing the number of traits in a domain will not necessarily lead to an increase in D_M. As narrower traits become more redundant, each of them adds less unique information about the difference between the groups, and contributes less to the overall size of D_M. In the limit, adding a trait that is just a linear combination of other traits (and hence completely redundant) has no effect whatsoever on D_M. Of course, if sample size is small relative to the number of traits, the estimated D_M may still be inflated by sampling error; the resulting bias can be corrected with the formula in Eq. 6.

In the Introduction, I cited a study by Hennessy et al. (2005), which found an overall effect size of $D_M = 3.20$ for sex differences in facial anatomy (bias-corrected estimate: $D_M = 2.84$).[7] In a fascinating analysis, the authors calculated the effect size at varying levels of anatomical resolution, from a minimum of 24 three-dimensional landmarks per face (e.g., the tip of the nose) to a maximum of 5,453 landmarks. The size of sex differences was $D_M \approx 2.20$ with 24 landmarks, went up to about 3.00 when using 100 landmarks, and reached a plateau of 3.20 with 200 landmarks. Increasing the anatomical resolution behind that point did not change the size of sex differences. Because D_M was always computed from a reduced version of the data (the first 18 PCA components) and not from the original coordinates, this procedure is not fully equivalent to increasing the number of traits k; still, the analysis nicely illustrates the "diminishing returns" of mapping a domain at an increasingly fine scale.

---

[7] This and other effect sizes from the Hennessy et al. (2005) study were digitally measured from Figure 3 of the paper. Note that the authors reported the squared D_M^2 instead of D_M.

Other Measures of Group Differences

The distance between average profiles is an important index of group differences, but the picture it provides is only partial. For a more complete perspective, one can also consider the average distance between the individuals in a group and the centroid of the *other* group. A related measure is the average distance between two individuals belonging to different groups. Figure 9 illustrates these distances in the same scenario of Figure 7 (orthogonal traits with univariate $d = 0.5$ on each trait), and compares them with the corresponding distances within the same group. In the simple scenario of Figure 9, between-group and within-group distances (D_1, D_2, and D_M) increase with constant ratios of about 1.06 (for pairs of individuals) and 1.12 (for individuals vs. centroids). These ratios are solely determined by the size of the univariate differences on each trait, which in this case are fixed to $d = 0.5$. If the shape distance D_S is computed from unstandardized scores, it does not follow a predictable pattern, and may suffer from interpretive issues if scores are not comparable across traits (see above). While it is possible to calculate the between-group pairwise D_S from standardized scores (Figure 9C), this measure converges to the same value as its within-group counterpart, making the comparison uninformative.

The pattern of between- and within-group distances illustrated in Figure 9 has some surprising implications. In this particular example, when the domain is measured with 100 traits the Mahalanobis distance is $D_M = 5.00$. The centroids are five standard deviations apart, corresponding to a distribution overlap of about 1.2% and an expected classification accuracy of more than 99%. But even if the two groups are sharply separated in multivariate space, a ratio of 1.06 means that the average distance between two individuals belonging to different groups (e.g., a random male and a random female) is only 6% larger than the average distance between two individuals in the same group (e.g., two random males).[8] Moreover, the average Mahalanobis distance between individuals is close to *fifteen* standard deviations. The bottom line is that, in high-dimensional domains, even large differences between groups will often coexist with much larger differences among individuals. However, within-group differences extend amorphously in all directions, whereas group differences exist on a specific axis; hence, the between-group signal can remain strong and clear on the background of massive individual variation. For example, human faces are strikingly variable and individually unique—and yet, the distance between the male and female distributions makes it easy to recognize people's sex from their faces with more than 95% accuracy (Bruce et al., 1993; Ng et al., 2015; O'Toole et al., 1998).

---

[8] Note that, when dimensionality becomes very high (e.g., thousands of dimensions), even a small between/within group ratio can yield clearly separated distributions of distances (see e.g., Murtagh, 2009). However, most psychological domains have a smaller number of traits, even when mapped at a fine scale.
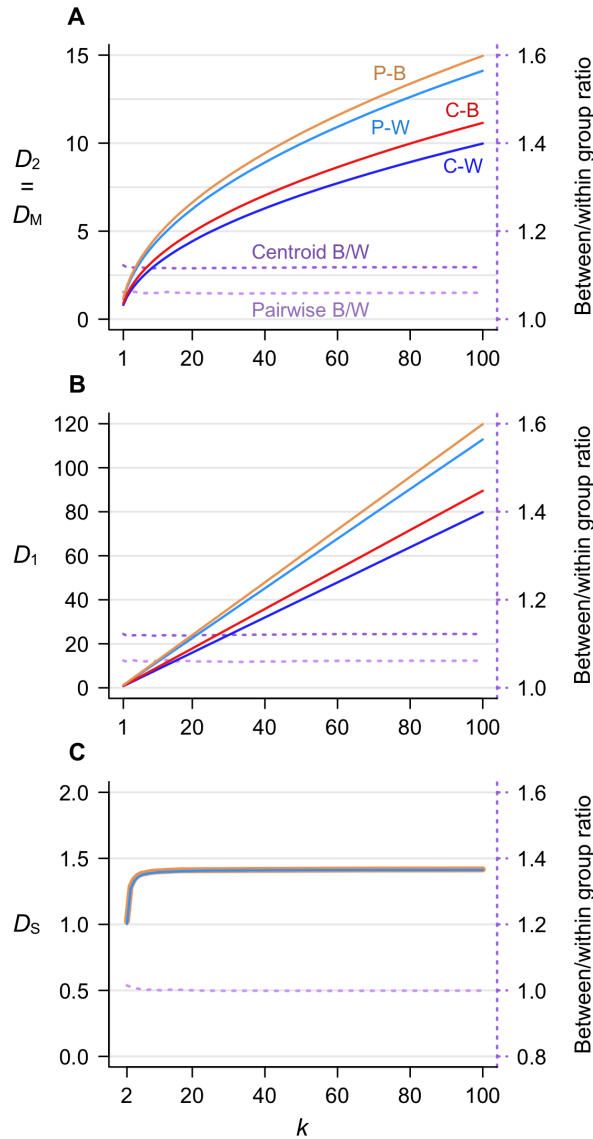
Figure 9. Simulated distances within and between two groups, measured with four alternative metrics. Plots are based on two standardized multivariate normal distributions (10,000 simulated points each) with $k = 1$ to 100 orthogonal traits, and a univariate difference of 0.5 standard deviations on each trait (same as in Figure 7). Solid lines are average distances; dotted lines show between/within group ratios (B/W) for the same distances (right vertical axis). P-B = pairwise distance between groups; P-W = pairwise distance within group; C-B = centroid distance between groups; C-W = centroid distance within group. D_2 = standardized Euclidean distance; D_M = Mahalanobis distance; D_1 = standardized city-block distance; D_S = shape distance (pairwise, from standardized scores).

The Effects of Measurement Error

Before ending this section, it is useful to briefly discuss how distances in multivariate domains are affected by measurement error. This is especially critical for psychological constructs, which are often measured with substantial uncertainty. In the language of classical

test theory, an individual's *observed score* on a trait is a sum of the underlying *true score* plus a certain amount of error. The *reliability* of a measure is the proportion of true score variance on the total variance; for instance, if a trait is measured with .80 reliability, it means that 80% of the variance in the scores is shared with the underlying construct, while the remaining 20% is accounted for by error (see Revelle, 2018). Here I make the usual simplifying assumptions that measurement error is unsystematic (i.e., its expected value is zero), does not depend on the value of the true score, and is uncorrelated across traits.[9]

To begin, consider the multivariate distribution of true scores for a single group. Adding measurement error inflates the variance of the scores without changing their means. And since error terms are uncorrelated across traits, trait correlations become attenuated, bringing the distribution closer to the orthogonal case and increasing the effective dimensionality of the data (see Del Giudice, 2020). As a result, both D_1 and D_2 concentrate faster with increasing k, unless the true scores are already orthogonal (Figures 5 and 6).

What happens to within-group distances (between pairs and from the centroid) partly depends on the shape of the distribution before and after the introduction of error. The simplest scenario is one in which true scores are multivariate normal *and* the errors are normally distributed, so that observed scores are also multivariate normal. The effects of error in this scenario are straightforward. Specifically, the average unstandardized D_1 and D_2 increase, owing to the variance inflation produced by measurement error. In contrast, the standardized D_1 remains unchanged, because standardization rescales all the variances to 1. The standardized D_2 is also unaffected if the traits are all orthogonal; but if traits are correlated, it increases slightly as correlations become attenuated (see Figures 5A, 5E, 5I). The within-group distributions of D_M and D_S are insensitive to changes in trait variances and correlations (Figure 5), and hence are not affected by the presence of measurement error.

Turning to group differences, the main index to consider is the distance between the centroids of two distributions. Now, all distances computed from *un*standardized scores remain unaffected: this includes the unstandardized D_1 and D_2, as well as the D_S between centroids. Because trait variance is inflated by measurement error, univariate standardized differences become attenuated, and the standardized D_1 and D_2 decrease accordingly. The story is a bit more complicated for D_M. On the one hand, the attenuation of standardized differences tends to decrease D_M just like D_1 and D_2. On the other hand, trait correlations also become attenuated, and this may either decrease or increase D_M, depending on whether the overall group difference goes "with" or "against the grain" of the correlational structure (Figure 8). Since these two effects can end up pulling in opposite directions, D_M may ultimately increase, decrease, or remain unchanged.

---

[9] If these assumptions are true in the population, they can be expected to hold to a reasonable approximation when sample size is sufficiently large. In small samples, however, fluctuations due to sampling may introduce systematic biases and sizable "nuisance correlations" among traits as a consequence of measurement error (see e.g., Stanley & Spence, 2014).

Correcting for Measurement Error

To get a more accurate picture of group differences at the level of the underlying constructs, researchers can employ two main error correction methods. The first and simpler approach is to obtain a data-based estimate of the reliability of the observed scores, and use it to disattenuate the univariate differences and correlations. To disattenuate the standardized difference between two groups on a given trait (Cohen's *d*), it is sufficient to divide it by the square root of the reliability of the observed score. Likewise, the correlations between two traits can be disattenuated by dividing it by the square root of the product of the two reliabilities. Naturally, a disattenuated correlation matrix yields lower ED estimates, reflecting the stronger degree of overlap among traits at the level of true scores (see Del Giudice, 2020).

The most common index of reliability in psychological research is Cronbach's *alpha* (α), which is based on the intercorrelations of the items that make up a scale (internal consistency). Despite its popularity, α suffers from important limitations. When applied to unidimensional scales, α tends to yield deflated estimates of reliability (Dunn et al., 2014; McNeish, 2018). More often, however, psychological scales are not fully unidimensional, and tap additional factors besides the trait they are supposed to measure. In these cases, α can be substantially inflated, and hence underestimate the amount of error in the data (Cortina, 1993; Crutzen & Peters, 2017). A better alternative to α is McDonald's *omega-hierarchical* (ω_h), which isolates the true score variance associated with the general factor of a scale (Dunn et al., 2014; McNeish, 2018; Revelle, 2018; Zinbarg et al., 2005). However, the method for calculating ω_h relies on factor analysis, and works poorly when traits are measured with a small number of items.

The second and more sophisticated approach is to use latent variable methods—most commonly structural equation modeling (SEM)—to explicitly model the factor structure of the measures, and estimate group differences based on true instead of observed scores (Brown, 2015; Kline, 2016). To illustrate, my colleagues and I examined the effect of different correction methods on sex differences in personality from the same dataset (15 primary factors of the 16PF model in a large United States sample). The Mahalanobis distance between average profiles (male and female centroids) was $D_M = 1.49$ without error correction, went up to 1.72 after disattenuation with α, and reached 2.71 when estimated via multigroup covariance and mean structure analysis (MG-CMSA), a variant of SEM specialized for group comparisons (Del Giudice et al., 2012). Another United States dataset based on an equivalent version of the 16PF yielded $D_M = 1.18$ without correction, 1.68 after disattenuation with α, and 2.06 when estimated via MG-CMSA (Kaiser et al., 2020). For more details on these methods and additional references, see Del Giudice (2022).[10]

An Empirical Example: Three Levels of Personality

In the previous section, I discussed the behavior of within- and between-group distances using results and simulations based on idealized distributions. I now demonstrate the same

---

[10] This section takes the standard perspective on reliability, according to which a trait lies at the "intersection" of lower-level components such as facets or single items (i.e., is the source of the common variance among all the facets/items). For an alternative perspective that conceptualizes traits as the "union" of their lower-level components, see McCrae (2015) and McCrae & Mõttus (2019).

concepts by reanalyzing an empirical dataset, a large sample of personality self-reports based on the Big Five model (Kaiser, 2019; original data by Johnson, 2015; retrieved from https://osf.io/9kpc5). To eschew the methodological complications of cross-cultural comparisons, I focus on the United States subsample, which comprises $N = 617,180$ online respondents (379,323 females; for details see Kaiser, 2019). The large size of this sample also obviates the need for small-sample corrections to indices such as D_M and n_1. All analyses were performed in R 3.6 (R Core Team, 2019); the code is available at https://doi.org/10.6084/m9.figshare.13070576.

Personality was assessed with the 120-item version of the IPIP-NEO (Johnson, 2014; see http://personal.psu.edu/~j5j/IPIP/). The items (on a 1-5 scale from "very inaccurate" to "very accurate") measure 30 narrow facets of the Big Five domains (Agreeableness, Conscientiousness, Extraversion, Neuroticism, and Openness), with six facets per domain (e.g., Extraversion comprises *Friendliness*, *Gregariousness*, *Assertiveness*, *Activity level*, *Excitement seeking*, and *Cheerfulness*). Facet scores were calculated as averages of the corresponding four items, and Big Five scores were calculated as averages of six facets each. Following DeYoung et al. (2007), I also used facets to derive scores for ten personality aspects: *Compassion* and *Politeness* (for Agreeableness), *Industriousness* and *Orderliness* (for Conscientiousness), *Enthusiasm* and *Assertiveness* (for Extraversion), *Volatility* and *Withdrawal* (for Neuroticism), and *Intellect* and *Openness* (for Openness). Each aspect score was calculated as the average of one to three facets (Section S2 of the Supplement). Since the original items were not selected to specifically measure the aspects described by DeYoung and colleagues, these scores should be regarded as approximations, imperfect but useful for the purpose of this demonstration.

Correlation matrices were calculated separately for males and females. A pooled correlation matrix (weighted by sample size) was also calculated for between-group comparisons (see Del Giudice, 2009). In the pooled matrix, (not corrected for measurement error) the average absolute correlations were .23 among the Big Five, .20 among the 10 aspects, and .20 among the 30 facets. The effective dimensionality of the dataset was estimated at $n_1 = 4.3$ for the Big Five, 7.7 for aspects, and 17.5 for facets.

Individual Differences

Figure 10 displays the average within- and between-sex distances for 5, 10, and 30 personality traits. Since the 1-4 scale of raw scores is largely arbitrary and depends on the specific items chosen to measure each trait, all distances were calculated on standardized scores, and D_S was only calculated for pairwise distances. As expected, D_1 grows approximately linearly with increasing number of traits (Figure 10C), while D_2 and D_M increase at a decelerating pace (Figures 10A and 10B). Also, the average pairwise D_S quickly converges on a value of about 1.41, corresponding to a profile correlation of zero (Figure 10D). Figure 11 illustrates the full density distributions of the four metrics, in the case of within-sex pairwise distances. (The other distances displayed in Figure 10 follow qualitatively similar distributions.) A comparison of Figures 11A and 11B shows that D_M is more concentrated around the mean than D_2; however, the distributions look more similar than one may expect from the simulations in Figure 5, based on the size of trait correlations and corresponding ED values. This is explained by deviations

from normality in the raw scores, which are especially pronounced at the level of facets and some of the aspects, owing to the small number of items in each scale.

The alternative metrics in Figure 10 have different strengths and weaknesses, and can be more or less useful depending on the specific research question at hand. If the question concerns the perceived similarity of personality profiles, there are reasons to believe that D_2 or D_M are going to be the appropriate metrics in most contexts (see above). As can be seen in Figure 10A, the average person lies two standard deviations away from the centroid in the 5-dimensional space of the Big Five, and more than five standard deviations away in the 30-dimensional space of personality facets. In line with the observation by van Tilburg (2019), average profiles are remarkably uncommon, and most people have "unusual" personalities according to this criterion. Even more strikingly, the expected distance between two people selected at random ranges from about three standard deviations for the Big Five to almost eight standard deviations for facets. (Note that all these figures are lower bound estimates, due to the presence of measurement error.) To be sure, the space of personality has ample room for individuality. Indeed, finding another person with a profile that closely matches one's own is surprisingly hard, even at the comparatively low-resolution level of the Big Five. At the finer-grained level of facets, individual profiles become so unique that it becomes almost impossible to find another person within three standard deviations of oneself[11] (Figures 11A and 11B).

If one focuses exclusively on shape differences between profiles while ignoring elevation and scatter, distances center around $D_S \approx 1.41$, a value of equivalent to a correlation of zero (Figures 10D and 11D). In practical terms, this means that about half of the other people have a personality profile with peaks and valleys that broadly resembles one's own (correlations > 0), while the other half show a discordant profile (correlations < 0). As one moves from Big Five to facets and the distribution of D_S concentrates more narrowly around 1.41 (Figure 11D), the personalities of other people increasingly look "just different"—neither particularly similar nor particularly discordant. One implication is that D_S and other correlation-based indices of (dis)similarity are most informative in low-dimensional domains, and tend to lose their resolving power as the number of traits increases. (The exception is when these indices are used to detect perfect or near-perfect shape concordance, as I suggested might be the case with face recognition of the same person at different ages. If so, their detection power *increases* with the number of traits, because the tail of the distribution becomes progressively thinner in the vicinity of $r = +1$ or $D_S = 0$.) To the extent that people are sensitive to differences and similarities in the shape of personality profiles, their sensitivity should be restricted to low-dimensional comparisons involving just a few traits, such as dominance and nurturance within the interpersonal circumplex (see above).

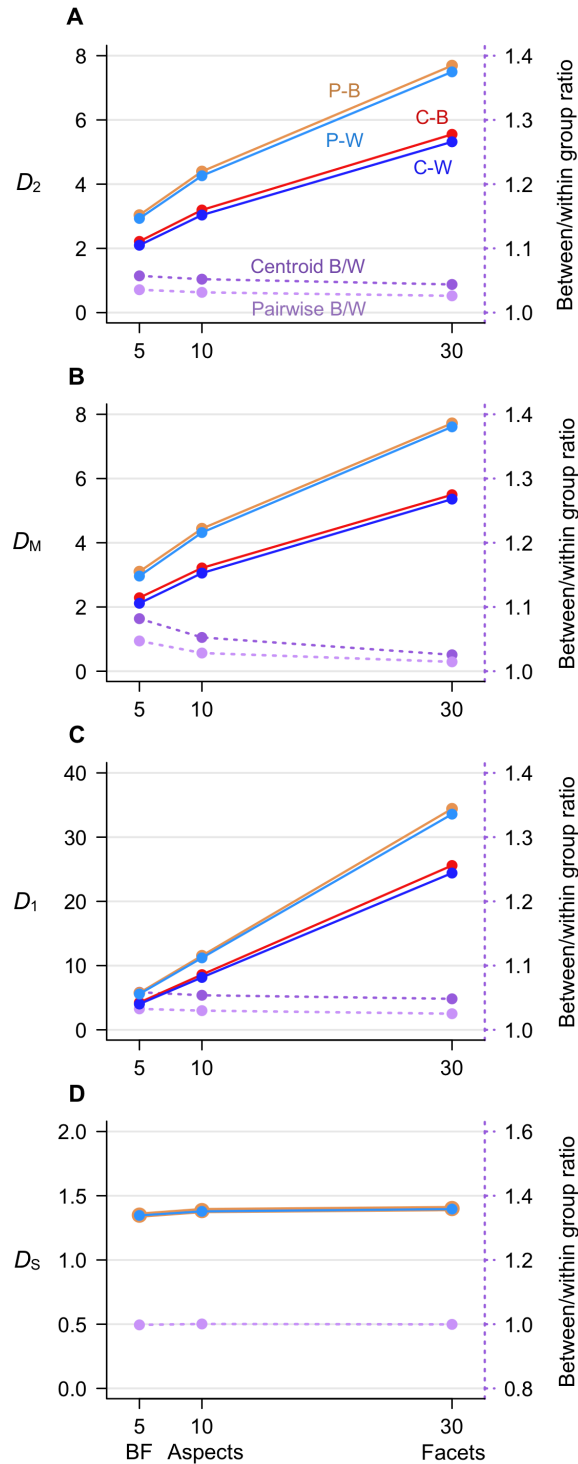[11] Which may be flattering or depressing, depending on one's disposition.

Figure 10. Empirical distance patterns in a large personality dataset. Traits were measured at three levels of resolution (5 Big Five, 10 aspects, and 30 facets). Solid lines are average distances; dotted lines show between/within sex ratios (B/W) for the same distances (right vertical axis). P-B = pairwise distance between the sexes; P-W = pairwise distance within sex; C-B = centroid distance between the sexes; C-W = centroid distance within sex. D_2 = standardized Euclidean distance; D_M = Mahalanobis distance; D_1 = standardized city-block distance; D_S = shape distance (pairwise, from standardized scores).
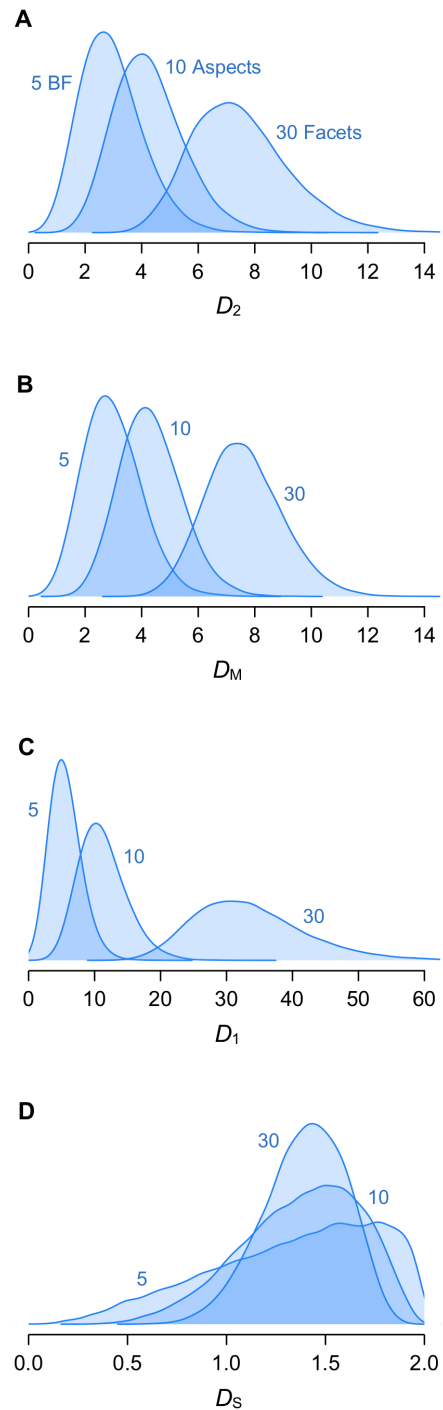
Figure 11. Empirical distribution of pairwise distances in a large personality dataset (within sex). Traits were measured at three levels of resolution (5 Big Five, 10 aspects, and 30 facets). The bandwidth for density estimation was set at 1/50th of the range. D_2 = standardized Euclidean distance; D_M = Mahalanobis distance; D_1 = standardized city-block distance; D_S = shape distance (from standardized scores).

Sex Differences

The average personality profiles of males and females in this sample are displayed in Figure 12 (raw score units). Consistent with the previous literature (see Del Giudice, 2015, 2022), females scored higher in Agreeableness and Neuroticism, with smaller differences in the other domains (Figure 12A). Aspects and facets revealed a more nuanced picture—for example, within the Openness domain, males had higher scores in Intellect, while females scored higher in the more aesthetic- and imagination-oriented Openness and the corresponding facets (Figures 12B and 12C; see Costa et al., 2001). Figure 12 shows that both males and females scored comparatively higher in Agreeableness and Conscientiousness and lower in Neuroticism, possibly due to self-evaluative biases and/or the specific mix of items included in different scales. Because it would be misleading to directly compare raw scores across traits, D_S is not a meaningful metric and is not shown in the figure.

Figure 13 shows the distance between the male and female centroids calculated with D_M and the standardized D_2 and D_1. Uncorrected distances are depicted as solid lines. Without error correction, univariate differences in the Big Five ranged from $d = -0.57$ (Extraversion) to -0.07 (Agreeableness; negative values indicate higher scores in females). The range was $d = -0.61$ to 0.22 for aspects, and $d = -0.62$ to 0.22 for facets. As noted in the previous sections, the Mahalanobis D_M is usually the most meaningful metric for comparing average profiles between groups. In this sample, the uncorrected D_M was 0.90 for the Big Five, 1.02 for aspects, and 1.23 for facets. The implied distribution overlap is about 65%, 61%, and 54%, respectively (Figure 13A). Notably, the Euclidean D_2 was smaller than D_M for the Big Five (0.53), virtually equal for aspects (1.01), and considerably larger for facets (2.57; see Figure 13B). This illustrates the importance of keeping correlations into account, and the fact that doing so may either increase or decrease the size of group differences.

Of course, uncorrected distances are deflated by measurement error—the more so for narrower traits, which are measured with fewer items. To partially compensate for this effect, I calculated the same distances after disattenuating univariate differences and correlations with Cronbach's α (dotted lines in Figure 13). The average coefficient α was .88 for the Big Five, .82 for aspects, and .77 for facets. After correction, univariate differences ranged from $d = -0.61$ to 0.08 for the Big Five, $d = -0.72$ to 0.26 for aspects, and $d = -0.76$ to 0.26 for facets. The estimated ED of the dataset showed a noticeable reduction, with $n_1 = 4.2$ for the Big Five, 6.8 for aspects, and 12.2 for facets. In total, the error-corrected Mahalanobis distance increased to $D_M = 1.05$ for the Big Five, 1.16 for aspects, and 1.74 for facets (implied overlap: 61%, 56%, and 39%, respectively). However, these values are still likely to underestimate the true differences between the sexes, owing to the limitations of α as an index of reliability. A "gold standard" for comparison is offered by the MG-CMSA analysis performed by Kaiser (2019) on the same dataset, which estimated a latent $D_M = 2.16$ from the 30 facets (corresponding to a distribution overlap of 28%; triangles in Figure 13A).
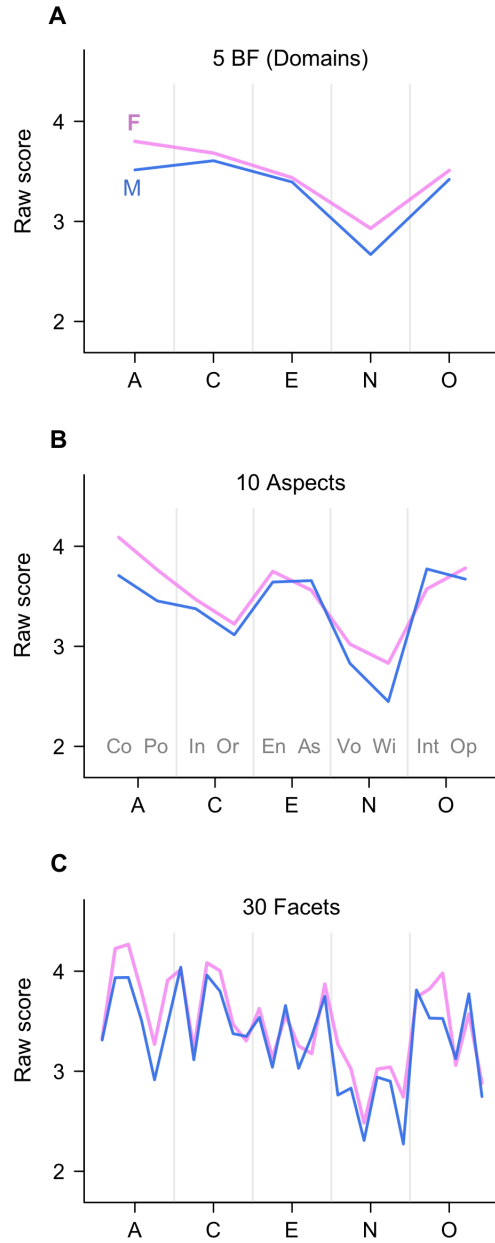
Figure 12. Mean personality profiles of males and females in a large dataset (raw scores). Traits were measured at three levels of resolution: 5 Big Five (panel A), 10 aspects (panel (B), and 30 facets (panel C). A = Agreeableness; C = Conscientiousness; E = Extraversion; N = Neuroticism; O = Openness; Co = Compassion; Po = Politeness; In = Industriousness; Or = Orderliness; En = Enthusiasm; As = Assertiveness; Vo = Volatility; Wi = Withdrawal; Int = Intellect; Op = Openness.
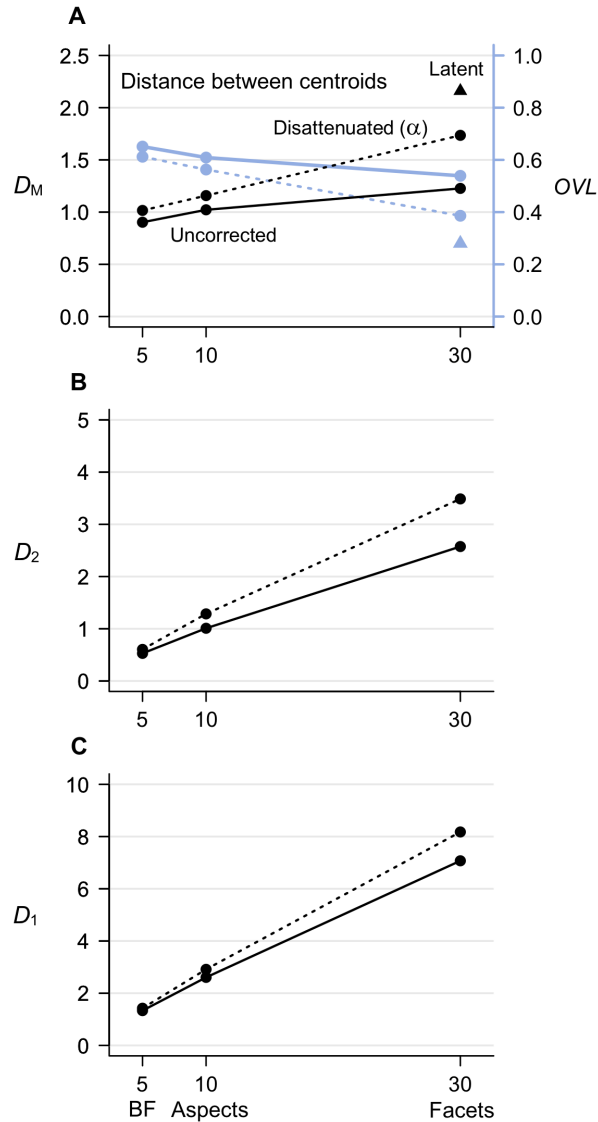
Figure 13. Empirical sex differences in personality in a large dataset, measured as distances between the male and female centroids. Traits were measured at three levels of resolution (5 Big Five, 10 aspects, and 30 facets). Panel (A) also shows the estimated proportion of overlap between distributions, measured with the overlapping coefficient *OVL* (right vertical axis). Solid lines are uncorrected distances; dotted lines are distances disattenuated with Cronbach's α; and triangles are latent differences estimated via multigroup covariance and mean structure analysis (MG-CMSA). D_2 = standardized Euclidean distance; D_M = Mahalanobis distance; D_1 = standardized city-block distance.

In sum, the overall magnitude of sex differences in personality is substantial by psychological standards (Del Giudice, 2022). Simply using participants' observed scores on this personality questionnaire, Eq. 5 estimates that one should be able to correctly classify them as male or female about 67% of the time with Big Five scores, about 70% of the time with aspect scores, and about 73% of the time with facet scores. This prediction was confirmed by LDA: with leave-one-out cross-validation and equal priors for males and females, the empirical

accuracy of the classifier was 68.2% with Big Five scores, 70.4% with aspect scores, and 73.8% with facet scores. Naturally, measurement error limits the accuracy that can be achieved from observed scores; if one could measure a person's "true" personality profile at the level of facets (as modeled in the MG-CMSA analysis by Kaiser [2019]), the expected classification accuracy would increase to a remarkable 86%.

Still, these sizable sex differences coexist with individual differences on an even larger scale, and an individual's average distance from a random member of the opposite sex is only marginally larger than the average distance from a random member of the same sex (Figures 10A-10C). As expected from simulations, between-sex distances were only slightly larger than their within-sex counterparts, with ratios consistently lower than 1.1. (Predictably, the standardized D_S was virtually identical between and within sexes, as shown in Figure 10D; compare with Figure 9C). The overall picture, then, is similar to the one I described in relation to faces. People can learn to identify male-typical and female-typical personalities with considerable accuracy, but still perceive each individual's personality as distinctive and unique, independently from their sex.

These phenomena contribute to explain an interesting "non-finding" in personality research: when unsupervised clustering algorithms are applied to personality data, they do not spontaneously recover two clusters of males vs. females—instead, they identify "types" that cut across the sexes, such as *resilient*, *overcontrolled*, and *undercontrolled* (e.g., Asendorpf et al., 2001; Donnellan & Robins, 2010; Gerlach et al., 2018).[12] This can be puzzling if one considers that there are robust differences between the average male and female profiles, to the point that one can identify people's sex with significant accuracy based on their personality scores. However, the fact that sex-based clusters fail to spontaneously emerge in unsupervised classification becomes less surprising when one considers the scale of within- and between-sex distances in the multivariate space of personality. Indeed, when the same algorithms are applied to human faces, the resulting clusters typically contain a mixture of males and females (Yang et al., 2013). In other words, it is very easy to distinguish male and female faces once the existence of two sexes is known, but it may be hard to discover the existence of two sexes from facial anatomy alone. This pattern can be expected to recur in a variety of multivariate domains, from personality and cognition to brain structure (e.g., Joel et al., 2018).[13]

Conclusion

Psychologists who study individual differences routinely deal with highly multivariate constructs, but the field has yet to fully appreciate the implications of variation in high-dimensional spaces. In this paper, I tried to clarify those implications, and demonstrate the existence of predictable phenomena that are likely to recur across different research areas. As it

---

[12] For discussion of the methodological challenges in the identification of meaningful personality types, see Gerlach et al. (2018), Freudenstein et al. (2019); Katahira et al. (2020); and Rosenström & Jokela (2017).

[13] In the neuroimaging study by Joel et al. (2018), brain structure could be used to reliably identify a person's sex (72-80% accuracy); however, unsupervised clustering failed to recover two clusters of male vs. female "brain types". The authors saw this as a demonstration that sex "is not a major predictor of the variability of human brain structure" (p. 1). I suggest that their findings can be better understood as reflecting the geometry (and relative scale) of individual and group distances in multivariate spaces.

turns out, the deceptively simple question "what happens when the number of traits increase?" also helps shed light on the properties of alternative distance metrics, both from the statistical and the cognitive perspective. An integrated understanding of multivariate differences and how to best measure them is going to greatly benefit the field, not least by suggesting many new and exciting questions for future research.

References

Aggarwal, C. C., Hinneburg, A., & Keim, D. A. (2001). On the surprising behavior of distance metrics in high dimensional space. In J. Van den Bussche & V. Vianu (Eds.), *International conference on database theory 2001* (pp. 420-434). Springer. https://doi.org/10.1007/3-540-44503-X_27

Altman, N., & Krzywinski, M. (2018). The curse(s) of dimensionality. *Nature Methods, 15,* 399-400. https://doi.org/10.1038/s41592-018-0019-x

Arabie, P. (1991). Was Euclid an unnecessarily sophisticated psychologist? *Psychometrika, 56,* 567-587. https://doi.org/10.1007/BF02294491

Asendorpf, J. B., Borkenau, P., Ostendorf, F., & Van Aken, M. A. (2001). Carving personality description at its joints: Confirmation of three replicable personality prototypes for both children and adults. *European Journal of Personality, 15,* 169-198. https://doi.org/10.1002/per.408

Ashby, F. G., & Perrin, N. A. (1988). Toward a unified theory of similarity and recognition. *Psychological Review, 95,* 124-150. https://doi.org/10.1037/0033-295X.95.1.124

Attneave, F. (1950). Dimensions of similarity. *American Journal of Psychology, 63,* 516-556. https://www.jstor.org/stable/1418869

Bennett, K. P., Fayyad, U., & Geiger, D. (1999, August). Density-based indexing for approximate nearest-neighbor queries. *Proceedings of the 5th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 233-243). https://doi.org/10.1145/312129.312236

Borg, I., Groenen, P. J. F., & Mair, P. (2018). *Applied multidimensional scaling and unfolding* (2nd ed.). Springer.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford.

Bruce, V. A., Burton, M., Hanna, E., Healey, P., Mason, O., Coombes, A., . . . Linney, A. (1993). Sex discrimination: How well do we tell the difference between male and female faces? *Perception, 22,* 131-152. https://doi.org/10.1068/p220131

Bruce, V., & Young, A. (2012). *Face perception*. Academic press.

Burton, A. M., Miller, P., Bruce, V., Hancock, P. J. B., & Henderson, Z. (2001). Human and automatic face recognition: A comparison across image formats. *Vision Research, 41,* 3185-3195. https://doi.org/10.1016/S0042-6989(01)00186-9

Burton, A. M., & Vokey, J. R. (1998). The face-space typicality paradox: Understanding the face-space metaphor. *Quarterly Journal of Experimental Psychology, 51,* 475-483. https://doi.org/10.1080/713755768

Carroll, J. D., & Wish, M. (1974). Multidimensional perceptual models and measurement methods. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of Perception, Vol. 2* (pp. 391-447). Academic Press.

Carroll, R. M., & Field, J. (1974). A comparison of the classification accuracy of profile similarity measures. *Multivariate Behavioral Research, 9,* 373-380. https://doi.org/10.1207/s15327906mbr0903_12

Cattell, H. E.P., & Schuerger, J. M. (2003). *Essentials of 16PF assessment*. Wiley.

Conroy-Beam, D., & Buss, D. M. (2017). Euclidean distances discriminatively predict short-term and long-term attraction to potential mates. *Evolution and Human Behavior, 38,* 442-450. https://doi.org/10.1016/j.evolhumbehav.2017.04.004

Conroy-Beam, D., Buss, D. M., Asao, K., Sorokowska, A., Sorokowski, P., Aavik, T., ... & Anjum, A. (2019a). Contrasting computational models of mate preference integration across 45 countries. *Scientific Reports, 9,* 1-13. https://doi.org/10.1038/s41598-019-52748-8

Conroy-Beam, D., Roney, J. R., Lukaszewski, A. W., Buss, D. M., Asao, K., Sorokowska, A., ... & Alm, C. (2019b). Assortative mating and the evolution of desirability covariation. *Evolution and Human Behavior, 40*, 479-491. https://doi.org/10.1016/j.evolhumbehav.2019.06.003

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78,* 98-104. https://doi.org/10.1037/0021-9010.78.1.98

Costa Jr, P. T., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: robust and surprising findings. *Journal of Personality and Social Psychology, 81,* 322-331. https://doi.org/10.1037/0022-3514.81.2.322

Cronbach, L. J., & Gleser, G. C. (1953). Assessing similarity between profiles. *Psychological Bulletin, 50,* 456-473. https://doi.org/10.1037/h0057173

Crutzen, R., & Peters, G. J. Y. (2017). Scale quality: alpha is an inadequate estimate and factor-analytic evidence is needed first of all. *Health Psychology Review, 11,* 242-247. https://doi.org/10.1080/17437199.2015.1124240

Davies, S. E., Connelly, B. S., Ones, D. S., & Birkland, A. S. (2015). The general factor of personality: The "Big One," a self-evaluative trait, or a methodological gnat that won't go away? *Personality and Individual Differences, 81,* 13-22. https://doi.org/10.1016/j.paid.2015.01.006

Deffenbacher, K. A., Vetter, T., Johanson, J., & O'Toole, A. J. (1998). Facial aging, attractiveness, and distinctiveness. *Perception, 27*, 1233-1243. https://doi.org/10.1068/p271233

Del Giudice, M. (2009). On the real magnitude of psychological sex differences. *Evolutionary Psychology, 7,* 264-279. https://doi.org/10.1177/147470490900700209

Del Giudice, M. (2015). Gender differences in personality and social behavior. In J. D. Wright (Ed.), *International encyclopedia of the social and behavioral sciences* (2nd ed.) (pp. 750-756). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.25100-3

Del Giudice, M. (2022). Measuring sex differences and similarities. In D. P. VanderLaan & W. I. Wong (Eds.), *Gender and sexuality development: Contemporary theory and research*. Springer.

Del Giudice, M. (2020). Effective dimensionality: A tutorial. *Multivariate Behavioral Research*. https://doi.org/10.1080/00273171.2020.1743631

Del Giudice, M., Booth, T., & Irwing, P. (2012). The distance between Mars and Venus: Measuring global sex differences in personality. *PLoS ONE, 7,* e29265. https://doi.org/10.1371/journal.pone.0029265

De Maesschalck, R., Jouan-Rimbaud, D., & Massart, D. L. (2000). The Mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems, 50,* 1-18. https://doi.org/10.1016/S0169-7439(99)00047-7

DeYoung, C. G., Peterson, J. B., & Higgins, D. M. (2002). Higher-order factors of the Big Five predict conformity: Are there neuroses of health? *Personality and Individual Differences, 33,* 533-552. https://doi.org/10.1016/S0191-8869(01)00171-4

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology, 93,* 880-896. https://doi.org/10.1037/0022-3514.93.5.880

DeYoung, C. G., Weisberg, Y. J., Quilty, L. C., & Peterson, J. B. (2013). Unifying the aspects of the Big Five, the interpersonal circumplex, and trait affiliation. *Journal of Personality, 81,* 465-475. https://doi.org/10.1111/jopy.12020

Donnellan, M. B., & Robins, R. W. (2010). Resilient, overcontrolled, and undercontrolled personality types: Issues and controversies. *Social and Personality Psychology Compass, 4,* 1070-1083. https://doi.org/10.1111/j.1751-9004.2010.00313.x

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology, 105,* 399-412. https://doi.org/10.1111/bjop.12046

Durrant, R. J., & Kabán, A. (2009). When is "nearest neighbor" meaningful: A converse theorem and implications. *Journal of Complexity, 25,* 385-397. https://doi.org/10.1016/j.jco.2009.02.011

Freudenstein, J. P., Strauch, C., Mussel, P., & Ziegler, M. (2019). Four personality types may be neither robust nor exhaustive. *Nature Human Behaviour, 3,* 1045-1046. https://doi.org/10.1038/s41562-019-0721-4

Furr, R. M. (2010). The double-entry intraclass correlation as an index of profile similarity: Meaning, limitations, and alternatives. *Journal of Personality Assessment, 92*, 1-15. https://doi.org/10.1080/00223890903379134

Garner, W. R. (1974). *The processing of information and structure*. Psychology Press.

Gerlach, M., Farb, B., Revelle, W., & Amaral, L. A. N. (2018). A robust data-driven approach identifies four personality types across four large data sets. *Nature Human Behaviour, 2,* 735-742. https://doi.org/10.1038/s41562-018-0419-z

Giraud, C. (2015). *Introduction to high-dimensional statistics*. CRC Press.

Goldstone, R. L., & Son, J. Y. (2013). Similarity. In K. J. Holyoak & R. G. Morrison (Eds.), *The Oxford handbook of thinking and reasoning* (pp. 155-176). Oxford University Press.

Gorban, A. N., Makarov, V. A., & Tyukin, I. Y. (2020). High-dimensional brain in a high-dimensional world: Blessing of dimensionality. *Entropy, 22,* 82. https://doi.org/10.3390/e22010082

Gronau, Q. F., & Lee, M. (2020). Bayesian Inference for Multidimensional Scaling Representations with Psychologically-Interpretable Metrics. *PsyArXiv,* https://doi.org/10.31234/osf.io/5zmep

Hennessy, R. J., McLearie, S., Kinsella, A., and Waddington, J. L. (2005). Facial surface analysis by 3D laser scanning and geometric morphometrics in relation to sexual dimorphism in cerebral–craniofacial morphogenesis and cognitive function. *Journal of Anatomy, 207,* 283-295. https://doi.org/10.1111/j.1469-7580.2005.00444.x

Huberty, C. J. (2005). Mahalanobis distance. In B. S. Everitt and D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (pp. 1110-1111). Wiley.

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60,* 581-592. https://doi.org/10.1037/0003-066X.60.6.581

Hyde, J. S. (2014). Gender similarities and differences. *Annual Review of Psychology, 65,* 373-398. https://doi.org/10.1146/annurev-psych-010213-115057

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2103). *An introduction to statistical learning with applications in R.* Springer.

Joel, D., Persico, A., Salhov, M., Berman, Z., Oligschläger, S., Meilijson, I., & Averbuch, A. (2018). Analysis of human brain structure reveals that the brain "types" typical of males are also typical of females, and vice versa. *Frontiers in Human Neuroscience, 12*, 399. https://doi.org/10.3389/fnhum.2018.00399

Johnson, J. A. (2014). Measuring thirty facets of the Five Factor Model with a 120-item public domain inventory: Development of the IPIP-NEO-120. *Journal of Research in Personality, 51*, 78–89. https://doi.org/10.1016/j.jrp.2014.05.003

Johnson, J. A. (2015). *Johnson's IPIP-NEO data repository*. Retrieved February 26, 2018, from the Open Science Framework website. URL: https://osf.io/tbmh5/

Jones, M., & Goldstone, R. L. (2013). The structure of integral dimensions: Contrasting topological and Cartesian representations. *Journal of Experimental Psychology: Human Perception and Performance, 39,* 111-132. https://doi.org/10.1037/a0029059

Jones, W. P., & Furnas, G. W. (1987). Pictures of relevance: A geometric analysis of similarity measures. *Journal of the American Society for Information Science, 38,* 420-442. https://doi.org/10.1002/(SICI)1097-4571(198711)38:6<420::AID-ASI3>3.0.CO;2-S

Just, C. (2011). A review of literature on the general factor of personality. Personality and *Individual Differences, 50*, 765-771. https://doi.org/10.1016/j.paid.2011.01.008

Kaiser, T. (2019). Nature and evoked culture: Sex differences in personality are uniquely correlated with ecological stress. *Personality and Individual Differences, 148,* 67-72. https://doi.org/10.1016/j.paid.2019.05.011

Kaiser, T., Del Giudice, M., & Booth, T. (2020). Global sex differences in personality: Replication with an open online dataset. *Journal of Personality, 88*, 415-429. https://doi.org/10.1111/jopy.12500

Katahira, K., Kunisato, Y., Yamashita, Y., & Suzuki, S. (2020). Commentary: A robust data-driven approach identifies four personality types across four large data sets. *Frontiers in Big Data, 3,* 8. https://doi.org/10.3389/fdata.2020.00008

Kemler Nelson, D. G. (1993). Processing integral dimensions: The whole view. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 1105–1113. https://doi.org/10.1037/0096-1523.19.5.1105

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford.

Krems, J. A., & Conroy-Beam, D. (2020). First tests of Euclidean preference integration in friendship: Euclidean friend value and power of choice on the friend market. *Evolution and Human Behavior,41,* 188-198. https://doi.org/10.1016/j.evolhumbehav.2020.02.003

Laub, J., Müller, K. R., Wichmann, F. A., & Macke, J. H. (2007). Inducing metric violations in human similarity judgements. *Advances in Neural Information Processing Systems, 19,* 777-784.

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research, 39,* 329-358. https://doi.org/10.1207/s15327906mbr3902_8

Lewis, J. P., Mo, Z., Anjyo, K., & Rhee, T. (2014). Probable and improbable faces. In K. Anjyo (Ed.), *Mathematical Progress in Expressive Image Synthesis* I (pp. 21-30). Springer. https://doi.org/10.1007/978-4-431-55007-5_4

McCrae, R. R. (2008). A note on some measures of profile agreement. *Journal of Personality Assessment, 90,* 105-109. https://doi.org/10.1080/00223890701845104

McCrae, R. R. (2015). A more nuanced view of reliability: Specificity in the trait hierarchy. *Personality and Social Psychology Review, 19,* 97-112. https://doi.org/10.1177/1088868314541857

McCrae, R. R., & Mõttus, R. (2019). What personality scales measure: A new psychometrics and its implications for theory and assessment. *Current Directions in Psychological Science, 28*, 415-420. https://doi.org/10.1177/0963721419849559

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37*, 1-10. https://doi.org/10.1016/j.intell.2008.08.004

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods, 23,* 412-433. https://doi.org/10.1037/met0000144

Meytlis, M. (2011). A model of face space. *Visual Cognition, 19,* 13-26. https://doi.org/10.1080/13506285.2010.512507

Meytlis, M., & Sirovich, L. (2007). On the dimensionality of face space. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29,* 1262-1267. https://doi.org/10.1109/TPAMI.2007.1033

Mõttus, R., Sinick, J., Terracciano, A., Hřebíčková, M., Kandler, C., Ando, J., ... & Jang, K. L. (2019). Personality characteristics below facets: A replication and meta-analysis of cross-rater agreement, rank-order stability, heritability, and utility of personality nuances. *Journal of Personality and Social Psychology, 117,* e35. https://doi.org/10.1037/pspp0000202

Murtagh, F. (2009). The remarkable simplicity of very high dimensional data: Application of model-based clustering. *Journal of Classification, 26*, 249-277. https://doi.org/10.1007/s00357-009-9037-9

Ng, C. B., Tay, Y. H., & Goi, B. M. (2015). A review of facial gender recognition. *Pattern Analysis and Applications, 18,* 739-755. https://doi.org/10.1007/s10044-015-0499-6

O'Toole, A. J., Deffenbacher, K. A., Valentin, D., McKee, K., Huff, D., & Abdi, H. (1998). The perception of face gender: The role of stimulus structure in recognition and classification. *Memory & Cognition, 26,* 146-160. https://doi.org/10.3758/BF03211378

Perrin, N. A., & Ashby, F. G. (1991). A test for perceptual independence with dissimilarity data. *Applied Psychological Measurement, 15,* 79-93. https://doi.org/10.1177/014662169101500109

R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/

Revelle, W. (2018). *An introduction to psychometric theory with applications in R*. Manuscript retrieved on October 24, 2018 from the Personality Project website: http://personality-project.org/r/book/

Revelle, W., Dworak, E. M., & Condon, D. M. (2021). Exploring the persome: The power of the item in understanding personality structure. *Personality and Individual Differences, 169,* 109905. https://doi.org/10.1016/j.paid.2020.109905

Rosenström, T., & Jokela, M. (2017). A parsimonious explanation of the resilient, undercontrolled, and overcontrolled personality types. *European Journal of Personality, 31,* 658-668. https://doi.org/10.1002/per.2117

Saucier, G. (2009). What are the most important dimensions of personality? Evidence from studies of descriptors in diverse languages. Soci*al and Personality Psychology Compass, 3,* 620-637. https://doi.org/10.1111/j.1751-9004.2009.00188.x

Sheehan, M. J., & Nachman, M. W. (2014). Morphological and population genomic evidence that human faces have evolved to signal individual identity. *Nature Communications, 5,* 1-10. https://doi.org/10.1038/ncomms5800

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237,* 1317-1323. https://doi.org/10.1126/science.3629243

Shepard, R. N. (1991). Integrality versus separability of stimulus dimensions: From an early convergence of evidence to a proposed theoretical basis, In J. R. Pomerantz & G. L. Lockhead (Eds.). *The perception of structure: Essays in honor of Wendell R. Garner* (pp. 53-71). American Psychological Association.

Sirovich, L., & Meytlis, M. (2009a). Symmetry, probability, and recognition in face space. *Proceedings of the National Academy of Sciences USA, 106,* 6895-6899. https://doi.org/10.1073/pnas.0812680106

Sirovich, L., & Meytlis, M. (2009b). *Geometry of face space*. Unpublished manuscript. Retrieved in July 2020 from: https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.84.5166

Skinner, H. A. (1978). Differentiating the contribution of elevation, scatter and shape in profile similarity. *Educational and Psychological Measurement, 38*, 297-308. https://doi.org/10.1177/001316447803800211

Stanley, D. J., & Spence, J. R. (2014). Expectations for replications: Are yours realistic? *Perspectives on Psychological Science, 9,* 305-318. https://doi.org/10.1177/1745691614528518

Stolier, R. M., Hehman, E., & Freeman, J. B. (2020). Trait knowledge forms a common structure across social cognition. *Nature Human Behaviour, 4,* 361-371. https://doi.org/10.1038/s41562-019-0800-6

Thomas, R. D. (1999). Assessing sensitivity in a multidimensional space: Some problems and a definition of a general *d'*. *Psychonomic Bulletin & Review, 6,* 224-238. https://doi.org/10.3758/BF03212328

Thomas, R. D. (2003). Further considerations of a general *d'* in multidimensional space. *Journal of Mathematical Psychology, 47*, 220-224. https://doi.org/10.1016/S0022-2496(02)00029-9

Tredoux, C. (2002). A direct measure of facial similarity and its relation to human similarity perceptions. *Journal of Experimental Psychology: Applied, 8*, 180-193. https://doi.org/10.1037/1076-898X.8.3.180

Valentine, T., Lewis, M. B., & Hills, P. J. (2016). Face-space: A unifying concept in face recognition research. *Quarterly Journal of Experimental Psychology, 69,* 1996-2019. https://doi.org/10.1080/17470218.2014.990392

van der Linden, D., Pekaar, K. A., Bakker, A. B., Schermer, J. A., Vernon, P. A., Dunkel, C. S., & Petrides, K. V. (2017). Overlap between the general factor of personality and emotional intelligence: A meta-analysis. *Psychological Bulletin, 143,* 36-52. https://doi.org/10.1037/bul0000078

van Tilburg, W. A. (2019). It's not unusual to be unusual (or: A different take on multivariate distributions of personality). *Personality and Individual Differences, 139,* 175-180. https://doi.org/10.1016/j.paid.2018.11.021

Wilson, H. R., Loffler, G., & Wilkinson, F. (2002). Synthetic faces, face cubes, and the geometry of face space. *Vision Research, 42,* 2909-2923. https://doi.org/10.1016/S0042-6989(02)00362-0

Yang, S., Bebis, G., Hussain, M., Muhammad, G., & Mirza, A. M. (2013). Unsupervised discovery of visual face categories. *International Journal on Artificial Intelligence Tools, 22*, 1250029. https://doi.org/10.1142/S0218213012500297

Zell, E., Krizan, Z., & Teeter, S. R. (2015). Evaluating gender similarities and differences using metasynthesis. *American Psychologist, 70,* 10-20. https://doi.org/10.1037/a0038208

Zimek, A., Schubert, E., & Kriegel, H. P. (2012). A survey on unsupervised outlier detection in high-dimensional numerical data. Statistical Analysis and Data Mining, 5, 363-387. https://doi.org/10.1002/sam.11161

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and McDonald's ω_H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika, 70,* 123-133. https://doi.org/10.1007/s11336-003-0974-7

Zuckerman, M., Kuhlman, D. M., Joireman, J., Teta, P., & Kraft, M. (1993). A comparison of three structural models for personality: The big three, the big five, and the alternative five. *Journal of Personality and Social Psychology, 65*, 757. https://doi.org/10.1037/0022-3514.65.4.757

Supplemental Material

**Individual and Group Differences in Multivariate Domains:
What Happens When the Number of Traits Increases?**

Marco Del Giudice

S1. Estimating the Effective Dimensionality of Facial Features

To obtain a rough estimate of the effective dimensionality (ED) of facial features, I reanalyzed published plots of the eigenvalues associated with "eigenfaces" (principal components of face pictures) from Meytlis & Sirovich (2007) and Sirovich & Meytlis (2009b). The assumptions that informed my analysis are:

- The first 100 components are sufficient to permit accurate face recognition (Burton et al., 2001; Meytlis & Sirovich, 2007; Sirovich & Meytlis, 2009a).

- The first three components usually capture broad differences in lighting and do not provide information about facial features; recognition improves after discarding them (Wilbraham et al., 2008).

- More than 30% of the remaining eigenfaces are "odd" (i.e., asymmetric around the left-right axis), and seem to represent lighting artifacts rather than facial traits; recognition improves after discarding them. This suggests that face recognition may actually rely on < 70 of the first 100 components (Sirovich & Meytlis, 2009a).

Based on these assumptions, I used published plots to estimate the first 100 (normalized) eigenvalues. I then estimated the ED after removing (a) the first 3 values, and (b) the first 3 values plus 30 randomly selected ones (to represent ~30% odd eigenfaces). I re-normalized the eigenvalues to a sum of 1, and used the resulting pseudo-probabilities (P) to calculate the n_1 index (see Del Giudice, 2020):

$$n_1 = \prod P_i^{-P_i} \tag{S1.1}$$

For each study, I repeated the randomization 1,000 times and calculated the average n_1 across repetitions. The R code for the analysis is available at https://doi.org/10.6084/m9.figshare.13070576.

Meytlis & Sirovich (2007)

Figure 1 in Meytlis & Sirovich (2007) shows a log-log plot of eigenvalues normalized to pseudo-probabilities (P) vs. component number (henceforth k). The plot is approximately linear up to about the 200th component, with slope −1.43; the log-probability of the second component

$(k = 2)$ is -1.73 (digitally measured from the plot). The log-probability of the first component was estimated at -1.30 by linear extrapolation. The equation of the line is:

$$\log_{10}(P) = -1.30 - 1.43 \log_{10}(k) \tag{S1.2}$$

which yields:

$$P = 10^{-1.30} k^{-1.43}. \tag{S1.3}$$

This equation was used to estimate the first 100 normalized eigenvalues. The average ED was $n_1 = 27.8$ when excluding 30 randomly selected values in addition to the first 3, and $n_1 = 39.4$ when only excluding the first 3 values.

Sirovich & Meytlis (2009b)

The top panel of Figure 3 in Sirovich & Meytlis (2009b) shows a similar log-log plot for the full picture set, including both male and female faces. The slope is -1.35, and the log-probability of the first component is -0.40. The equation of the line is:

$$\log_{10}(P) = -0.40 - 1.35 \log_{10}(k) \tag{S1.4}$$

which yields:

$$P = 10^{-0.40} k^{-1.35}. \tag{S1.5}$$

This equation was used to estimate the first 100 normalized eigenvalues. The average ED was $n_1 = 30.5$ when excluding 30 randomly selected values in addition to the first 3, and $n_1 = 43.4$ when only excluding the first 3 values.

References

Burton, A. M., Miller, P., Bruce, V., Hancock, P. J. B., & Henderson, Z. (2001). Human and automatic face recognition: A comparison across image formats. *Vision Research, 41,* 3185-3195. https://doi.org/10.1016/S0042-6989(01)00186-9

Del Giudice, M. (2020). Effective dimensionality: A tutorial. *Multivariate Behavioral Research*. https://doi.org/10.1080/00273171.2020.1743631

Meytlis, M., & Sirovich, L. (2007). On the dimensionality of face space. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29,* 1262-1267. https://doi.org/10.1109/TPAMI.2007.1033

Sirovich, L., & Meytlis, M. (2009a). Symmetry, probability, and recognition in face space. *Proceedings of the National Academy of Sciences USA, 106,* 6895-6899. https://doi.org/10.1073/pnas.0812680106

Sirovich, L., & Meytlis, M. (2009b). *Geometry of face space*. Unpublished manuscript. Retrieved in July 2020 from: https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.84.5166

Wilbraham, D. A., Christensen, J. C., Martinez, A. M., & Todd, J. T. (2008). Can low level image differences account for the ability of human observers to discriminate facial identity? *Journal of Vision, 8,* 5-5. https://doi.org/10.1167/8.15.5

S2. Aspect Scores in the IPIP-NEO-120

To compute scores for the 10 aspects of the Big Five (DeYoung et al., 2007) from the facets of the IPIP-NEO-120 (Johnson, 2014), I inspected the content of each IPIP-NEO-120 facet for correspondence with the items of the Big Five Aspect Scales (BFAS), as listed in Table 4 of DeYoung et al. (2007). If at least one item of a facet closely matched one of those included in the BFAS, the facet was counted as a match for the corresponding aspect. To obtain aspect scores, the IPIP-NEO-120 facets that matched each of the aspects were averaged together, yielding a score on a 1-4 range.

Agreeableness:
- *Altruism* and *Sympathy* matched the content of Compassion
- *Modesty, Cooperation,* and *Morality* matched the content of Politeness

Conscientiousness:
- *Self-discipline* matched the content of Industriousness
- *Orderliness* matched the content of Orderliness

Extraversion:
- *Friendliness* and *Cheerfulness* matched the content of Enthusiasm
- *Assertiveness* matched the content of Assertiveness

Neuroticism:
- *Anger* matched the content of Volatility
- *Anxiety, Depression,* and *Vulnerability* matched the content of Withdrawal

Openness:
- *Intellect* matched the content of Intellect
- *Imagination* and *Artistic* matched the content of Openness

References

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology, 93,* 880-896. https://doi.org/10.1037/0022-3514.93.5.880

Johnson, J. A. (2014). Measuring thirty facets of the Five Factor Model with a 120-item public domain inventory: Development of the IPIP-NEO-120. *Journal of Research in Personality, 51,* 78–89. https://doi.org/10.1016/j.jrp.2014.05.003